



# Discovery of a Novel Periodontal Disease-Associated Bacterium

Pedro J. Torres<sup>1</sup> · John Thompson<sup>1</sup> · Jeffrey S. McLean<sup>2</sup> · Scott T. Kelley<sup>1</sup> · Anna Edlund<sup>3</sup> 

Received: 1 March 2018 / Accepted: 30 April 2018  
© The Author(s) 2018

## Abstract

One of the world's most common infectious disease, periodontitis (PD), derives from largely uncharacterized communities of oral bacteria growing as biofilms (a.k.a. plaque) on teeth and gum surfaces in periodontal pockets. Bacteria associated with periodontal disease trigger inflammatory responses in immune cells, which in later stages of the disease cause loss of both soft and hard tissue structures supporting teeth. Thus far, only a handful of bacteria have been characterized as infectious agents of PD. Although deep sequencing technologies, such as whole community shotgun sequencing have the potential to capture a detailed picture of highly complex bacterial communities in any given environment, we still lack major reference genomes for the oral microbiome associated with PD and other diseases. In recent work, by using a combination of supervised machine learning and genome assembly, we identified a genome from a novel member of the Bacteroidetes phylum in periodontal samples. Here, by applying a comparative metagenomics read-classification approach, including 272 metagenomes from various human body sites, and our previously assembled draft genome of the uncultivated *Candidatus Bacteroides perioicalifornicus* (CBP) bacterium, we show CBP's ubiquitous distribution in dental plaque, as well as its strong association with the well-known pathogenic "red complex" that resides in deep periodontal pockets.

**Keywords** Periodontitis · Metagenomics · Oral microbiome · Bacteroidetes · Candidate phyla

## Introduction

Initial studies of periodontitis (PD) relied on culturing methods and traditional culture independent methods (i.e., DNA-DNA hybridization, cloning, and targeted sequencing) [1, 2], neither of which allow microbial diversity to be fully understood. The advent of culture independent high-throughput sequencing technology has

increased our understanding of the diversity of oral bacteria through two commonly used approaches: sequencing of conserved 16S ribosomal RNA genes and untargeted ("shotgun") sequencing of all ("meta") microbial genomes ("genomics"). However, because we lack reference genome sequence data for large portions of the microbial tree of life, there remains a high potential for overlooking microbes that are truly present in any given environment. To fill in some of these knowledge gaps and bypass the need for sequence homology for taxonomic classification, studies have employed contig binning, i.e., short reads assembled into contiguous sets of overlapping reads (contigs), which can be grouped into taxa based on sequence composition, similarity or read coverage [3]. One such approach includes supervised binning, which assigns contigs into taxonomic classes using a model trained with available reference sequences [4]. In a previous study, we employed this methodology on metagenomics sequence data obtained from microbial samples collected from 12 subjects with severe periodontal disease [5]. Briefly, all quality-trimmed reads were de novo assembled using

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00248-018-1200-6>) contains supplementary material, which is available to authorized users.

---

✉ Anna Edlund  
aedlund@jcvl.org

<sup>1</sup> Department of Biology, San Diego State University, San Diego, CA 92182, USA

<sup>2</sup> Department of Periodontics, University of Washington School of Dentistry, Seattle, WA 98195, USA

<sup>3</sup> J. Craig Venter Institute, Genomic Medicine Group, La Jolla, CA 92037, USA

SPAdes v 2.40 [6, 7]. SPAdes was chosen as assembly algorithm, since this program has demonstrated exceptionally high genome assembly quality as compared to other available assemblers, both single-cell assemblers as well as assemblers for multi-cell data (e.g. Velvet and SoapDeNovo) [6, 7]. We conducted post-assembly processing of contigs, which included taxonomic classification based on a machine learning algorithm using the MG Taxa tool as described earlier [4]. Several large contigs that were presented in a number of libraries had the same k-mer frequency and were originally classified at a low score to uncultivated phylum OD1, indicating they were distantly related to any previously sequenced genome [4]. These contigs, from a single sampling subject, were then sorted into a bin and further inspected for k-mer frequency consistency and used for downstream genome analyses. The assembled draft genome is 2.53 Mb and consists of 49 major contigs (sizes range between 18,374 to 129,525 bp), with an overall GC content of 59.4% (GenBank accession number: LIK00000000). Gene annotation using the Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) provided by the National Center for Biotechnology Information (NCBI) identified a total of 1875 genes, consisting of 1678 coding sequences, 39 tRNAs, and 1 rRNA operon (5S) [4]. Due to that metagenomic assembly methodologies cannot distinguish nearly identical sequences, which may originate from different genomes within the sample, our draft genome may represent several closely related bacterial strains.

Uncultivated groups, such as Candidate bacterial phyla are prevalent in the oral cavity, including Saccharibacteria/TM7, Gracilibacteria (GN02), SR1, and WPS-2 clades. Recently, strain TM7x, a member of the elusive TM7 Candidate phylum, associated with severe PD and other inflammatory conditions, was isolated from a human saliva sample [8]. Its cultivation facilitated the sequencing of a complete genome and revealed its clearly symbiotic lifestyle as the genome did not contain any amino acid biosynthetic pathways. Aside from this rare example, the ecological and clinical role of uncultivated bacteria and archaea in PD still remains a challenge.

PD, one of the world's most common infectious diseases, is a progressive polymicrobial infection that if untreated can progress to moderate and severe periodontitis. Overall, the disease refers to the inflammatory process that occurs in the tissues surrounding the teeth in response to the growth of bacterial biofilms, or dental plaque, along the gumline. Eventually, PD results in the breakdown of the periodontal ligament and alveolar bone, and can lead to loss of teeth. PD affects the majority of adults worldwide and may contribute to various

systemic diseases, including atherosclerosis (ATH), cardiovascular disease, type 2 diabetes, and rheumatoid arthritis [9]. Despite decades of research, the substantial differences among periodontitis patients in disease incidence, progressivity, and response to treatment are poorly understood.

Subgingival microbiota of periodontally healthy subjects has been shown to differ from that found in subjects with periodontal disease [10]. Studies also show that there is a striking change in the composition of the microbial profiles with greater disease severity. The shift is particularly marked for the known pathogens in the so-called red complex, i.e., *Porphyromonas gingivalis*, *Treponema denticola*, and *Tannerella forsythia* [11], whose numbers increase with pocket depth [12]. Researchers have found little or no relationship to pocket depth for the majority of other microbial species; however, most members of the so-called orange complex, which includes *Fusobacterium nucleatum* among other species, and all species of the red complex strongly associate with deeper periodontal pockets and disease severity [13]. In fact, the abundances of all members of the red complex are highly correlated, and studies have shown that co-infection with multiple members cause more severe PD than individual infections [11, 14]. The establishment of periodontal biofilms on tooth surfaces is initiated by fast growing bacterial community members of the yellow complex, such as *Streptococcus mitis* and *S. oralis*, while bridging species of the orange complex, i.e., *Fusobacterium* and late colonizers of the red complex require longer periods of time to grow [13]. Co-culturing studies have shown that members of the orange complex, particularly *F. nucleatum*, significantly enhance the growth of the more severe periodontal pathogens in the red complex [15]. It is important to note that most studies require cultivation or rely on reference databases of known sequences (16S rRNA gene and whole genomes); any bacteria that do not readily culture or are missing in the databases are ignored or missed.

Here, we were able to further characterize a recently discovered member of the *Bacteroidetes* phylum, CBP, by analyzing a total of 272 previously published metagenomes from various human body sites representing both healthy adults and adults with PD. We found that CBP is orally ubiquitous, existing in both healthy and diseased individuals, but not present in gut or skin samples. CBP also increases with increased pocket depth, co-exists with both *F. nucleatum*, *T. denticola*, and *P. gingivalis*. Its abundance is strongly correlated with members of the red complex, but not healthy commensals, all of which suggest that CBP is a novel candidate member of the symbiotic and pathogenic red complex.

## Methods

### *Candidatus Bacteroides periocalifornicus* Draft Genome Information

The draft genome LIIK00000000 was accessed via Bioproject Accession PRJNA289925, Biosample Accession SAMN03859889. Other relevant information (e.g. annotations) associated with the LIIK00000000 genome can be found via NCBI Taxon ID 1702214, IMG Submission ID: 77482, GOLD ID in IMG Database Study ID: Gs0118016 Project ID: Gp0126827, GOLD Analysis Project Id: Ga0104344.

### Maximum Likelihood Tree

Seventy-eight genomes representing major lineages from the *Bacteroidetes* phylum were downloaded from NCBI. Thirty-one taxa-specific marker genes, which were previously determined as single copy genes and unique at the nucleotide level [16] were concatenated and analyzed for optimal tree topography under evolutionary criteria by using the Molecular Evolutionary Genetics Analysis (MEGA) software, version 6.0 [17]. Five thousand bootstrap iterations were performed.

### HMP Dataset

The National Institutes of Health (NIH) Human Microbiome Project (HMP) was established by the NIH Common Fund (<http://commonfund.nih.gov/hmp/>) to provide a public resource to facilitate human microbiome research [18]. Two hundred and fourteen metagenomes were obtained from the HMP whole metagenomics shotgun sequencing website (<https://www.hmpdacc.org/HMIWGS/healthy/>). This included 18 gut, 14 left retroauricular crease a.k.a. skin, 6 saliva, 16 subgingival, and 160 supragingival datasets (Table S1).

### Human Oral Microbiome Datasets

Published metagenomic libraries, representative of both healthy and diseased subjects, were obtained from the Human Oral Microbiome Database (HOMD) under the submission number 20130522 ([ftp://ftp.homd.org/publication\\_data/20130522/](ftp://ftp.homd.org/publication_data/20130522/)) and from a study by Duran-Pinedo and colleagues [19], respectively. In all studies, healthy and periodontitis subjects were diagnosed by a clinician. The datasets included subgingival samples from six healthy individuals and seven individuals diagnosed with periodontitis. One healthy individual (Metagenome\_Healthy2) was excluded from the analysis due to its abnormally high level of CBP (0.63 versus a mean of 0.02 for all samples).

### American Indian/Alaskan Native Study Dataset

This dataset included metagenomes generated from 22 subgingival samples from 12 different patients recruited from an American Indian/Alaskan Native population in Southern California [20]. See previous publication for details on sampling methods, disease classification, DNA extraction, and study population [21]. The samples included ten sample pairs from the same patient, one obtained before, and one after standard periodontal treatment. Participants were classified to various degrees of periodontitis based on periodontal pocket depth (PPD), clinical attachment loss (CAL), plaque score, and bleeding on probing (BOP). Individuals with PPD  $\leq$  4 mm, CAL  $\leq$  3, and BOP  $>$  10% were classified as having gingivitis; individuals with PPD  $\geq$  5 mm, CAL  $\geq$  4, and BOP  $\geq$  30% were classified as having mild-moderate periodontitis; and individuals with PPD  $\geq$  7 mm, CAL  $\geq$  6, and BOP  $\geq$  30% were classified as having severe periodontitis.

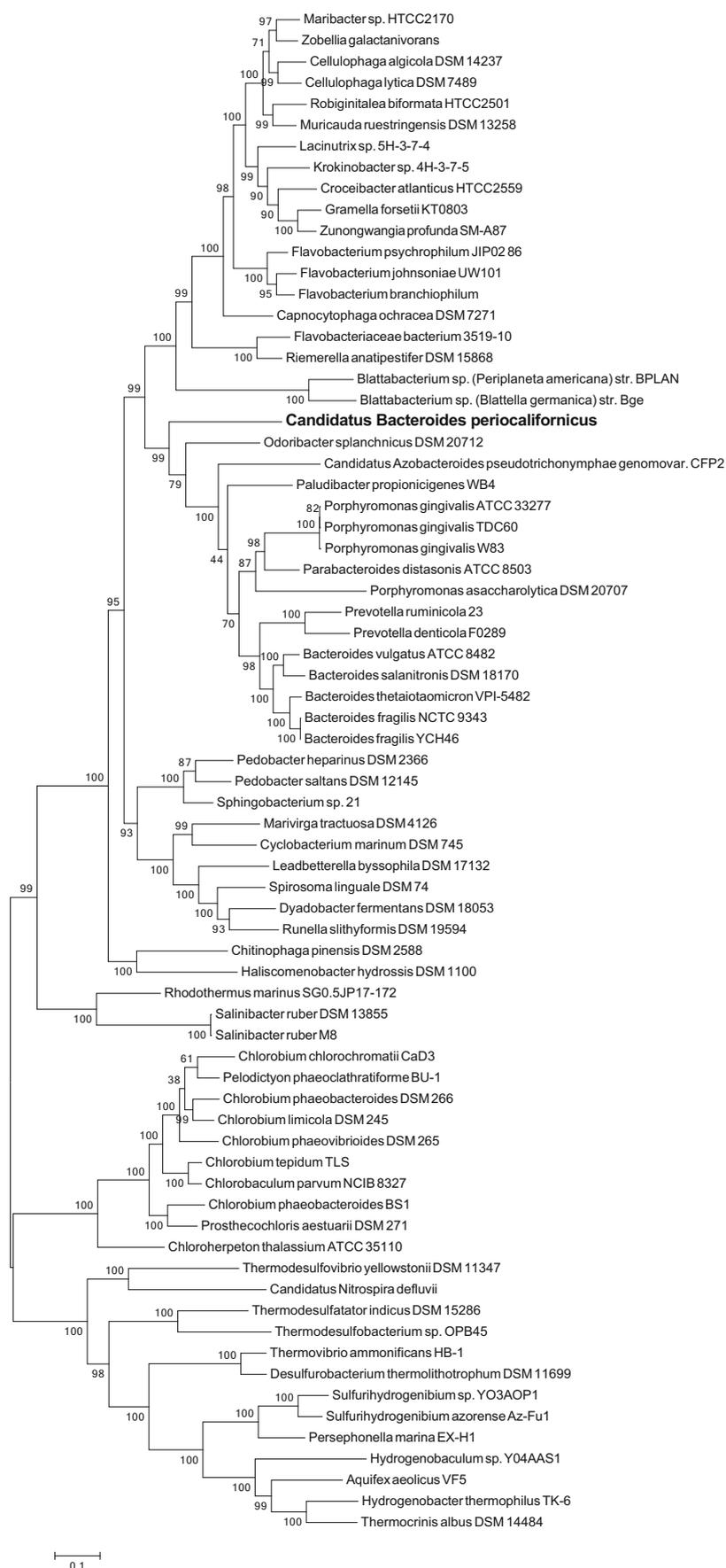
### University of Southern California Study Dataset

This data set includes metagenomes generated from 24 subgingival samples from patients treated at the graduate periodontology clinic at the Herman Ostrow School of Dentistry of the University of Southern California (USC). Information on molecular and clinical methods and study population can be found in a previous study by Califf and colleagues [22]. Participants were recruited as part of a study investigating the effectiveness of dilute sodium hypochlorite on periodontitis. Each participant received a comprehensive clinical examination, and was randomly assigned to a control or treatment group [22, 23]. No scaling was performed before or during the treatment. Each patient exhibited at least four separate teeth with a pocket depth of  $\geq$  6 mm. Pocket depth categories were as follows: class A = periodontal pocket depth up to 6 mm, class b = 6–8 mm, and class c  $>$  8 mm.

### Metagenomic Sequence Processing and Analysis

Metagenome reads were trimmed using the Trimmomatic (v.0.36) [24] default settings (<http://www.usadellab.org/cms/?page=trimmomatic>). Metagenomes were then subjected to stringent error filtering using PRINSEQ (v.0.20.4) [25] with the following parameters: minimum sequence length of 60 bp, minimum mean quality score of 25, sequences containing any “N’s” were removed, and low-complexity threshold of 50 (using entropy). Human DNA was filtered out using the DeconSeq software (coverage  $>$  90, identity  $>$  90) (v.0.4.3) [26]. After DeconSeq, paired-end files were rewritten to make sure all reads had a mate and separated out singletons using FASTQ Pair, available at <https://github.com/linsalrob/EdwardsLab/>. BBMerge (v.3.36) [27] was used to merge overlapping pairs of reads using

**Fig. 1** Phylogenetic relatedness of CBP to bacterial lineages within the phylum *Bacteroidetes*. Phylogenetic tree from Maximum Likelihood estimation of 31 marker genes [16] representing 78 different lineages of bacteria within the Bacteroidetes phylum. Five thousand bootstrap iterations were performed. The analysis placed CBP (boldface in figure) deeply within the *Bacteroides*, as sister-group to the genera *Odoribacter*, *Paludibacter*, *Porphyromonas*, *Parabacteroides*, and *Prevotella*>



default parameters. Forward reads showed very high quality scores; therefore, those that did not merge were extracted from the bbmerge unmerged output file ([https://github.com/pjtorres/xtract\\_forward](https://github.com/pjtorres/xtract_forward)) as to not discard useful data.

Metagenomes were analyzed using Kraken [28]. Kraken is able to assign taxonomic labels to short DNA sequences with high sensitivity and speed by utilizing exact alignment of short subsequences of length  $k$ , called  $k$ -mers (default size  $k = 31$ ), and a novel classification algorithm. Kraken first uses a reference database and builds a new database by adding phylogenetic information to every  $k$ -mer in its database. Kraken then classifies reads by breaking each read into overlapping  $k$ -mers. Each  $k$ -mer is then mapped to the lowest common ancestor of the genomes containing that  $k$ -mer in the precomputed database. For our study, we built a custom database containing all complete bacterial reference genomes from the NCBI refseq database (<ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/bacteria/>) and CBP.

## Statistical Analysis

Pearson's product-moment correlation was performed when analyzing CBP relative abundance over pocket depth using the RStudio statistical package (version 1.0.153). Kruskal-Wallis nonparametric tests were used to determine whether the relative abundance of CBP, *P. gingivalis*, or *F. nucleatum* differed between two groups and this was followed by post-hoc Dunn's

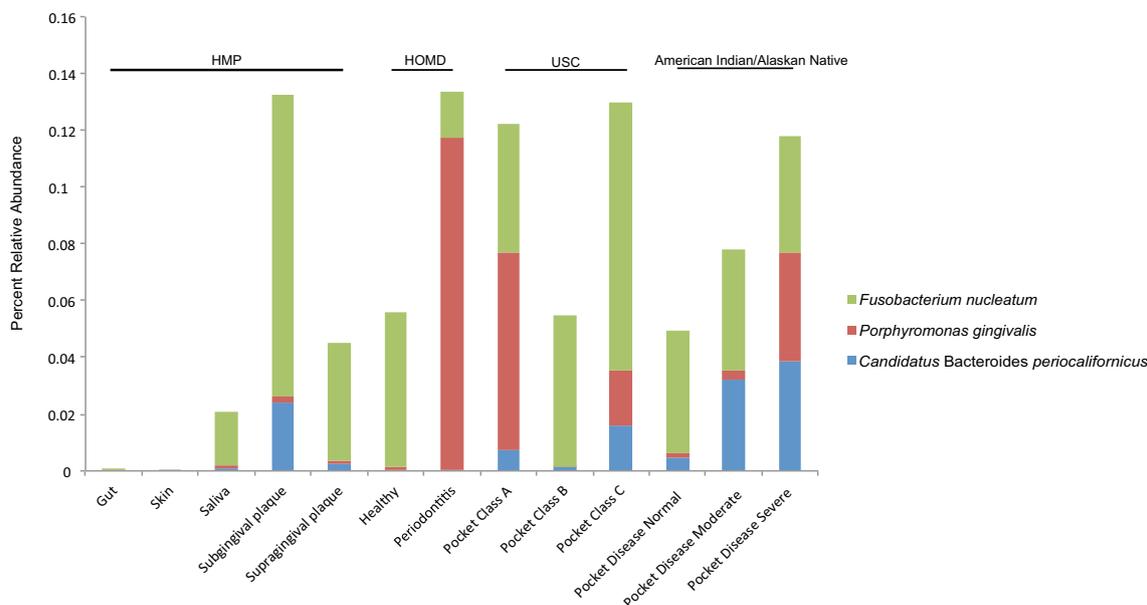
multiple comparisons test when comparing three or more groups.

## Genome Mining of Virulence Factors and Biosynthetic Gene Clusters

JGI IMG genome portal analysis pipeline (<https://img.jgi.doe.gov/>) was used to assess virulence properties of the CBP genome. The antiSMASH tool (the Bacterial version) [29] was applied to search the genome for biosynthetic gene clusters. Default settings were applied.

## Results and Discussion

We used the metagenomic taxonomic classification tool Kraken [28] to investigate the relationship of the uncultivated CBP to periodontal disease and to members of key periodontal pathogens in the orange and red complex. CBP had previously been identified in a metagenome from a patient with severe PD in high relative abundance [5]. By performing Maximum Likelihood estimates using multiple marker genes from this genome and 78 additional sequenced genomes, representing a broad diversity of bacteria within the Bacteroidetes phylum; we found that CBP was deeply placed within the *Bacteroides*, as sister-group to the genera *Odoribacter*, *Paludibacter*, *Porphyromonas*, *Bacteroides*, *Parabacteroidetes*, and *Prevotella* (Fig. 1). Its closest neighbor genome was



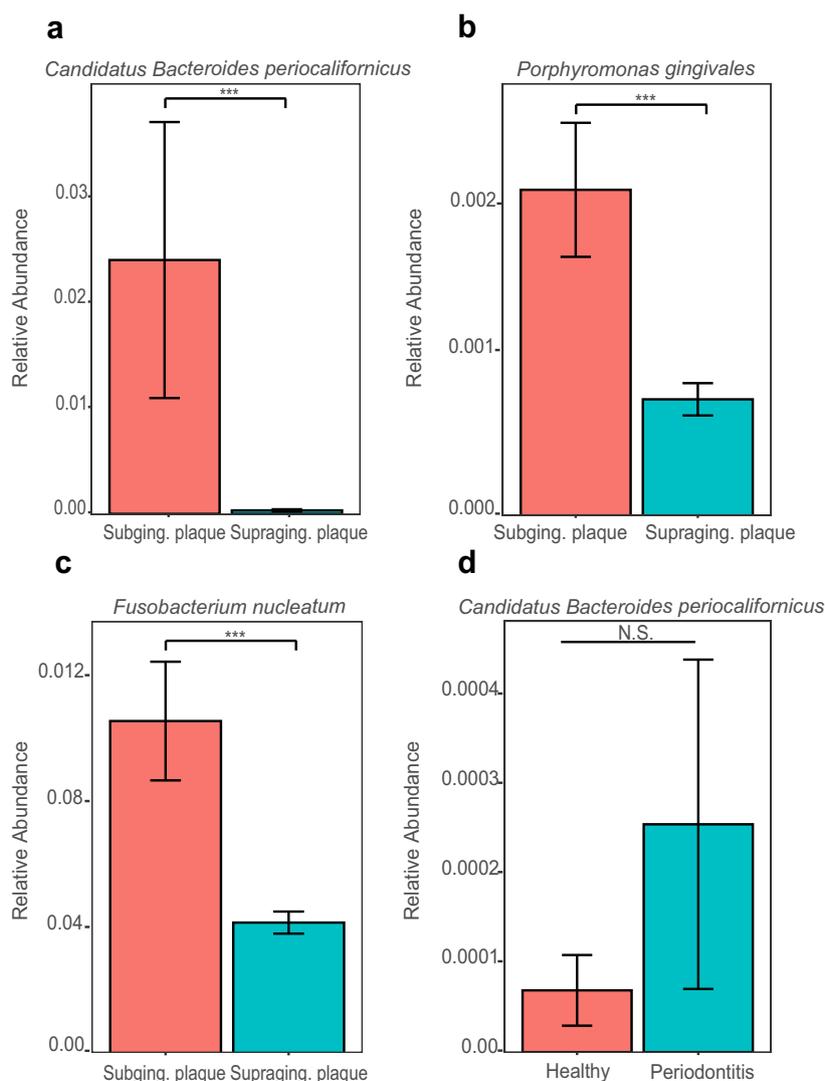
**Fig. 2** *B.periocaliformicus* (CBP) detected in oral sites but not the skin or gut. Stacked bar plot representing the relative abundance (%) of CBP, *P. gingivalis*, and *F. nucleatum* from a combined total of 272 metagenomes generated from different body sites, including gut ( $n = 18$ ), skin ( $n = 14$ ), saliva ( $n = 6$ ), sub- ( $n = 16$ ), and supragingival plaque ( $n = 160$ ) from the HMP datasets; healthy

individuals ( $n = 5$ ) and individuals diagnosed with periodontitis ( $n = 7$ ) from HODM datasets [19]; oral disease class A ( $n = 10$ ), disease class B ( $n = 4$ ), and disease class C ( $n = 10$ ) from USC dataset [22]; oral pocket disease state ranging from normal ( $n = 3$ ), moderate ( $n = 10$ ), and severe ( $n = 9$ ) from the American Indian/Alaskan Native dataset [21]

*Alistipes putredinis* with an average nucleotide identity of 68% [5].

To deepen our understanding of the ecology and biogeographic distribution of CBP, we used the Kraken tool to determine its general abundance in the oral cavity and whether it was specific to the oral cavity or found commonly in other niches of the human microbiome. We analyzed a total of 272 previously published metagenomes from various human body sites representing both healthy and non-healthy adults (Table S1), including: gut ( $n = 18$ ), skin ( $n = 14$ ) saliva ( $n = 6$ ), sub- ( $n = 21$ ) and supragingival plaque ( $n = 160$ ), and adults

with periodontal disease (PD) ( $n = 53$ ). By performing multiple comparative nonparametric statistical tests and correlation analyses on the normalized relative abundance values obtained by Kraken, we found that CBP was unique to the oral cavity (Fig. 2), and that the relative abundance of CBP was significantly higher in subgingival plaque compared to supragingival plaque and saliva samples (Fig. S1 and Fig. 3a). These results strongly indicate that CBP is specifically adapted to the subgingival environment, where its relative abundance 0.024 (2.4%) is similar to that of *P. gingivalis* 0.002 (0.2%) and *F. nucleatum* 0.01 (1.0%) (Fig. 3b, c). Comparisons of



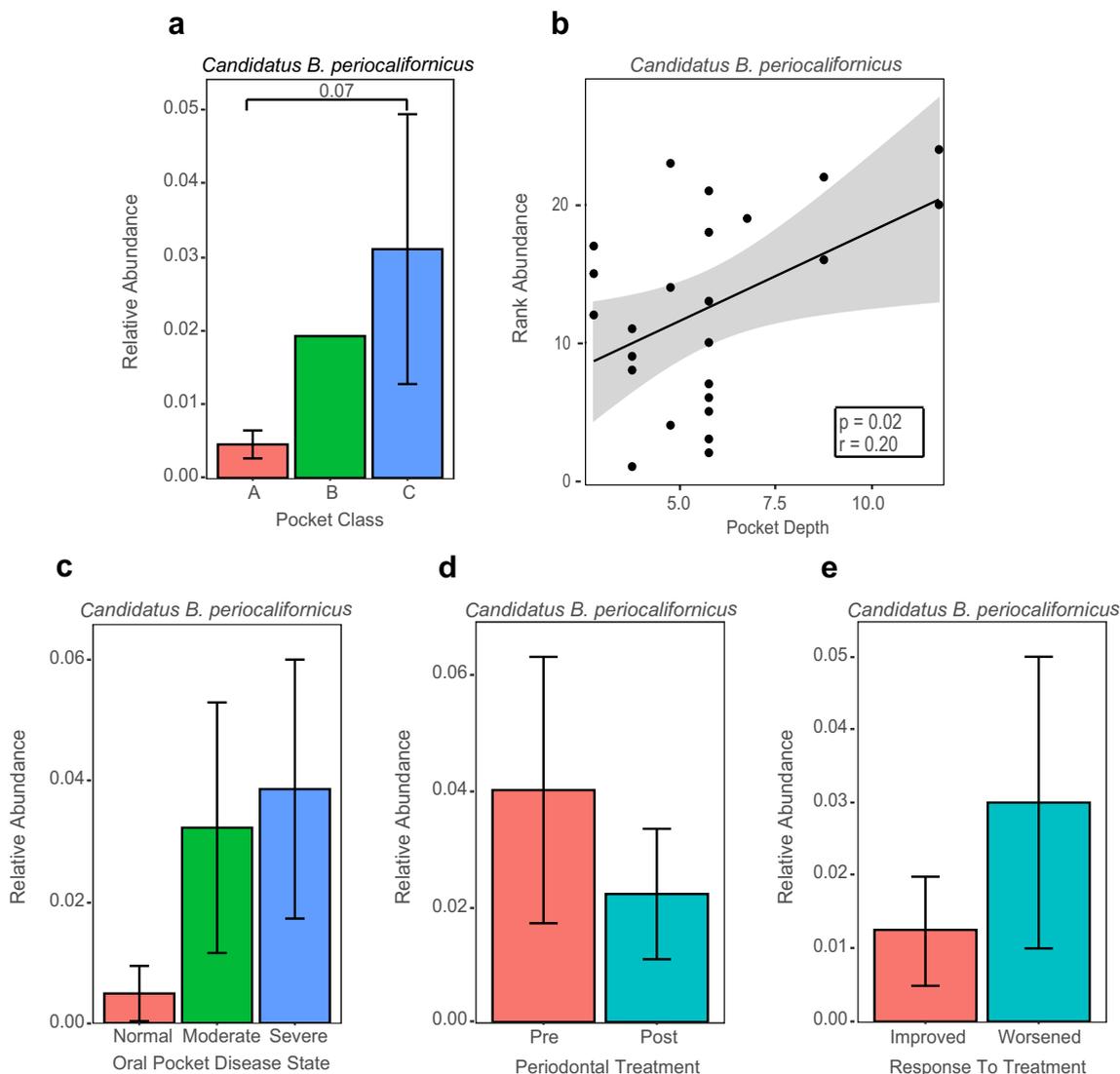
**Fig. 3** Relative abundance comparisons of *B. periocalifornicus* (CBP) in subgingival plaque. Relative proportions of CBP in healthy individuals was compared to individuals diagnosed with periodontal disease (PD), as well as proportions of *P. gingivalis* and *F. nucleatum*. CBP is highly abundant in subgingival as compared to supragingival plaque. Bar plot (mean  $\pm$  SEM) showing the relative abundance of **a** *B. californicus*, **b** *P. gingivalis*, and **c** *F. nucleatum* in sub- ( $n = 16$ ) and supragingival plaque ( $n = 160$ ). Datasets were obtained from the publicly available National

Institutes of Health Human Microbiome Project. Dataset for the relative abundance of **d** CBP in subgingival samples from healthy individuals ( $n = 5$ ) and individuals diagnosed with PD ( $n = 7$ ) were obtained from the Human Oral Microbiome Database (HOMD) under the submission number 20130522 (Table S1). ([ftp://ftp.homd.org/publication\\_data/20130522/](ftp://ftp.homd.org/publication_data/20130522/)). Kruskal-Wallis nonparametric test was performed to compare means among the two groups; \*\*\*  $p \leq 0.001$

sequence libraries between healthy subjects and subjects with PD found CBP to be enriched in patients with disease (Fig. 3d).

The association of CBP and the periodontal pocket milieu was further explored by analyzing the presence of CBP in sequence libraries obtained from subjects with different levels of severe PD (detailed information on how PD was diagnosed and how samples were collected can be found in a previous study by Califf and colleagues [22]). The mean relative

abundance of CBP was greater in deeper pockets (Kruskal-Wallis,  $p = 0.07$ ; Fig. 4a), and a positive correlation between CBP relative abundance and pocket depth was observed (Pearson's;  $p = 0.02$ ; Fig. 4b), as well as a trend for more severely disease periodontal pockets and a higher relative abundances of CBP (Kruskal-Wallis,  $p = 0.37$ ; Fig. 4c). The latter correlation was not significant, which is likely due to the high within group variability (a common feature of human



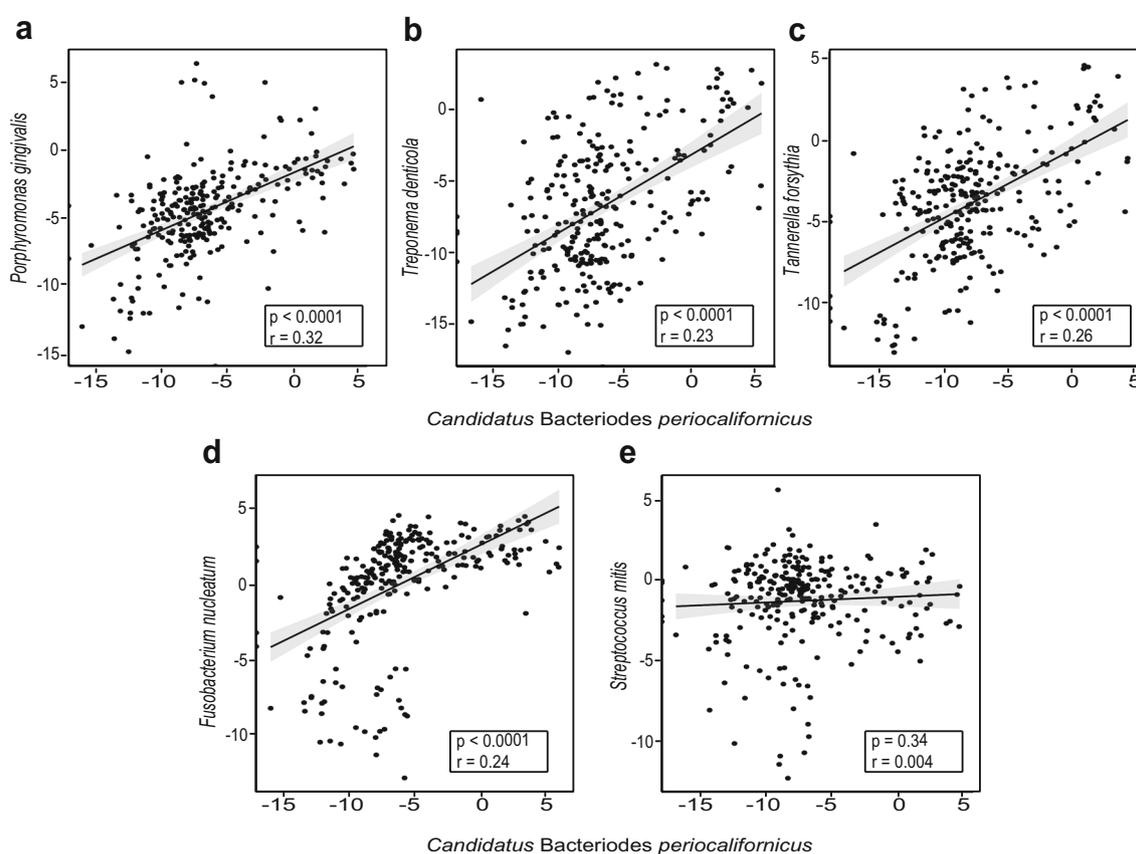
**Fig. 4** Relative abundance of *Candidatus B. periocaliformicus* (CBP) in patients with severe periodontal disease (PD). Relative abundance of CBP increases with worsening oral disease class, pocket class, pocket depth and is more abundant in patients with severe oral pocket disease state; decreased posttreatment. Bar plot (mean ± SEM) showing the relative abundance of *Candidatus Bacteroides periocaliformicus* from 24 subjects grouped based on **a** Subjects were also grouped based on pocket class A ( $n = 18$ ), pocket class B ( $n = 1$ ) and pocket class C ( $n = 5$ ) (Kruskal-Wallis,  $p$  value = 0.07). **b** Scatterplot and trend line showing the relationship between pocket depth and ranked CBP relative abundance. Results of Pearson's correlation ( $p$  value = 0.02) and correlation coefficient are shown in box inset with gray shaded area

indicating the 95% confidence interval for the line of best fit. Bar plot (mean ± SEM) representing the relative abundance of CBP from 22 subjects based on **c** oral pocket disease state ranging from normal ( $n = 3$ ), moderate ( $n = 10$ ), and severe ( $n = 9$ ) (Kruskal-Wallis,  $p$  value = 0.37). **d** There were 11 subjects in the pre- and 11 in the posttreatment groups [22] (Kruskal-Wallis,  $p$  value = 0.65). **e** Five subjects improved after standard periodontal treatment and six worsened (Kruskal-Wallis,  $p$  value = 0.47). Kruskal-Wallis nonparametric tests followed by post hoc Dunn's multiple comparisons tests were used when comparing three groups. One-way ANOVA on ranked data was used to compare the means between the two groups

microbiome sequence data). In addition, by comparing sequence libraries from PD-patients who were subjected to a 0.25% sodium hypochlorite (diluted bleach) treatment, and whose conditions either improved or worsened after treatment [22], a positive trend was observed, showing a higher abundance of CBP in samples before treatment, and in samples that worsened after treatment (Kruskal-Wallis,  $p = 0.65$ ; Fig. 4d,  $p = 0.47$ ; Fig. 4e). The relative abundance of CBP was also significantly correlated with all members of the red complex (Fig. 5a–c) as well as the orange complex member *F. nucleatum* (Fig. 5d). Furthermore, no correlation was observed between the abundance of CBP and *S. mitis*, a common oral commensal bacterium (Fig. 5e). Intriguingly, our comparative metagenomic read abundance analysis approach of the CBP draft genome not only reveals an orally ubiquitous bacterium, specifically adapted to the subgingival plaque environment, but also a novel candidate member of the pathogenic red complex.

To further elucidate the functional capacity of CBP we employed the JGI IMG genome portal analysis pipeline available at <https://img.jgi.doe.gov/>, and identified a number of

virulence-associated genes. This analysis showed that the CBP genome encode the flagellar assembly proteins CheA, CheB, CheR, CheW, and CheY, which are involved in chemotaxis (i.e., direct movement toward an attractant or away from a repellent), suggesting that CBP is motile and also harbors genes that are key in adhesion to a host and in host invasion [30]. The genome also includes multiple genes involved in beta-lactam resistance, which is in line with numerous studies showing that *Bacteroides* species have the broadest spectrum of resistance to commonly used antimicrobial agents, especially to beta-lactam compounds [31]. In addition, the genome harbors the *rfbA*, *rfbB*, *rfbC*, and *rfbCD* genes, which encode enzymes that are involved in the biosynthesis of dTDP-rhamnose for the assembly of lipopolysaccharide (LPS), suggesting that CBP may have antagonist LPS structures, similar to other *Bacteroidetes*, such as *P. gingivalis* and *Tannerella* [32]. Two antioxidant enzymes were also identified (a peroxiredoxin, and a 1-Cys-peroxiredoxin), which are known to control cytokine-induced peroxide levels and are thereby mediating signal transduction in mammalian cells [33]. Furthermore, by performing BLAST analysis of the



**Fig. 5** *Candidatus B. periocaliformicus* (CBP) relative abundance is strongly correlated with all three members of the red complex. Scatterplot and trend line showing the relationship between CBP and **a** *P. gingivalis*, **b** *T. denticola*, **c** *T. forsythia*, **d** *F. nucleatum*, and **e** *S. mitis*.

Results of Pearson's correlation ( $p$  value and correlation coefficient) are shown in the box inset with gray shaded area indicating the 95% confidence interval for the line of best fit ( $n = 272$ ). Relative abundance data was log<sub>2</sub> transformed (normalized)

CBP genome against the well annotated *P. gingivalis* ATCC 33277 genome, we identified the following shared virulence-associated genes: C25 domains encoding gingipains, which are well-known *P. gingivalis* proteases that target outer membranes via the *Bacteroidetes*-specific type 9 secretion system, *ragA* and *ragB* surface antigen genes, and hemolysin encoding genes (Table S2). To further explore the capacity of CBP to produce bioactive small molecules, we applied the antiSMASH software [29]. This analysis predicted that the genome harbors seven putative biosynthetic gene clusters (Table S3) of which one was associated with S-layer glycan biosynthesis, that supports glycosylation of proteins, and gives the cell membrane fluidity, i.e., it is important for gliding motility. Another cluster encode an arylpolyene-like molecule, which corresponds to flexirubin—a pigment associated with all *Bacteroidetes* bacteria, and that is known to protect the cell from oxidative stress [34]. An O-antigen biosynthetic gene cluster, encoding a group of molecules that is known for being important in interactions with other bacterial cells and with human host cells was also identified [35].

Based on all the above findings, we suggest that the CBP genome harbors several genes and pathways similar to the other known oral pathogens belonging to the *Bacteroidetes* phylum, which are involved in PD-associated virulence, including host cell modulation, which strengthens the evidence that this bacterium is a new member of the red complex. A goal is to further elucidate CBP's role in oral health and disease by attempting various cultivation approaches which, if successful, would allow us to study CBP in the research laboratory and to obtain a complete genome sequence. Also, by applying fluorescent staining techniques, targeting CBP and other oral bacteria, we could learn more about its spatial distribution and physical interactions with other biofilm community members.

**Acknowledgements** We thank Rob Edwards for the use of the Anthill computational cluster at San Diego State University.

**Author Contributions** All authors contributed extensively to the work presented in this paper.

**Funding Information** Part of this work was funded by the NIH grant number U26IHS300292 to Dan Calac and Roberta Gottlieb. NIH NIDCR Awards DE023810 (J.S.M.), R00DE024543 (A.E.).

## Compliance with Ethical Standards

**Competing Interests** The authors declare that they have no competing interests.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Chen H, Jiang W (2014) Application of high-throughput sequencing in understanding human oral microbiome related with health and disease. *Front. Microbiol.* 5:508. <https://doi.org/10.3389/fmicb.2014.00508>
- Woyke T, Teeling H, Ivanova NN, Huntemann M, Richter M, Gloeckner FO, Boffelli D, Anderson IJ, Barry KW, Shapiro HJ, Szeto E, Kyrpides NC, Mussmann M, Amann R, Bergin C, Ruehland C, Rubin EM, Dubilier N (2006) Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* 443:950–955. <https://doi.org/10.1038/nature05192>
- Mande SS, Mohammed MH, Ghosh TS (2012) Classification of metagenomic sequences: methods and challenges. *Brief. Bioinform.* 13:669–681. <https://doi.org/10.1093/bib/bbs054>
- McLean JS, Lombardo MJ, Badger JH, Edlund A, Novotny M, Yee-Greenbaum J, Vyahhi N, Hall AP, Yang Y, Dupont CL, Ziegler MG, Chitsaz H, Allen AE, Yooseph S, Tesler G, Pevzner PA, Friedman RM, Nealson KH, Venter JC, Lasken RS (2013) Candidate phylum TM6 genome recovered from a hospital sink biofilm provides genomic insights into this uncultivated phylum. *Proc. Natl. Acad. Sci. U. S. A.* 110:E2390–E2399. <https://doi.org/10.1073/pnas.1219809110>
- McLean JS, Liu Q, Thompson J, Edlund A, Kelley S (2015) Draft genome sequence of “Candidatus *Bacteroides* pericalifornicus,” a new member of the *Bacteroidetes* phylum found within the oral microbiome of periodontitis patients. *Genome Announc* 3: e01485–e01415. <https://doi.org/10.1128/genomeA.01485-15>
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19: 455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Pribelski AD, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, Clingenpeel SR, Woyke T, McLean JS, Lasken R, Tesler G, Alekseyev MA, Pevzner PA (2013) Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J. Comput. Biol.* 20:714–737. <https://doi.org/10.1089/cmb.2013.0084>
- He X, McLean JS, Edlund A, Yooseph S, Hall AP, Liu SY, Dorrestein PC, Esquenazi E, Hunter RC, Cheng G, Nelson KE, Lux R, Shi W (2015) Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. *Proc. Natl. Acad. Sci. U. S. A.* 112:244–249. <https://doi.org/10.1073/pnas.1419038112>
- Kim J, Amar S (2006) Periodontal disease and systemic conditions: a bidirectional relationship. *Odontology* 94:10–21. <https://doi.org/10.1007/s10266-006-0060-6>
- Ximenez-Fyvie LA, Haffajee AD, Socransky SS (2000) Microbial composition of supra- and subgingival plaque in subjects with adult periodontitis. *J. Clin. Periodontol.* 27:722–732
- Socransky SS, Haffajee AD, Cugini MA, Smith C, Kent Jr RL (1998) Microbial complexes in subgingival plaque. *J. Clin. Periodontol.* 25:134–144
- Mineoka T, Awano S, Rikimaru T, Kurata H, Yoshida A, Ansai T, Takehara T (2008) Site-specific development of periodontal disease is associated with increased levels of *Porphyromonas gingivalis*, *Treponema denticola*, and *Tannerella forsythia* in subgingival plaque. *J. Periodontol.* 79:670–676. <https://doi.org/10.1902/jop.2008.070398>
- Teles R, Teles F, Frias-Lopez J, Paster B, Haffajee A (2013) Lessons learned and unlearned in periodontal microbiology. *Periodontol.* 62:95–162. <https://doi.org/10.1111/prd.12010>

14. Hajishengallis G, Lamont RJ (2012) Beyond the red complex and into more complexity: the polymicrobial synergy and dysbiosis (PSD) model of periodontal disease etiology. *Mol Oral Microbiol* 27:409–419. <https://doi.org/10.1111/j.2041-1014.2012.00663.x>
15. Ebersole JL, Feuille F, Kesavalu L, Holt SC (1997) Host modulation of tissue destruction caused by periodontopathogens: effects on a mixed microbial infection composed of *Porphyromonas gingivalis* and *Fusobacterium nucleatum*. *Microb. Pathog.* 23:23–32. <https://doi.org/10.1006/mpat.1996.0129>
16. Wu M, Eisen JA (2008) A simple, fast, and accurate method of phylogenomic inference. *Genome Biol.* 9:R151. <https://doi.org/10.1186/gb-2008-9-10-r151>
17. Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725–2729. <https://doi.org/10.1093/molbev/mst197>
18. Group NHW, Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA, Deal C, Baker CC, Di Francesco V, Howcroft TK, Karp RW, Lunsford RD, Wellington CR, Belachew T, Wright M, Giblin C, David H, Mills M, Salomon R, Mullins C, Akolkar B, Begg L, Davis C, Grandison L, Humble M, Khalsa J, Little AR, Peavy H, Pontzer C, Portnoy M, Sayre MH, Starke-Reed P, Zakhari S, Read J, Watson B, Guyer M (2009) The NIH human microbiome project. *Genome Res.* 19:2317–2323. <https://doi.org/10.1101/gr.096651.109>
19. Duran-Pinedo AE, Chen T, Teles R, Starr JR, Wang X, Krishnan K, Frias-Lopez J (2014) Community-wide transcriptome of the oral microbiome in subjects with and without periodontitis. *ISME J* 8:1659–1672. <https://doi.org/10.1038/ismej.2014.23>
20. Kumar PKV, Gottlieb RA, Lindsay S, Delange N, Penn TE, Calac D, Kelley ST (2018) Metagenomic analysis uncovers strong relationship between periodontal pathogens and vascular dysfunction in American Indian population. *bioRxiv In Rev.* <https://doi.org/10.1101/250324>
21. Schwarzberg K, Le R, Bharti B, Lindsay S, Casaburi G, Salvatore F, Saber MH, Alonaizan F, Slots J, Gottlieb RA, Caporaso JG, Kelley ST (2014) The personal human oral microbiome obscures the effects of treatment on periodontal disease. *PLoS One* 9:e86708. <https://doi.org/10.1371/journal.pone.0086708>
22. Califf KJ, Schwarzberg-Lipson K, Garg N, Gibbons SM, Caporaso JG, Slots J, Cohen C, Dorrestein PC, Kelley ST (2017) Multi-omics analysis of periodontal pocket microbial communities pre- and posttreatment. *mSystems* 2. <https://doi.org/10.1128/mSystems.00016-17>
23. Galvan M, Gonzalez S, Cohen CL, Alonaizan FA, Chen CT, Rich SK, Slots J (2014) Periodontal effects of 0.25% sodium hypochlorite twice-weekly oral rinse. A pilot study. *J. Periodontol Res.* 49:696–702. <https://doi.org/10.1111/jre.12151>
24. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
25. Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27:863–864. <https://doi.org/10.1093/bioinformatics/btr026>
26. Schmieder R, Edwards R (2011) Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS One* 6:e17288. <https://doi.org/10.1371/journal.pone.0017288>
27. Bushnell B, Rood J, Singer E (2017) BBMerge—accurate paired shotgun read merging via overlap. *PLoS One* 12:e0185056. <https://doi.org/10.1371/journal.pone.0185056>
28. Wood DE, Salzberg SL (2014) Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 15:R46. <https://doi.org/10.1186/gb-2014-15-3-r46>
29. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* 39:W339–W346. <https://doi.org/10.1093/nar/gkr466>
30. Olsen JE, Hoegh-Andersen KH, Casadesus J, Rosenkranz J, Chadfield MS, Thomsen LE (2013) The role of flagella and chemotaxis genes in host pathogen interaction of the host adapted *Salmonella enterica* serovar Dublin compared to the broad host range serovar *S. Typhimurium*. *BMC Microbiol.* 13:67. <https://doi.org/10.1186/1471-2180-13-67>
31. Fang H, Edlund C, Hedberg M, Nord CE (2002) New findings in beta-lactam and metronidazole resistant *Bacteroides fragilis* group. *Int. J. Antimicrob. Agents* 19:361–370
32. Murray DA, Wilton JM (2003) Lipopolysaccharide from the periodontal pathogen *Porphyromonas gingivalis* prevents apoptosis of HL60-derived neutrophils in vitro. *Infect. Immun.* 71:7232–7235
33. Sun CC, Dong WR, Shao T, Li JY, Zhao J, Nie L, Xiang LX, Zhu G, Shao JZ (2017) Peroxiredoxin 1 (Prx1) is a dual-function enzyme by possessing Cys-independent catalase-like activity. *Biochem. J.* 474:1373–1394. <https://doi.org/10.1042/BCJ20160851>
34. McBride MJ, Xie G, Martens EC, Lapidus A, Henrissat B, Rhodes RG, Goltsman E, Wang W, Xu J, Hunnicutt DW, Staroscik AM, Hoover TR, Cheng YQ, Stein JL (2009) Novel features of the polysaccharide-digesting gliding bacterium *Flavobacterium johnsoniae* as revealed by genome sequence analysis. *Appl. Environ. Microbiol.* 75:6864–6875. <https://doi.org/10.1128/AEM.01495-09>
35. Lerouge I, Vanderleyden J (2002) O-antigen structural variation: mechanisms and possible roles in animal/plant-microbe interactions. *FEMS Microbiol. Rev.* 26:17–47