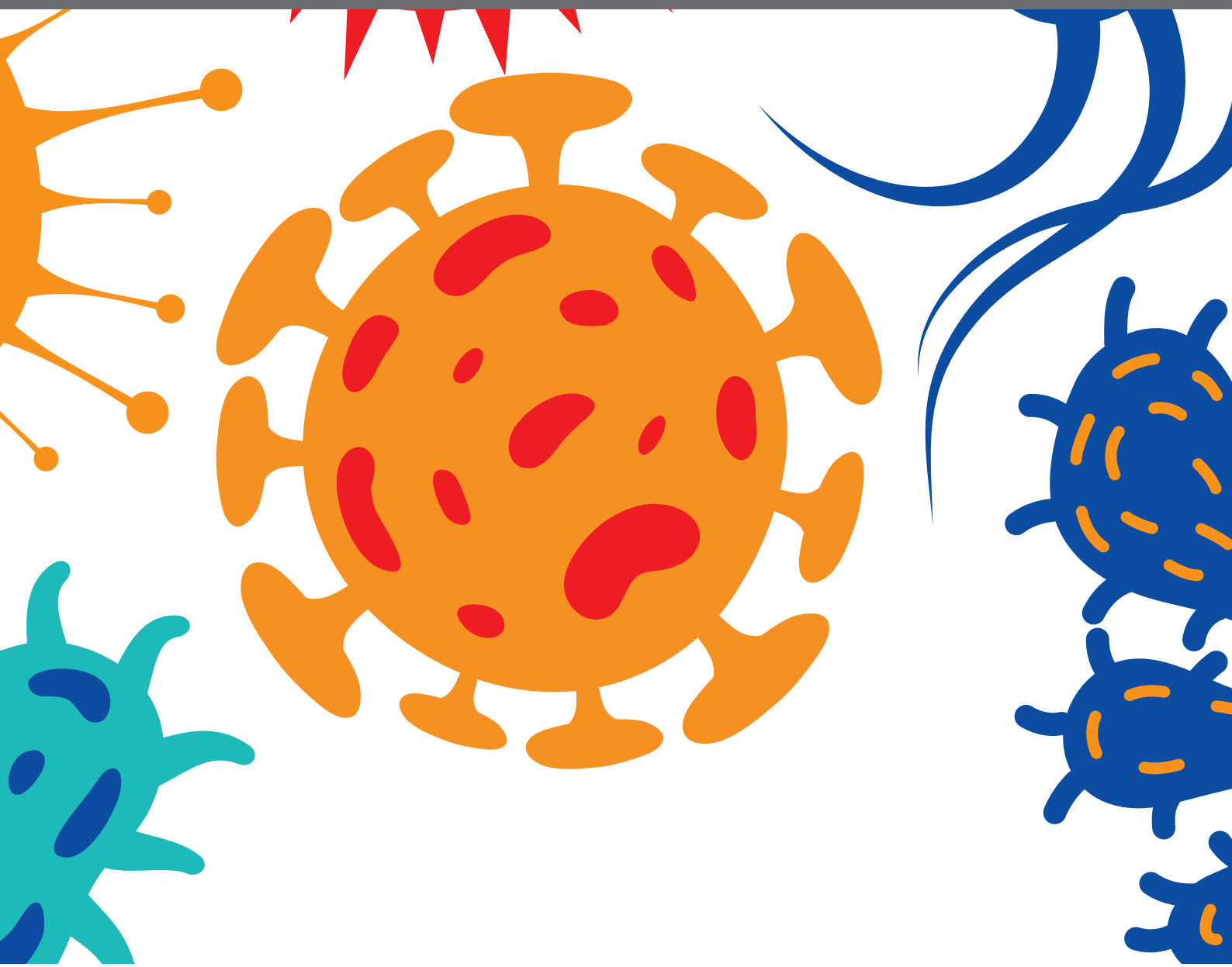
APPLICATIONS OF NEXT GENERATION SEQUENCING (NGS) TECHNOLOGIES TO DECIPHER THE ORAL MICROBIOME IN SYSTEMIC HEALTH AND DISEASE

EDITED BY: Thuy Do, Dongmei Deng and Naile Dame-Teixeira

PUBLISHED IN: Frontiers in Cellular and Infection Microbiology



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88974-308-7
DOI 10.3389/978-2-88974-308-7

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

APPLICATIONS OF NEXT GENERATION SEQUENCING (NGS) TECHNOLOGIES TO DECIPHER THE ORAL MICROBIOME IN SYSTEMIC HEALTH AND DISEASE

Topic Editors:
Thuy Do, University of Leeds, United Kingdom
Dongmei Deng, Academic Centre for Dentistry Amsterdam, VU Amsterdam, Netherlands
Naile Dame-Teixeira, University of Brasilia, Brazil

Citation: Do, T., Deng, D., Dame-Teixeira, N., eds. (2022). Applications of Next Generation Sequencing (NGS) Technologies to Decipher the Oral Microbiome in Systemic Health and Disease. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88974-308-7

Table of Contents

frontiers in Cellular and Infection Microbiology

EDITORIAL
published: 22 December 2021
doi: 10.3389/fcimb.2021.801122

Editorial: Applications of Next Generation Sequencing (NGS) Technologies to Decipher the Oral Microbiome in Systemic Health and Disease

Thuy Do[1]*, *Naile Dame-Teixeira*[1,2] *and Dongmei Deng*[3]

[1] Division of Oral Biology, School of Dentistry, University of Leeds, Leeds, United Kingdom, [2] Department of Dentistry, School of Health Sciences, University of Brasília, Brasilia, Brazil, [3] Department of Preventive Dentistry, Academic Centre for Dentistry Amsterdam (ACTA), University of Amsterdam and VU University Amsterdam, Amsterdam, Netherlands

OPEN ACCESS

Edited and reviewed by:
Andrew T. Gewirtz,
Georgia State University,
United States

***Correspondence:**
Thuy Do
n.t.do@leeds.ac.uk

Specialty section:
This article was submitted to Microbiome in Health and Disease, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: 24 October 2021
Accepted: 30 November 2021
Published: 22 December 2021

Citation:
Do T, Dame-Teixeira N and Deng D (2021) Editorial: Applications of Next Generation Sequencing (NGS) Technologies to Decipher the Oral Microbiome in Systemic Health and Disease. Front. Cell. Infect. Microbiol. 11:801122. doi: 10.3389/fcimb.2021.801122

Keywords: NGS, oral microbiome, biofilms, dysbiosis, health, systemic diseases

Editorial on the Research Topic

Applications of Next Generation Sequencing (NGS) Technologies to Decipher the Oral Microbiome in Systemic Health and Disease

Advances in next generation sequencing (NGS) technologies have revolutionised microbiology in the past few decades enabling researchers to elucidate the composition of the human microbiome. Recently, more effort is being directed at capturing both structural and functional aspects of the microbiome through gathering information on genes, transcripts, proteins and metabolites (Oresic et al., 2004; Muller et al., 2021). Combining several omics approaches with patient information is set to become the norm in future studies and will provide invaluable information to bring our understanding of symbiotic and dysbiotic processes to the next level. In this special issue "Applications of next generation sequencing (NGS) technologies to decipher the oral microbiome in systemic health and disease", we have compiled 17 papers including mostly original research articles, but also systematic/scoping reviews and a brief research report, that give focus to the oral microbiome and its impact on our systemic health.

The oral cavity harbours a rich and diverse microbial population, which varies greatly within and between individuals. However, it remains relatively stable throughout adulthood (Kilian et al., 2016), despite regular disruptions through diet, oral hygiene, and occasional medication including antibiotics and polypharmacy, if any.

Changes in systemic conditions and associated treatments may influence the oral microbiota, for example hypertension, hyperglycaemia and dysbiosis occurring at other body sites like the gut *via* the oral-gut axis. Several studies in this e-book provide such evidence. Significant alterations in the oral and gut microbial profiles were determined by Shi et al. and were correlated with the severity of

rheumatic heart disease. In the study by Kageyama et al., tongue cancer patients undergoing surgical resection were found to have a significantly different salivary microbiota. The mechanisms that underpin the oral-gut microbiome axis and other cross-talks with distant organs such as the lungs, liver, brain, skin and genital tract remain unclear and deserve full investigations (Zhao et al.) (Martinez et al., 2021; Park et al., 2021).

There is also growing evidence that the oral microbiota can influence systemic health (Jia et al., 2018). Specifically, the oral microbiota has been associated with a wide array of conditions such as dementia, depression, obesity, cancer, arthritis, diabetes, gut and skin diseases (Frid et al.) (Maitre et al., 2020; Sedghi et al., 2021; Wingfield et al., 2021). However, from the evidence presented, it is not clear whether changes in microbial profiles associated with oral dysbiosis are a manifestation of disease, or whether they drive the disease process in the oral cavity and elsewhere.

Deeper insights into how the microbiome can influence our wellbeing will help design tailored strategies for disease prevention and treatment. The paper by Zhou et al. is a good example of the use of oral microbiota data to help diagnose carcinomas through a non-invasive and low-cost approach. The acquisition of increasingly large NGS data coupled with biological and medical information is paving the way to the design of more accurate predictive models of health and disease which will enable a comprehensive and personalised medicine approach in our near future. Xie et al. described the use of elastic net models on supra-gingival microbiome data to accurately predict the presence of metabolites. More work is needed to make sense of the complex, multi-dimensional interactions occurring between the host and the microbiome. Computational methods linking microbes with metabolites (either produced or degraded) will be invaluable to learn patterns specific to homeostasis and pathogenesis, with therapeutic implications. Moreover, the ease of access to the oral cavity makes health checks convenient. Attempts at manipulating and controlling the oral microbiota through modulation strategies to reduce local and systemic inflammation, microbial load and maintain oral homeostasis will hopefully lead to similar repercussions systemically and promote overall health.

High throughput sequencing has helped generate large, publicly available datasets of descriptive microbiome profiles which are increasingly combined with functional microbial analyses. Investigations using NGS data combined with proteomics as the one described in Bao et al. are harbinger of future conventional methodologies. Re-analyses and meta-analyses of large datasets are important work which advance the robustness of our analysis methods and our overall knowledge of microbes and their interactions (Cai et al.; de Cena et al.; Jiang et al.; Kang et al.). The inclusion of patients' clinical data (medical history, systemic health or disease parameters) would also be helpful in deciphering the mechanisms of actions through which the oral microbiome exerts its impact on human health. Most journals impose requirements for NGS studies to provide sequencing data *via* links to repositories. While dealing with manuscripts in this e-book, we have come across two important issues: 1) Incomplete information within presented datasets are common. Kang et al. stated that only 57% of publicly available datasets with accession numbers provided both sequencing and metadata. 2) Clinical data are not always specified in the provided metadata. Within the datasets that provided both sequencing and metadata, we noticed that many manuscripts did not specify patients' clinical data, for example diagnosis criteria, clinical parameters or medical history. This makes future rigorous re-analyses and meta-analyses difficult for the scientific community, if possible, at all. As we are entering a personalized medicine era, it is crucial that all available data are optimally utilized for comparative analyses of larger and varied patient populations (Garcia et al., 2013; Berg et al., 2020).

Furthermore, taking in account clinical parameters as potential confounding factors for health or disease will be useful in preventing bias when examining omics data. For example, when comparing oral microbial communities in individuals with and without caries, other oral and systemic conditions, e.g. hyposalivation, periodontitis, candidiasis (Lyu et al.) and any sign of inflammation should be taken into consideration (Dame-Teixeira et al., 2021). The study by Bostanci et al. clearly showed how oral dysbiosis fluctuates during the menstrual cycle and detailed the impact of smoking and dietary sugar as risk factors. Inflammatory and immunological impairments also shape the oral microbiome. Lettieri et al. correlated the salivary microbiome with a significant inflammophylic profile in the oral phenotype of syndromic individuals with ineffective cathepsin C and impairment of neutrophils, where traditional periodontal therapy is not efficient.

The field of microbiome research has grown rapidly and has covered many disciplines in the past decade. There is still a lack of consensus in the terminology used to describe some of the methods and microbial communities (Marchesi and Ravel, 2015). We noticed that while most of the papers in our collection used the term '16S rRNA gene sequencing', some refer to the incorrect term '16S rDNA sequencing'. Hence, we recommend the standardization of the vocabulary used in microbiome studies (Berg et al., 2020).

In our e-book, we have collected papers on the current trends in oral microbiome research. The studies described integration of meta-omics data with complex biological context of health and disease. Future collaborative efforts are expected to carry out more accurate, rigorous computational methods using increasingly larger, and detailed datasets, which will enable us to predict treatment outcomes and make personalised medicine achievable and accessible.

AUTHOR CONTRIBUTIONS

TD, ND-T and DD contributed to the conception of the work, the drafting, writing and critical revision of the article.

FUNDING

This work was funded by the UK's Academy of Medical Sciences Newton International Fellowship (NIF\R5\242).

REFERENCES

Berg, G., Rybakova, D., Fischer, D., Cernava, T., Verges, M. C., Charles, T., et al. (2020). Microbiome Definition Re-Visited: Old Concepts and New Challenges. *Microbiome* 8, 103. doi: 10.1186/s40168-020-00875-0

Dame-Teixeira, N., de Lima, A. K. A., Do, T., and Stefani, C. M. (2021). Meta-Analysis Using NGS Data: The Veillonella Species in Dental Caries. *Front. Oral. Health* 2. doi: 10.3389/froh.2021.770917

Garcia, I., Kuska, R., and Somerman, M. J. (2013). Expanding the Foundation for Personalized Medicine: Implications and Challenges for Dentistry. *J. Dent. Res.* 92, 3S–10S. doi: 10.1177/0022034513487209

Jia, G., Zhi, A., Lai, P. F. H., Wang, G., Xia, Y., Xiong, Z., et al. (2018). The Oral Microbiota - a Mechanistic Role for Systemic Diseases. *Br. Dent. J.* 224, 447–455. doi: 10.1038/sj.bdj.2018.217

Kilian, M., Chapple, I. L., Hannig, M., Marsh, P. D., Meuric, V., Pedersen, A. M. L., et al. (2016). The Oral Microbiome - An Update for Oral Healthcare Professionals. *Br. Dent. J.* 221, 657–666. doi: 10.1038/sj.bdj.2016.865

Maitre, Y., Micheneau, P., Delpierre, A., Mahalli, R., Guerin, M., Amador, G., et al. (2020). Did the Brain and Oral Microbiota Talk to Each Other? A Review of the Literature. *J. Clin. Med.* 9 (12), 3876. doi: 10.3390/jcm9123876

Marchesi, J. R., and Ravel, J. (2015). The Vocabulary of Microbiome Research: A Proposal. *Microbiome* 3, 31. doi: 10.1186/s40168-015-0094-5

Martinez, J. E., Vargas, A., Perez-Sanchez, T., Encio, I. J., Cabello-Olmo, M., and Barajas, M. (2021). Human Microbiota Network: Unveiling Potential Crosstalk Between the Different Microbiota Ecosystems and Their Role in Health and Disease. *Nutrients* 13 (9), 2905. doi: 10.3390/nu13092905

Muller, E., Algavi, Y. M., and Borenstein, E. (2021). A Meta-Analysis Study of the Robustness and Universality of Gut Microbiome-Metabolome Associations. *Microbiome* 9, 203. doi: 10.1186/s40168-021-01149-z

Oresic, M., Clish, C. B., Davidov, E. J., Verheij, E., Vogels, J., Havekes, L. M., et al. (2004). Phenotype Characterisation Using Integrated Gene Transcript, Protein and Metabolite Profiling. *Appl. Bioinformatics* 3, 205–217. doi: 10.2165/00822942-200403040-00002

Park, S. Y., Hwang, B. O., Lim, M., Ok, S. H., Lee, S. K., Chun, K-S., et al. (2021). Oral-Gut Microbiome Axis in Gastrointestinal Disease and Cancer. *Cancers (Basel)* 13 (9), 2124. doi: 10.3390/cancers13092124

Sedghi, L., DiMassa, V., Harrington, A., Lynch, S. V., and Kapila, Y. L. (2021). The Oral Microbiome: Role of Key Organisms and Complex Networks in Oral Health and Disease. *Periodontol. 2000* 87, 107–131. doi: 10.1111/prd.12393

Wingfield, B., Lapsley, C., McDowell, A., Miliotis, G., McLafferty, M., O'Neill, S. M., et al. (2021). Variations in the Oral Microbiome Are Associated With Depression in Young Adults. *Sci. Rep.* 11, 15009. doi: 10.1038/s41598-021-94498-6

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cellular and Infection Microbiology

ORIGINAL RESEARCH
published: 02 October 2020
doi: 10.3389/fcimb.2020.588155

Proteome and Microbiome Mapping of Human Gingival Tissue in Health and Disease

Kai Bao[1], Xiaofei Li[2], Lucy Poveda[3], Weihong Qi[3], Nathalie Selevsek[4], Pinar Gumus[5], Gulnur Emingil[5], Jonas Grossmann[3], Patricia I. Diaz[6], George Hajishengallis[2], Nagihan Bostanci[1] and Georgios N. Belibasakis[1]**

[1] Division of Oral Diseases, Department of Dental Medicine, Karolinska Insitutet, Huddinge, Sweden, [2] Department of Basic and Translational Sciences, School of Dental Medicine, Philadelphia, PA, United States, [3] Functional Genomic Centre, ETH Zurich and University of Zurich, Zürich, Switzerland, [4] Swiss Integrative Center for Human Health, Fribourg, Switzerland, [5] Department of Periodontology, School of Dentistry, Ege University, Izmir, Turkey, [6] Department of Oral Biology, University at Buffalo, State University of New York, Buffalo, NY, United States

OPEN ACCESS

Edited by:
Thuy Do,
University of Leeds, United Kingdom

Reviewed by:
Jeffrey Ebersole,
University of Nevada, Las Vegas, United States
Daniel Champlin Propheter, University of Texas Southwestern Medical Center, United States

***Correspondence:**
Kai Bao
kai.bao@ki.se
Georgios N. Belibasakis
george.belibasakis@ki.se

Specialty section:
This article was submitted to Microbiome in Health and Disease, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: *28 July 2020*
Accepted: *02 September 2020*
Published: *02 October 2020*

Citation:
Bao K, Li X, Poveda L, Qi W, Selevsek N, Gumus P, Emingil G, Grossmann J, Diaz PI, Hajishengallis G, Bostanci N and Belibasakis GN (2020) Proteome and Microbiome Mapping of Human Gingival Tissue in Health and Disease. Front. Cell. Infect. Microbiol. 10:588155. doi: 10.3389/fcimb.2020.588155

Efforts to map gingival tissue proteomes and microbiomes have been hampered by lack of sufficient tissue extraction methods. The pressure cycling technology (PCT) is an emerging platform for reproducible tissue homogenisation and improved sequence retrieval coverage. Therefore, we employed PCT to characterise the proteome and microbiome profiles in healthy and diseased gingival tissue. Healthy and diseased contralateral gingival tissue samples (total $n = 10$) were collected from five systemically healthy individuals (51.6 ± 4.3 years) with generalised chronic periodontitis. The tissues were then lysed and digested using a Barocycler, proteins were prepared and submitted for mass spectrometric analysis and microbiome DNA for 16S rRNA profiling analysis. Overall, 1,366 human proteins were quantified (false discovery rate 0.22%), of which 69 proteins were differentially expressed (≥2 peptides and $p < 0.05$, 62 up, 7 down) in periodontally diseased sites, compared to healthy sites. These were primarily extracellular or vesicle-associated proteins, with functions in molecular transport. On the microbiome level, 362 species-level operational taxonomic units were identified. Of those, 14 predominant species accounted for >80% of the total relative abundance, whereas 11 proved to be significantly different between healthy and diseased sites. Among them, *Treponema* sp. HMT253 and *Fusobacterium naviforme* and were associated with disease sites and strongly interacted ($r > 0.7$) with 30 and 6 up-regulated proteins, respectively. Healthy-site associated strains *Streptococcus vestibularis*, *Veillonella dispar, Selenomonas* sp. HMT478 and *Leptotrichia* sp. HMT417 showed strong negative interactions ($r < -0.7$) with 31, 21, 9, and 18 up-regulated proteins, respectively. In contrast the down-regulated proteins did not show strong interactions with the regulated bacteria. The present study identified the proteomic and intra-tissue microbiome profile of human gingiva by employing a PCT-assisted workflow. This is the first report demonstrating the feasibility to analyse full proteome profiles of gingival tissues in both healthy and disease sites, while deciphering the tissue site-specific microbiome signatures.

Keywords: tissue proteomic analysis, biofilm, gingiva, periodontitis, microbiome

INTRODUCTION

As a consequence of the microbial challenge, periodontitis causes destruction of underline connecting tissue, including gingival epithelial layer, which builds a barrier to the external challenge. Thus, microbial invasion of the periodontal tissues may take place during the respective pathological processes (Colombo et al., 2007). In earlier transmission electron microscopy studies, invasion of spirochetes and other microorganisms were evident in the gingival epithelium and connective tissues, especially in patients with acute necrotising gingivitis (Listgarten, 1965; Courtois et al., 1983). This tissue invasive feature is different from that of endocytosis by non-phagocytic host cells, through which bacteria evade phagocytic elimination by the immune system. However, bacterial invasion has traditionally been considered to take place at relatively late stages. Based on this rationale, most periodontal microbiome studies have been focused on the characterisation of biofilms. It is plausible that, at least in part, bacterial invasion is involved in the pathogenesis of periodontitis. Interestingly, one study showed that bacteria could form a biofilm-like structure within the gingival tissue (Baek et al., 2018). Further, 16s rRNA profiling analysis has shown that *Fusobacterium nucleatum* and *Porphyromonas gingivalis* were highly enriched within the tissue compared with the plaque (Baek et al., 2018). Although most available microbiome studies were mainly focused on the bacterial plaque (biofilm), an overall microbiome map directly derived from human gingival tissues is necessary to draw the whole picture for understanding this virulence aspect of periodontal disease.

With the help of mass spectrometry, researchers can identify thousands of proteins for a given sample in a single run (Bostanci and Bao, 2017), which is ideal for delivering a snapshot of protein regulations within the gingival tissue. However, attempts to map gingival tissue proteomes have been hampered by lacking sufficient protein extraction workflows. Bertoldi et al. (2013) reported 13 gingival proteins differentially regulated in diseased sites, compared with their neighboring inter-proximal healthy sites. This included the upregulation of annexin A2, actin cytoplasmic 1, carbonic anhydrase 1 and 2; Ig kappa chain C region and flavinreductase as well as downregulation of 4-3-3 protein sigma and zeta/delta, heat-shock protein beta-1, triosephosphateisomerase, peroxiredoxin-1, fatty acid-binding protein-epidermal, and galectin-7 in pathological tissues. Monari et al. identified 32 different protein spots and elevation of S100A9, 14-3-3 protein zeta/delta, Heat shock protein beta-1 and Galectin-7 in gingival tissues from periodontal patients compared with those from healthy individuals (Monari et al., 2015). Whereas, Yaprak et al. (2018) identified 47 proteins from healthy gingival tissue, including 14-3-3 protein sigma, S100A9 and Galectin-7, which also identified in the works of Bertoldi et al. and Monari et al. Yet, although transcriptomic and proteomic patterns are rarely similar (Wang et al., 2017), transcriptomic analysis of gingival tissues has identified as many as 12,744 expressing genes (Demmer et al., 2008), indicating that there is plenty of space for the improvement for proteomic identification.

A sensitive pressure cycling technology (PCT)-assistant workflow with proven efficiency in gingival tissue disruption (Bao et al., 2019) was used in this study. The present study aimed to concomitantly characterise the gingival tissue proteome and microbiome of systematically healthy individuals with periodontitis, by comparing healthy and diseased sites. Label-free quantitative proteomics and 16SrRNA gene sequencing platforms were applied to dissect the relationship between bacterial abundance and protein regulation among these gingival tissues.

RESULTS

Proteome Profiles of Gingival Tissue Samples Cluster Based on Clinical State

The gingival tissue proteome charted in this study derived from 10 gingival tissues (one healthy and one diseased site per individual) obtained from 5 individuals with stage III periodontitis. Prevalence of teeth with one or more sites with probing pocket depth (PPD) $>$ 5 mm and PPD $>$ 5 mm were % 44 $\pm$ 5.3 and % 35.6 $\pm$ 6.3. Approximately 82% of sites with PPD $>$ 5 mm had bleeding on probing (BOP). The mean PPD and clinical attachment loss (CAL) scores of the sampled diseased sites were significantly higher than the healthy ones [$p < 0.05$, PPD (mm): 2.2 $\pm$ 0.8 vs. 7.0 $\pm$ 0.7, $p < 0.001$, CAL (mm): O vs. 8.0 $\pm$ 0.7, $p < 0.0001$].

Following a PCT-assisted label-free quantification work-flow, we obtained an overview of the gingiva proteome of 1,369 proteins (including 2 contaminant and 3 decoy proteins), with a protein false discovery rate (FDR) of 0.22%. Each quantified protein consisted of at least two unique peptides identified and quantified (**Appendix Table 1**). Although unsupervised hierarchical clustering analysis of the tissue proteomes could not distinguish healthy from diseased sites based on their normalised abundances (**Figure 1A**), this became possible by the utilisation of sPLS-DA (**Figure 1B**). Considering that protein regulation among individuals may vary, we assessed the differentially expressed protein levels by comparing intra-individually healthy and diseased sites, using paired *t*-test. Of all quantified proteins, 62 qualified as higher [log2 (FC) $\geq$ 0, $P \leq 0.05$], whereas only 7 qualified as lower [log2 (FC) $\leq$ 0, $P \leq 0.05$] in diseased sites compared to the respective healthy sites (**Figure 1C**, **Appendix Table 2**).

Gene Ontology (GO) Analysis of the Regulated Proteins in the Gingival Tissue

The GO functions of differentially expressed proteins were annotated using the METACORE online software (https://portal.genego.com, Thomson Reuters). The top enriched GO terms for localisations, processes and molecular functions of all 69 regulated proteins were recognised and ranked according to their statistical significance (**Table 1**, **Appendix Table 3**). The major cellular localisation of the regulated proteins was "extracellular"-or "vesicle"-related (**Table 1A**). For instance, "extracellular exosome," "extracellular vesicle," and "extracellular organelle" were the top three enriched terms (**Table 1A**). Many of these proteins were linked to "localisation" or "transport" processes (e.g., "cellular localisation," intracellular

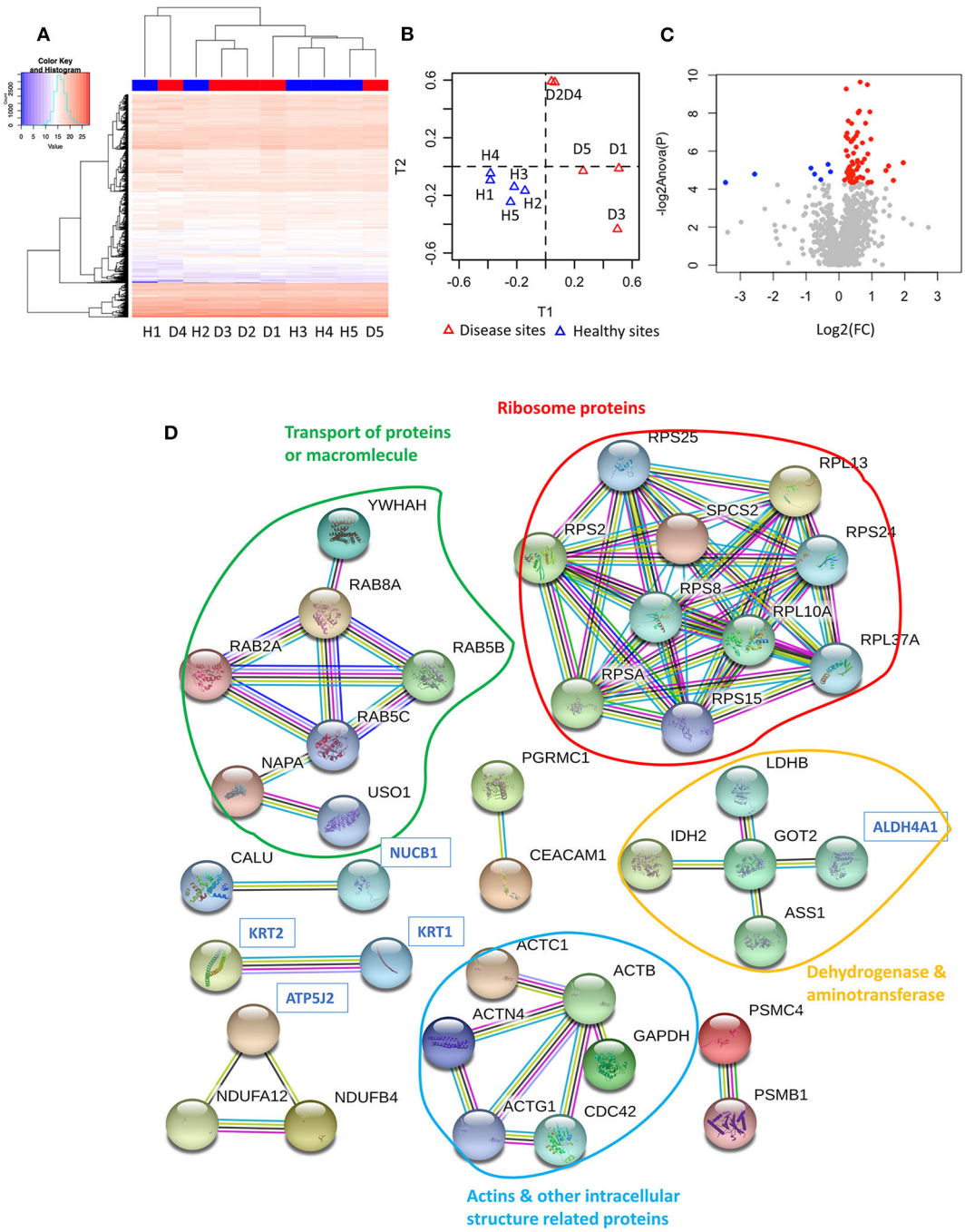


FIGURE 1 | Proteome of gingival tissues from healthy and diseased sites. **(A)** The heatmap of normalised abundance for identified and quantified proteins in gingival tissue. Samples isolated from healthy sites were highlighted in blue colour, while samples isolated from diseased sites were highlighted in red. **(B)** The sPLS-DA plots represented the normalised abundance of all the 1,366 proteins from healthy (blue triangles) or diseased sites (red triangles). **(C)** 62 proteins were upregulated [log2 (FC) $\geq$ 0, $P \leq 0.05$] in the disease compared with healthy sites (red dots), and 7 were downregulated [log2 (FC) $\leq$ 0, $P \leq 0.05$]. **(D)** String analysis for the interaction between regulated proteins. Network established using STRING with interaction confident scores more than 0.9. The upregulated proteins were labelled in black, while the downregulated proteins were labelled in blue. The methods for acquiring the different protein-protein interactions were illustrated by different lines. The interaction confirmed by the curated database and experimental results were shown in blue and purple line, respectively. The interaction predicated by gene neighborhood, gene fusions, and gene co-occurrence were shown in green, red, and dark blue lines, respectively. While interaction was determined by text-mining, co-expression and protein homology were shown in yellow, black, and light blue lines, respectively. Healthy sites: H. Diseased sites: D.

TABLE 1 | Enriched GO terms of regulated proteins.

A: Top 10 enriched GO localisation of regulated proteins

#	GO terms (Localisations)	Regulated	*P*-value
1	Extracellular exosome	47/2,932	2.8778E-25
2	Extracellular vesicle	47/2,951	3.807E-25
3	Extracellular organelle	47/2,963	4.5382E-25
4	Extracellular space	49/4,428	1.7958E-19
5	Extracellular region part	50/4,693	2.8805E-19
6	Vesicle	50/4,975	3.6834E-18
7	Extracellular region	52/5,860	1.0373E-16
8	Intracellular organelle part	62/10,709	7.1559E-12
9	Organelle part	63/11,087	8.3233E-12
10	Cytoplasmic part	64/11,535	1.2786E-11

B: Top 10 enriched GO processes of regulated proteins

#	GO terms (Processes)	Regulated	*P*-value
1	Establishment of localisation in the cell	40/2,418	5.532E-19
2	Cellular component organisation or biogenesis	63/7,803	2.3678E-16
3	Cellular localisation	41/3,189	1.3431E-15
4	Intracellular protein transport	27/1,234	2.5142E-15
5	Cellular protein localisation	33/2,036	4.6355E-15
6	Cellular component biogenesis	44/3,862	5.2894E-15
7	Cellular macromolecule localisation	33/2,048	5.4923E-15
8	Intracellular transport	33/2,068	7.2675E-15
9	Supramolecular fiber organisation	20/636	2.7295E-14
10	Establishment of protein localisation	32/2,036	3.3676E-14

C: Top 10 enriched GO molecular functions of regulated proteins

#	GO terms (Molecular functions)	Regulated	*P*-value
1	Structural constituent of ribosome	9/202	1.4028E-08
2	Structural molecule activity	17/1,044	1.6659E-08
3	Cytoskeletal protein binding	16/1,100	2.2308E-07
4	Cadherin binding	9/335	1.0296E-06
5	Protein binding	62/14,119	1.3072E-06
6	RNA binding	19/1,919	4.6905E-06
7	Heterocyclic compound binding	40/6,982	6.5837E-06
8	Cell adhesion molecule binding	10/558	9.2596E-06
9	Organic cyclic compound binding	40/7,084	9.7525E-06
10	Actin binding	9/483	1.9857E-05

protein transport) (**Table 1B**), whereas their top molecular functions belonged to the "structural constituent of ribosome" and "structural molecule activity" (**Table 1C**).

To further understand their inter-relationships, the protein-protein interactions were analysed using STRING (https://string-db.org/) (**Appendix Table 4**). When applying the highest confidence score (0.9), 40 among the 69 regulated proteins, formed 73 pairs of known such interactions, as illustrated by string networks (**Figure 1D**). The largest cluster of protein interactions consisted of 10 different proteins, which were mainly ribosomal ones. The second-largest cluster identified consisted of 8 proteins with assigned macromolecular transport properties (e.g., Ras-related proteins, general vesicular transport factor p115). In addition, among the interactions illustrated in the network, some were identified between actin and other intracellular-structural proteins, or between five dehydrogenases and aminotransferases, as well as more sparse interactions of only two or three proteins.

Microbiome Profiles of Gingival Tissue Samples Cluster Based on Clinical State

Our next approach was to examine microorganisms present in the gingival tissues by establishing a microbial catalogue from 450,668 sequences that binned within 97% sequence identity from all ten-tissue samples (**Appendix Table 5**). On average, more than 119609.8 reads were identified from each sample, with a standard deviation of 77658.61. To analyse the alpha diversity, rarefaction curves were plotted based on the observed OTUs (**Figure 2A**), with calculated coverages for disease and health of 99.32 and 99.50%, respectively (**Figure 2B**), while the inverse Simpson diversity for disease and health were 7.18 and 5.05, respectively (**Figure 2C**). Furthermore, 362 non-rare OTUs were discovered among all ten-samples (data are available via ENA), with no significant differences in the number of detected OTUs between disease and health ($P = 0.277$, 77, and 74 average OTUs for diseased and healthy sites, respectively) using paired *t*-test (**Figure 2D**). To visualise differences in community structure between the groups, an NMDS plot of the thetayc distance was generated (**Figure 2E**), yielding sample clustering based on sites (healthy tissue vs. diseased tissue), not by sample pairing. Such sample clustering indicated the presence of different OTUs between diseased and healthy tissues. In addition, unsupervised hierarchical clustering analysis of OTU abundance also pointed to global microbiome differences between the two types of gingival tissues, where samples were also clustered according to tissue type (**Figure 2F**).

Only 14 species comprised more than 80% of 16S rRNA gene reads (**Figure 2G**). In healthy sites, *Streptococcus vestibularis* was the most abundant species, followed by *Haemophilus parahaemolyticus* and *Veillonella dispar* (**Appendix Table 6**). For diseased sites, the most abundant species was an as-yet-uncultured species *Treponema* sp. Human Microbial Taxon (HMT) 253, whereas the abundance of *S. vestibularis* declined from more than 20% in health sites, to <10% in diseased sites (**Figure 2H**). Eleven OTUs were significantly different ($P < 0.05$) between healthy and diseased sites, including five from the 14 most abundant species (i.e., *S. vestibularis, Treponema* sp. HMT 253*, V. dispar, Fusobacterium naviforme,* and *Selenomonas* sp. HMT 478) (**Figure 2G**).

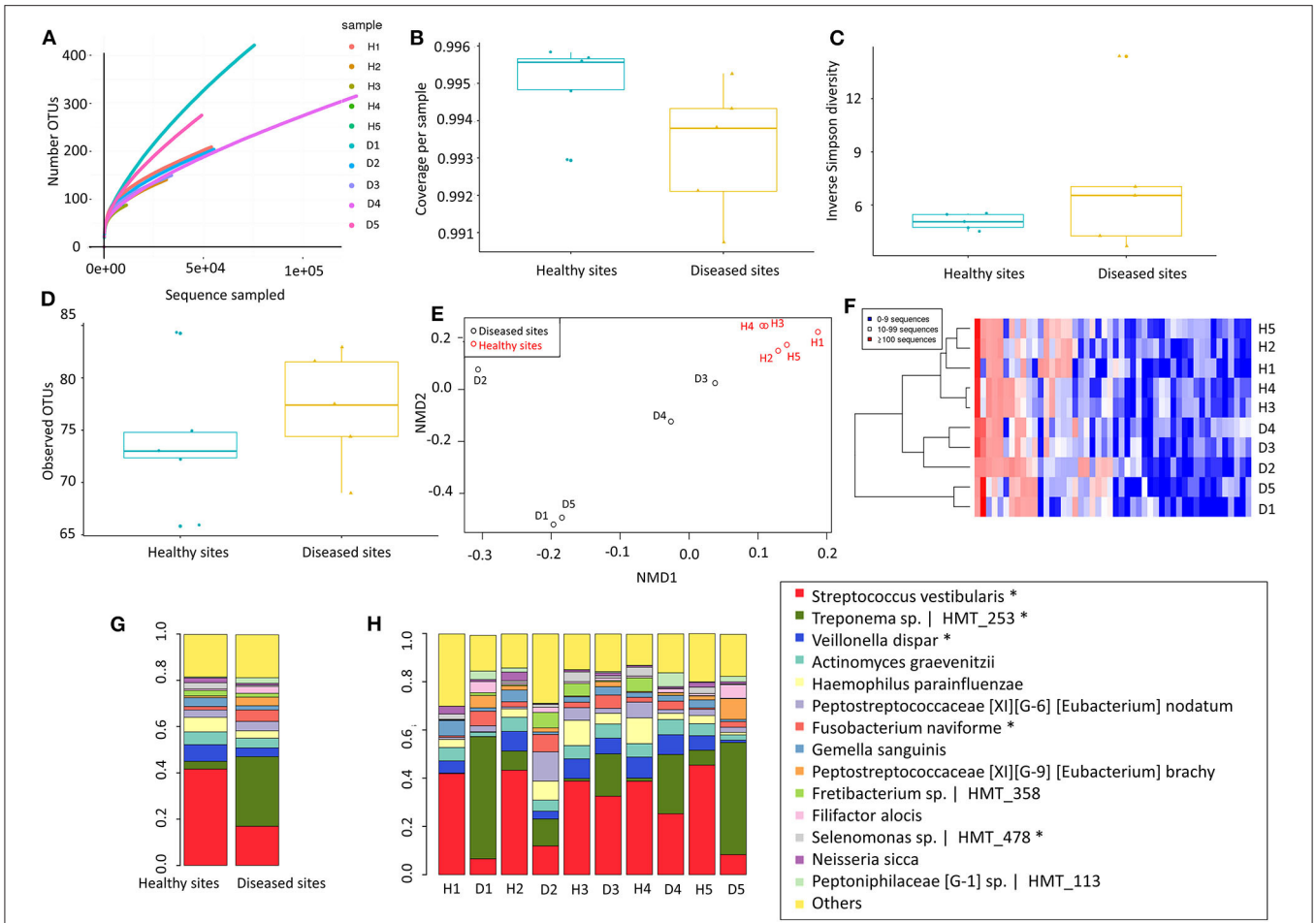


FIGURE 2 | Microbiome of gingival tissues from healthy and diseased sites. **(A)** Rarefaction curve for the number of OTUs as a function of sampling effort. **(B)** Sample coverage. **(C)** Inverse Simpson diversity. **(D)** Observed OTU per group. **(E)** Non-metric multidimensional scaling (NMDS) for distance matrix of between disease and healthy samples. **(F)** OTU abundance across samples_ Abundant OTUs only (after removing rare OTUs). The bacterial relative abundance of the gingival tissue in **(G)** all healthy and disease tissues, or in **(H)** individual gingival tissues. The abundance that significant shift ($P < 0.05$) between two sites were highlighted in "*". Healthy sites: H. Diseased sites: D.

Gingival Tissue Interactome: Correlations Between Proteomes and Microbiomes of Gingival Tissues

To adequately address the interactome of the gingival tissues, potential correlations between the 69 regulated proteins and 11 regulated species were further analysed using the mixOmics package (**Figure 3**, **Appendix Table 7**). We found *S. vestibularis*, *V. dispar*, *Leptotrichia* sp. HMT_417, and *Selenomonas* sp. HMT_478 were clustered with all 7 downregulated proteins (i.e., Nucleobindin-1, Delta-1-pyrroline-5-carboxylate dehydrogenase, ATP synthase subunit f, Aldehyde dehydrogenase class 2, Keratin, type II cytoskeletal 1, Vacuolar protein sorting-associated protein 29, and Cytokeratin-2e), and negatively correlated with the 62 upregulated proteins (**Figure 3A**). On the contrary, five regulated species (more abundant in diseased sites), namely *Treponema* sp. HMT_253, *Streprococcus salivarius, Peptostreptococcaceae [XI][G-6] [Eubacterium] nodatum*, *Variovorax paradoxus* and *F. naviforme*, were strongly associated ($r > 0.7$) with 28, 29, 20, 1, and 4 upregulated proteins, respectively (**Figure 3B**, **Appendix Table 7**).

DISCUSSION

Different proteome and microbiome studies have been performed to understand periodontal diseases, yet few have focused on gingival tissues. Earlier studies indicated that the microbial content of the periodontal pocket in non-human primates or human determines the gene expression patterns in the gingival tissues (Papapanou et al., 2009; Ebersole et al., 2020). In the present study, we successfully applied a contemporary PCT-assisted workflow to dissect the gingival proteome and microbiome of both diseased and healthy sites from patients with periodontitis. The tissue recipient sites included both maxilla and mandible, which may exhibit different degrees of keratinisation. It should also be acknowledged that the stringent requirement

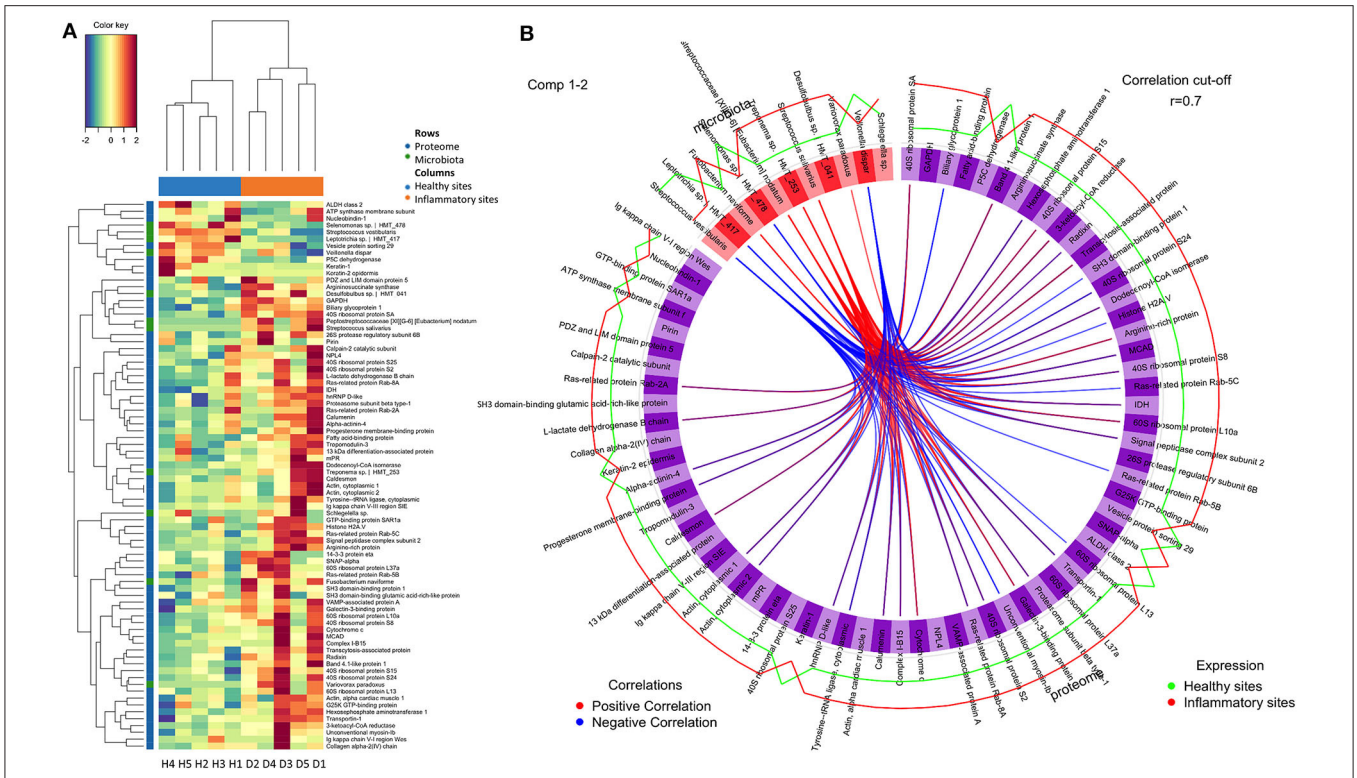


FIGURE 3 | Correlations between regulated proteins and microbiotas. **(A)** The colour-coded clustered image map that present correlation between regulated proteins and microbiotas following pair-wise variable associations for canonical correlation analysis. **(B)** The Circos Plot showed variable correlations among regulated proteins and microbiotas (**Appendix Table 7**). Positive correlations were indicated with red lines, while negative correlations were indicated with blue lines. Only correlations more than 0.7 were showed. The levels of expression in healthy sites were indicated in green line, while the levels of expression in inflammatory sites were indicated in red. Healthy sites: H. Diseased sites: D.

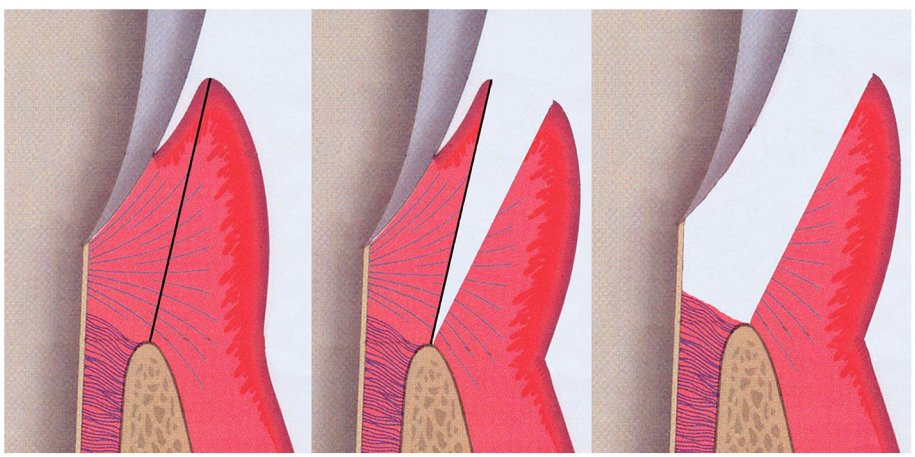


FIGURE 4 | Illustration of the sample sits for gingival tissue. Both pocket epithelium and underlying connective tissue were included in the gingival specimen taken from the approximal sites of the selected teeth (grey; tooth, pink; gingival connective tissue, purple-red; gingival epithelium, brown-beige; alveolar bone).

of obtaining both a healthy and a diseased tissue specimen from the same donor prohibited us from limiting further the sampling criteria to either jaw, else it would have been more cumbersome to identify suitable patients.

The study also presents the most comprehensive quantitative proteome map of human gingival tissue to date, by quantifying 1,366 proteins, while earlier proteome analyses quantified <50 proteins at a time. Interestingly, only 69 of over 1,300 quantified

proteins were significantly differentially expressed, whereas many of the inflammation-related proteins, such as Ig gamma-1 chain C region and Protein S100-A9, were abundantly expressed at both sites. The fact that the global proteomic profiles were similar between healthy and diseased sites denotes that protein composition of clinically healthy gingival tissue in the periodontal patient may have already been altered, even in the absence of evident clinical signs of inflammation at those sites. Of note, we used a data-dependent acquisition (DDA) strategy in this study. DDA only samples a subset of the most abundant ions detected during the first MS scan for the further fragmentation and sequential MS scans, while discards the rest. Hence, many proteins with demonstrated roles in periodontitis, including various cytokines, might have been masked due to their lower abundance. The high biological variabilities among patients may also contribute to low number of differentially expressed proteins. Nevertheless, this observation on overall protein regulations is consistent with transcriptomic changes observed in gingival tissue in periodontitis (Kebschull and Papapanou, 2011). Furthermore, the prediction of the protein functions shows that most of the regulated proteins were localised in the extracellular space. Similarly, increased numbers of secreted proteins were previously identified in human experimental gingivitis (Bostanci et al., 2013) and murine ligature-induced periodontitis (Bao et al., 2019). Hence, it is not surprising to find that many of the regulated proteins had transport-related functions. We also observed increases in the ribosome-related proteins (e.g., 60S ribosomal protein L37a, 40S ribosomal protein S25, 40S ribosomal protein S8, etc.), indicating ribosomal biosynthetic activity in the inflamed sites (Zhou et al., 2015).

Under homeostatic conditions, the host is in a balanced relationship with commensal oral species or potential pathobionts (Hajishengallis and Lamont, 2014). Hence, the concomitant study of the proteome and the microbiome is high relevant. One the microbiome aspect, 14 species accounted for 80% of the abundances, only 11 significantly differentiated between health and disease, six of which belonging to the rarely abundant group. This denotes that both high and low abundance species are to be considered in the future for defining signatures of target organisms capable of distinguishing between clinical health and disease. The oral microbiome data obtained in this study from gingival tissues identified potentially invasive species of the periodontium. Nevertheless, it cannot be definitively confirmed that the detected bacteria were all actual tissue invaders, and not superficial persisters after the washing steps. Yet, only a limited portion of the tissue interface has been in direct contact with the biofilm, which is dispersed during the homogenisation process of the specimen, thus down-playing the representation of the non-invaded species. Although their precise effects on the tissue may currently be unclear, different species were earlier found to co-exist within the gingival tissue (Baek et al., 2018). The relative abundances of different taxa may denote their potential roles in health and disease, or their invasive capacity. *Treponema* sp. HMT 253 and *F. naviforme* were significantly increased in diseased compared to healthy sites. For *Treponema s*p. HMT253, the only information currently available on this as-yet-uncultured species is its 16S rRNA gene clone library, derived from dental plaque of subjects with periodontitis and acute necrotising ulcerative gingivitis (Dewhirst et al., 2000). It is closely related to *Treponema denticola*, a potential pathogen implicated in periodontal disease (Dewhirst et al., 2010). Different models have shown that *T. denticola* is able to invade the epithelium and basement membrane (Grenier et al., 1990; Lux et al., 2001; Chi et al., 2003), as well as to secrete a chymotrypsin-like protease that can digest host components including type IV collagen, laminin and fibronectin (Grenier et al., 1990). *T. denticola* dentilisin was also reported to disrupt the epithelial cell monolayer (Chi et al., 2003). *Fusobacterium naviforme* (formerly F. nucleatum ssp. naviforme) has been identified and isolated from subgingival plaque samples (Colombo et al., 2009). Based on phylogenetic analysis (Dewhirst et al., 2010), it is expected to display functional similarities *F. nucleatum*. Although *Treponema* sp. HMT 253 and *F. naviforme* were not usually found as an abundant constituent of the subgingival biofilm in patients with periodontitis, their abundance within the tissue documented in this study suggests a greater invasion potential than their cultivated and characterised relatives. Of further note, other species that were elevated in diseased compared to healthy tissue, but did not reach statistical significance, are worth mentioning. Such were *F. alocis* and *Fretibacterium* sp. HMT 358, suggesting that they are potentially invasive of the gingival tissues. Both *F. alocis* and *Fretibacterium* sp. have been increasingly associated with periodontal disease. *Fretibacterium* sp. belongs to the phylum *Synergistetes* and is shown to be increased in the saliva of patients with periodontitis (Belibasakis et al., 2013) and in dental biofilms of patients with ANUG (Baumgartner et al., 2012), an invasive form of periodontal disease.

Some potentially invasive but generally less pathogenic species, including *S. vestibularis*, *V. dispar*, and *Selenomonas* sp. HMT 478, were enriched in healthy sites. Even though originally identified in the oral cavity (Whiley and Hardie, 1988), the presence of *S. vestibularis* has not been reported in periodontitis, but in other infectious diseases (Duan et al., 2017; Yilmaz et al., 2017). *V. dispar* is found in subgingival plaque from chronic periodontitis patients (Moon et al., 2015) and plays an import role when saliva is the main nutritional source of oral biofilm (Kolenbrander, 2011). The presence of *Selenomonas* spp. are reported in the salivary (Duan et al., 2017) or dental plaque microbiome (Paster et al., 2001; Faveri et al., 2008) of periodontal patients, but at a lower prevalence compared with other putative pathogens (Goncalves et al., 2012). Previous studies have shown that *Streptococcus* (Teles et al., 2012), *Veillonella* (Kolenbrander, 2011) and *Selenomonas* (Goncalves et al., 2012) contribute to the structural organisation of oral biofilm. *Streptococcus* spp. have the potential to colonise or invade the gingival tissue, but with no known association to gingival inflammation, which is well in line with our findings. It should be noted that, although diseased sites showed higher abundances of *Treponema* sp. HTM 253, most healthy sites fostered a fairly high proportion of this species. Perhaps this species allows other less invasive microorganisms like *Veillonella* spp. and *Streptococcus* spp. to penetrate the tissue barrier.

Thus far there has been inconsistent evidence to support or exclude the invasive properties of oral species in the pathogenesis of periodontal disease (Mendes et al., 2015). From

an epidemiological perspective, there is sufficient evidence on the role of specific species as etiological agents of periodontitis, but the disease may be better understood as dysbiotic inflammation resulting from the concerted interaction of correlations (Lopez et al., 2015; Hajishengallis and Lamont, 2016). The interactome analysis of the present study indicates that groups of significantly elevated species and proteins tend to correlate with one another in health or disease. Information derived from such studies may decipher biological signatures in periodontal disease, which will help us understand its etiopathogenesis on the tissue level and may confer future diagnostic and prognostic value (Belibasakis and Mylonakis, 2015).

MATERIALS AND METHODS

Study Population and Design

Gingival tissues ($n = 10$) were collected from two sites (one healthy and one diseased) of each five systematically healthy individuals with stage III periodontitis (age range from 45 to 56 years with a mean age 51.6 ± 4.5 years, F:M: 2:3). The study was approved by the Ethics Committee of Ege University (number 17–11.1/34) and conducted following the guidelines of the World Medical Association Declaration of Helsinki. Patients first attended the Department of Oral Diagnosis and Radiology for the completion of clinical and radiological examination procedures and were then directed to the specialised within the University Dental Clinics, for further assessment and treatment. Exclusion criteria were the use of tobacco products, presence of cardiovascular and respiratory diseases, diabetes mellitus, HIV infection, systemic inflammatory conditions or non-plaque-induced oral inflammatory conditions, immunosuppressive chemotherapy, and current pregnancy or lactation. None of the patients had a history of periodontal therapy or had taken medication such as antibiotics or anti-inflammatory drugs that could affect their periodontal status for at least 6 months prior to the study. Informed consent was obtained from all participants. Full-mouth and site-specific periodontal parameters including PPD, CAL, dichotomous presence of BOP, and plaque for each patient were recorded. The full-mouth means PPD (mm) and CAL (mm) were 5.1 ± 0.4, 5.7 ± 0.4, respectively. The full-mouth mean plaque and BOP scores were 72.0 ± 6.7 and 75.0 ± 5.0, respectively.

Collection of Gingival Tissue Samples

Gingival tissue samples, including both pocket epithelium and underlying connective tissue, were taken from the approximal sites of the selected teeth prior to non-surgical periodontal therapy (**Figure 4**). The "healthy sites" had no clinical signs of gingival inflammation (no BOP), exhibited a PPD of ≤ 3 mm and had no radiographic evidence of alveolar bone loss and no CAL. These healthy tissues were sampled when the premolars were scheduled to have periodontal crown lengthening surgery. The "diseased sites" showed BOP, had an interproximal PPD of ≥6 mm, and a concomitant CAL of ≥6 mm. Two gingival tissue samples from each participant were obtained and washed with sterile normal saline solution to remove any blood or detached biofilms on the tissue surface. Tissues were then placed in a sterile tube containing a tissue protectant solution (RNAlater, Sigma-Aldrich) and stored at +4°C overnight, before long-term storage at −70°C until later usage.

Protein Extraction and Digestion

Gingival tissues were washed three times, each for 5 min in PBS to remove any residues prior to lysis. The tissues were then lysed and digested using a Barocycler NEP2320 (Pressure BioSciences) at 33°C as described previously (Bao et al., 2019). In brief, 2.5 to 3 mg of samples ($n = 10$) were placed in MicroTubes (Pressure BioSciences) and lysed with a 60–cycle barocyling process. The exacted proteins were then reduced and alkylated using tris (2-carboxyethyl) phosphine (Sigma) and iodoacetamide (Sigma). Later, extracted proteins were digested using Lys-C (Wako) at an enzyme-to-protein (estimated to be 10% of the wet sample weight) ratio of a 1:45 with a 45–cycle barocyling process. These resultant solutions were further diluted and then digested again using trypsin (Promega) at an enzyme-to-protein ratio of 1:50 with a 90-cycle process. Each barocyling-cycle mentioned above consisted of a 50 s ultra-high pressure phase (45, 20, 20 thousand pounds per square inch (KPSI) for 60-, 45-, 90-cycle process, respectively) followed with a 10 s ambient pressure phase in each cycle. Resultant solutions were acidified by trifluoroacetic acid (TFA) (Sigma) to a final concentration of 0.8% w/v, desalted using reverse-phase cartridges Finisterre SPE C18 (Wicom International AG), dried with vacuum centrifuged, and kept in −20°C freezer until further use.

Mass Spectrometry (MS) and Data Analysis

All frozen peptides were reconstituted in 3% acetonitrile ACN in 0.1% formic acid and adjusted to 0.5 μg/μl using NanoDrop 1000 spectrophotometer (WITEC AG). One microlitre of desalted peptide was analysed on an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific) for proteomic analysis as described previously (Bao et al., 2017).

Label-free quantification was performed by Progenesis QI for proteomics (Non-linear Dynamics) as described previously (Bao et al., 2017). In brief, .raw files of individual samples were aligned with a pooled sample to create a Mascot files (.mgf). This .mgf files was searched with Mascot (version 2.4.1, Matrix Science) using the following search parameters: precursor tolerance: ± 10 ppm; fragment ion tolerance: ± 0.6 Da; enzyme: trypsin; maximum missed cleavages: 2; fixed medication: carbamidomethylation of cysteine; variable modification: deamidated (NQ), oxidation (M) and acetylation on protein N-termini. Data were searched against a FASTA file (40,510 sequences and 22,667,481 residues), consisting of the human proteome from UniProt (isoforms included, retrieved December 9th 2016), contaminant database from the FGCZ, the resulting.dat file was imported into Scaffold v4.0 (Proteome Software) to generate spectrum report, with protein false discovery rate (FDR) of 10%, minimal one peptide and peptide FDR of 5%. Finally, the spectrum report was imported back into Progenesis QI for identifying the quantified proteins. The "within-subject" option was used for experiment design set up. Proteins with a minimum of two unique quantified peptides and

a significant ANOVA p-value smaller than 0.05 were considered as differentially regulated ones.

Data Clustering and Heat Maps for Regulated Proteins

Unsupervised clustering analysis and heat maps of regulated proteins were generated using the R software (R: A Language and Environment for Statistical Computing, R Development Core Team) in particular the Quantable packages (https://cran.r-project.org/web/packages/quantable/index.html) to obtain a global visualisation and regulation trends of protein profiles. Apparent outliers were excluded from the quantification. Sparse Partial Least Squares Discriminant Analysis (sPLS-DA) was used to visualise the similarity between healthy and inflammation sites using the SPLS package (Chun and Keles, 2010).

Functional Analysis of the Regulated Proteins

The regulated proteins were subjected to Metacore online database (29th May 2019) for "gene ontology (GO) enrichment analysis." Enriched GO terms were recognised and ranked according to their statistical significance (–log2P value), using a hypergeometric distribution.

Sample Disruption and DNA Extraction

Other gingival tissues (approximal 2.5 to 3 mg, $n = 10$) were lysed using a Barocycler NEP2320 (Pressure BioSciences) with only a 60–cycle barocyling process (each consisting 50 s at 45 KPSI followed by 10 s at ambient pressure) at 33°C for DNA extraction. The genomic DNA from lysates was extracted using the GenElute™ Bacterial Genomic DNA Kit (Sigma) and stored at −20°C until further use.

Sample Preparation for 16S rRNA Gene Sequencing

The hypervariable regions 7 to 9 (V7-9) of the 16S rRNA gene were amplified in the first round of PCR from isolated genomic DNA using universal bacterial primers that also contained the Illumina Truseq primer binding site (**Appendix Table 8**). Amplification reactions were performed in a total volume of 25 μl containing 5X KAPA HiFi Buffer, 10 mM KAPA dNTP Mix, 0.5 U KAPA HiFi DNA Polymerase (KAPA Biosystems), 4 μM of primers ordered from Microsynth (Balgach), and 22.4 ng DNA diluted in DNA-free water. The PCR amplification was performed on a Verity thermocycler (Thermo Fisher Scientific) with the following cycling conditions: 95°C for 5 min, 25 cycles at 98°C each for 20 s, 70°C for 30 s, 72°C for 30 s and a final extension at 72°C for 10 min. The PCR reactions were run on a 2% agarose gel, the amplicon band was cut and extracted using MinElute Gel Extraction Kit (Qiagen) and eluted in 50 μl DNase-free water.

In the second round of PCR, the remaining Illumina Truseq adaptors together with dual indexing Truseq barcodes were incorporated into the previously amplified material (**Appendix Table 9**). Amplification reactions were performed in a total volume of 50 μl containing 5X KAPA HiFi Buffer, 10 mM KAPA dNTP Mix, 1 U KAPA HiFi DNA Polymerase (KAPA Biosystems), 4 μM of the primers ordered from Microsynth (Balgach) and 22.5 ng of the previously amplified material diluted in DNA-free water. PCR amplification was performed on a Verity thermocycler (Thermo Fisher Scientific) with the following cycling conditions: 98°C for 5 min, five cycles at 98°C each for 30 s, 54°C for 30 s, 72°C for 30 s and a final extension at 72°C for 5 min. PCR products were gel-purified and eluted in 50 μl DNase-free water. The quality and quantity of resulting amplicon libraries were validated using Qubit® (1.0, Invitrogen) Fluorometer and the Tapestation (Agilent). The amplicons from the different samples were normalised to 4 nM in Tris-Cl 10 mM, pH8.5 with 0.1% Tween 20 and as they contain dual indexes, they were equimolarly pooled and paired-end sequenced in a Miseq Illumina Instrument (Illumina CA) using a 600cycle V3 kit.

Processing and Taxonomic Classification of 16S rRNA Gene Reads

MiSeq paired-end (PE) reads were first filtered based on average quality (>= Q20) using Trimmomatic (version 0.36) (Bolger et al., 2014). Quality checked PE reads were processed using mothur (version 1.38.1) (Schloss et al., 2009), following the MiSeq SOP (Standard Operation Protocol) (https://www.mothur.org/wiki/MiSeq_SOP). In detail, the quality-filtered PE reads were joined into contig sequences. Identical sequences were merged and the counts of all unique sequences were recorded. Unique sequences were aligned guided by the Silva bacterial 16S reference alignment (Release 102) (Quast et al., 2013). After alignment, the bulk of the sequences started at position 34,476 and ended at position 43,116 of the reference alignment. Sequences aligned at the different start and/or stop sites, as well as sequences with homopolymers longer than 8 nt were filtered out. Sites containing only gap characters were also removed. Sequences were pre-clustered allowing for up to three base differences. Chimaera sequences were removed using the UCHIME algorithm (Edgar et al., 2011). Sequences were initially classified by comparing them to the mothur-formatted RDP training set (v.9), with cutoff values set at genus level (Cole et al., 2014). This taxonomic information was used to remove undesired contaminants (Chloroplast, Mitochondria, unknown Archaea and Eukaryota) and to split the sequences into 16S genus bins (taxlevel=6) and one un-classified bin. Each of the 166 bins was then clustered into operational taxonomic units (OTUs) using the single-linkage clustering algorithm implemented in hpc-clust (Matias Rodrigues and von Mering, 2014), with 97% sequence similarity as the cutoff. The mothur-compatible OTU list was prepared using the utility script "makeotus_mothur" in the same software and imported into mothur for OTU-based analysis. To taxonomically classify OTUs, representative sequences (the most abundant) were compared against the Human Oral Microbiome Database (HOMD) (Dewhirst et al., 2010) using BLASTN in ncbi-blast-2.6.0+ (Altschul et al., 1990). The taxonomy of the best match (with >96.99% homology) was assigned to the corresponding OTU. If the representative sequence had <97% homology to the HOMD reference, the genus name was used to taxonomically designate the OTU.

Microbiome Data Analysis

In all analysis where normalisation was applied, standardised datasets were generated by randomly selecting 3,597 sequences 1,000 times from each sample. To analyse the alpha diversity of the samples, rarefaction curves describing the number of OTUs observed as a function of sampling effort were plotted. The numbers of sequences, the sample coverage, the number of observed OTUs, and the Inverse Simpson diversity were calculated. To compare the membership and structure of the samples between groups, distance matrices for the classical Jaccard (1908) and Yue and Clayton theta values (Yue and Clayton, 2005) were calculated. The distance matrices were also visualised using NMDS (non-metric multidimensional scaling) (Clarke, 1993). The R software, in particular, the packages "gplots" and "stats," were used to generate unsupervised clustering analysis and heatmaps of non-rare OTUs (abundance $\geq$ 50 across all samples). Differentially represented OTUs were evaluated via paired Student's *t*-test using their relative abundances $p < 0.05$. Benjamini-Hochberg corrected *P*-values and power calculations were provided for each OTUs.

Correlations Between Microbiome and Protein Datasets

To visualise the correlation between differentially presented OTUs and proteins, Circos plots and cluster-imagine maps were generated for r values (Pair-wise variable associations for canonical correlation analysis correlation between variables, defined by a generalisation of the cosine angle between the center of the circle and each variable point; Gonzalez et al., 2012) using the mixOmics package in R (Rohart et al., 2017).

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found here: https://www.ebi.ac.uk/ena, PRJEB27125 (ERP109167); http://www.proteomexchange.org/, PXD018029.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of Ege University (number 17-11.1/34). The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

KB contributed to data acquisition, analysis, interpretation, drafted, and critically revised the manuscript. XL, LP, and WQ contributed to data analysis and interpretation as well as critically revised the manuscript. NS, JG, PD, and GH critically revised the manuscript. PG and GE contributed to conception and design, data acquisition, and critically revised the manuscript. NB and GB contributed to conception and design, data interpretation, drafted, and critically revised the manuscript. All authors gave final approval and agreed to be accountable for all aspects of the work.

FUNDING

This work was supported by the strategic funds from the Karolinska Institutet, research grants from Swedish Research Council (VR:2017-01198) and DE028561 from the National Institutes of Health/National Institute of Dental and Craniofacial Research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2020.588155/full#supplementary-material

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Baek, K., Ji, S., and Choi, Y. (2018). Complex intratissue microbiota forms biofilms in periodontal lesions. *J. Dent. Res.* 97, 192–200. doi: 10.1177/0022034517732754

Bao, K., Bostanci, N., Thurnheer, T., and Belibasakis, G. N. (2017). Proteomic shifts in multi-species oral biofilms caused by *Anaeroglobus geminatus*. *Sci. Rep* 7:4409. doi: 10.1038/s41598-017-04594-9

Bao, K., Li, X., Kajikawa, T., Toshiharu, A., Selevsek, N., Grossmann, J., et al. (2019). Pressure cycling technology assisted mass spectrometric quantification of gingival tissue reveals proteome dynamics during the initiation and progression of inflammatory periodontal disease. *Proteomics* 2019:e1900253. doi: 10.1002/pmic.201900253

Baumgartner, A., Thurnheer, T., Luthi-Schaller, H., Gmur, R., and Belibasakis, G. N. (2012). The phylum synergistetes in gingivitis and necrotizing ulcerative gingivitis. *J. Med. Microbiol.* 61(Pt. 11), 1600–1609. doi: 10.1099/jmm.0.047456-0

Belibasakis, G. N., and Mylonakis, E. (2015). Oral infections: clinical and biological perspectives. *Virulence* 6, 173–176. doi: 10.1080/21505594.2015.1025191

Belibasakis, G. N., Ozturk, V. O., Emingil, G., and Bostanci, N. (2013). Synergistetes cluster A in saliva is associated with periodontitis. *J. Periodont. Res.* 48, 727–732. doi: 10.1111/jre.12061

Bertoldi, C., Bellei, E., Pellacani, C., Ferrari, D., Lucchi, A., Cuoghi, A., et al. (2013). Non-bacterial protein expression in periodontal pockets by proteome analysis. *J. Clin. Periodontol.* 40, 573–582. doi: 10.1111/jcpe.12050

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bostanci, N., and Bao, K. (2017). Contribution of proteomics to our understanding of periodontal inflammation. *Proteomics* 17:1500518. doi: 10.1002/pmic.201500518

Bostanci, N., Ramberg, P., Wahlander, A., Grossman, J., Jonsson, D., Barnes, V. M., et al. (2013). Label-free quantitative proteomics reveals differentially regulated proteins in experimental gingivitis. *J. Proteome Res.* 12, 657–678. doi: 10.1021/pr300761e

Chi, B., Qi, M., and Kuramitsu, H. K. (2003). Role of dentilisin in *Treponema denticola* epithelial cell layer penetration. *Res. Microbiol.* 154, 637–643. doi: 10.1016/j.resmic.2003.08.001

Chun, H., and Keles, S. (2010). Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *J. R. Stat. Soc. Series B Stat. Methodol.* 72, 3–25. doi: 10.1111/j.1467-9868.2009.00723.x

Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Austral J. Ecol.* 18, 117–143. doi: 10.1111/j.1442-9993.1993.tb00438.x

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., et al. (2014). Ribosomal database project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 42, D633–D642. doi: 10.1093/nar/gkt1244

Colombo, A. P., Boches, S. K., Cotton, S. L., Goodson, J. M., Kent, R., Haffajee, A. D., et al. (2009). Comparisons of subgingival microbial profiles of refractory periodontitis, severe periodontitis, and periodontal health using the human oral microbe identification microarray. *J. Periodontol.* 80, 1421–1432. doi: 10.1902/jop.2009.090185

Colombo, A. V., da Silva, C. M., Haffajee, A., and Colombo, A. P. (2007). Identification of intracellular oral species within human crevicular epithelial cells from subjects with chronic periodontitis by fluorescence *in situ* hybridization. *J. Periodont. Res.* 42, 236–243. doi: 10.1111/j.1600-0765.2006.00938.x

Courtois, G. J. III., Cobb, C. M., and Killoy, W. J. (1983). Acute necrotizing ulcerative gingivitis. A transmission electron microscope study. *J. Periodontol.* 54, 671–679. doi: 10.1902/jop.1983.54.11.671

Demmer, R. T., Behle, J. H., Wolf, D. L., Handfield, M., Kebschull, M., Celenti, R., et al. (2008). Transcriptomes in healthy and diseased gingival tissues. *J. Periodontol.* 79, 2112–2124. doi: 10.1902/jop.2008.080139

Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C., Yu, W. H., et al. (2010). The human oral microbiome. *J. Bacteriol.* 192, 5002–5017. doi: 10.1128/JB.00542-10

Dewhirst, F. E., Tamer, M. A., Ericson, R. E., Lau, C. N., Levanos, V. A., Boches, S. K., et al. (2000). The diversity of periodontal spirochetes by 16S rRNA analysis. *Oral Microbiol. Immunol.* 15, 196–202. doi: 10.1034/j.1399-302x.2000.150308.x

Duan, X., Wu, T., Xu, X., Chen, D., Mo, A., Lei, Y., et al. (2017). Smoking may lead to marginal bone loss around non-submerged implants during bone healing by altering salivary microbiome: a prospective study. *J. Periodontol.* 88, 1297–1308. doi: 10.1902/jop.2017.160808

Ebersole, J., Kirakodu, S., Chen, J., Nagarajan, R., and Gonzalez, O. A. (2020). Oral microbiome and gingival transcriptome profiles of ligature-induced periodontitis. *J. Dent. Res.* 99, 746–757. doi: 10.1177/0022034520906138

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381

Faveri, M., Mayer, M. P., Feres, M., de Figueiredo, L. C., Dewhirst, F. E., and Paster, B. J. (2008). Microbiological diversity of generalized aggressive periodontitis by 16S rRNA clonal analysis. *Oral Microbiol. Immunol.* 23, 112–118. doi: 10.1111/j.1399-302X.2007.00397.x

Goncalves, L. F., Fermiano, D., Feres, M., Figueiredo, L. C., Teles, F. R., Mayer, M. P., et al. (2012). Levels of *Selenomonas* species in generalized aggressive periodontitis. *J. Periodont. Res.* 47, 711–718. doi: 10.1111/j.1600-0765.2012.01485.x

Gonzalez, I., Cao, K. A., Davis, M. J., and Dejean, S. (2012). Visualising associations between paired 'omics' data sets. *BioData Min* 5:19. doi: 10.1186/1756-0381-5-19

Grenier, D., Uitto, V. J., and McBride, B. C. (1990). Cellular location of a *Treponema denticola* chymotrypsinlike protease and importance of the protease in migration through the basement membrane. *Infect. Immun.* 58, 347–351. doi: 10.1128/IAI.58.2.347-351.1990

Hajishengallis, G., and Lamont, R. J. (2014). Breaking bad: manipulation of the host response by *Porphyromonas gingivalis*. *Eur. J. Immunol.* 44, 328–338. doi: 10.1002/eji.201344202

Hajishengallis, G., and Lamont, R. J. (2016). Dancing with the stars: how choreographed bacterial interactions dictate nososymbiocity and give rise to keystone pathogens, accessory pathogens, and pathobionts. *Trends Microbiol.* 24, 477–489. doi: 10.1016/j.tim.2016.02.010

Jaccard, P. (1908). Nouvelles recherches Sur La distribution florale. *Bull. Soc. Vaudoise Sci. Nat.* 44, 223–270.

Kebschull, M., and Papapanou, P. N. (2011). Periodontal microbial complexes associated with specific cell and tissue responses. *J. Clin. Periodontol.* 38(Suppl. 11), 17–27. doi: 10.1111/j.1600-051X.2010.01668.x

Kolenbrander, P. E. (2011). Multispecies communities: interspecies interactions influence growth on saliva as sole nutritional source. *Int. J. Oral Sci.* 3, 49–54. doi: 10.4248/IJOS11025

Listgarten, M. A. (1965). Electron microscopic observations on the bacterial flora of acute necrotizing ulcerative gingivitis. *J. Periodontol.* 36, 328–339. doi: 10.1902/jop.1965.36.4.328

Lopez, R., Hujoel, P., and Belibasakis, G. N. (2015). On putative periodontal pathogens: an epidemiological perspective. *Virulence* 6, 249–257. doi: 10.1080/21505594.2015.1014266

Lux, R., Miller, J. N., Park, N. H., and Shi, W. (2001). Motility and chemotaxis in tissue penetration of oral epithelial cell layers by Treponema denticola. *Infect. Immun.* 69, 6276–6283. doi: 10.1128/IAI.69.10.6276-6283.2001

Matias Rodrigues, J. F., and von Mering, C. (2014). HPC-CLUST: distributed hierarchical clustering for large sets of nucleotide sequences. *Bioinformatics* 30, 287–288. doi: 10.1093/bioinformatics/btt657

Mendes, L., Azevedo, N. F., Felino, A., and Pinto, M. G. (2015). Relationship between invasion of the periodontium by periodontal pathogens and periodontal disease: a systematic review. *Virulence* 6, 208–215. doi: 10.4161/21505594.2014.984566

Monari, E., Cuoghi, A., Bellei, E., Bergamini, S., Lucchi, A., Tomasi, A., et al. (2015). Analysis of protein expression in periodontal pocket tissue: a preliminary study. *Proteome Sci*. 13:33. doi: 10.1186/s12953-015-0089-y

Moon, J. H., Lee, J. H., and Lee, J. Y. (2015). Subgingival microbiome in smokers and non-smokers in Korean chronic periodontitis patients. *Mol. Oral Microbiol.* 30, 227–241. doi: 10.1111/omi.12086

Papapanou, P. N., Behle, J. H., Kebschull, M., Celenti, R., Wolf, D. L., Handfield, M., et al. (2009). Subgingival bacterial colonization profiles correlate with gingival tissue gene expression. *BMC Microbiol* 9:221. doi: 10.1186/1471-2180-9-221

Paster, B. J., Boches, S. K., Galvin, J. L., Ericson, R. E., Lau, C. N., Levanos, V. A., et al. (2001). Bacterial diversity in human subgingival plaque. *J. Bacteriol.* 183, 3770–3783. doi: 10.1128/JB.183.12.3770-3783.2001

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

Rohart, F., Gautier, B., Singh, A., and Le Cao, K. A. (2017). mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* 13:e1005752. doi: 10.1371/journal.pcbi.1005752

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09

Teles, F. R., Teles, R. P., Uzel, N. G., Song, X. Q., Torresyap, G., Socransky, S. S., et al. (2012). Early microbial succession in redeveloping dental biofilms in periodontal health and disease. *J. Periodont. Res.* 47, 95–104. doi: 10.1111/j.1600-0765.2011.01409.x

Wang, J., Ma, Z., Carr, S. A., Mertins, P., Zhang, H., Zhang, Z., et al. (2017). Proteome profiling outperforms transcriptome profiling for coexpression based gene function prediction. *Mol. Cell. Proteomics* 16, 121–134. doi: 10.1074/mcp.M116.060301

Whiley, R. A., and Hardie, J. M. (1988). *Streptococcus vestibularis* sp. nov. from the human oral cavity. *Int. J. Syst. Evol. Microbiol.* 38:5. doi: 10.1099/00207713-38-4-335

Yaprak, E., Kasap, M., Akpinar, G., Kayaalti-Yuksek, S., Sinanoglu, A., Guzel, N., et al. (2018). The prominent proteins expressed in healthy gingiva: a pilot exploratory tissue proteomics study. *Odontology* 106, 19–28. doi: 10.1007/s10266-017-0302-9

Yilmaz, F., Bora, F., and Ersoy, F. (2017). "*Streptococcus vestibularis*": a rare cause of peritoneal dialysis-related peritonitis. *Ther. Apher. Dial.* 21, 418–419. doi: 10.1111/1744-9987.12538

Yue, J. C., and Clayton, M. K. (2005). A similarity measure based on species proportions. *Commun. Stat. Theor. Methods* 34, 2123–2131. doi: 10.1080/STA-200066418

Zhou, X., Liao, W. J., Liao, J. M., Liao, P., and Lu, H. (2015). Ribosomal proteins: functions beyond the ribosome. *J. Mol. Cell Biol.* 7, 92–104. doi: 10.1093/jmcb/mjv014

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cellular and Infection Microbiology

ORIGINAL RESEARCH
published: 04 November 2020
doi: 10.3389/fcimb.2020.602239

Salivary Oral Microbiome of Children With Juvenile Idiopathic Arthritis: A Norwegian Cross-Sectional Study

Paula Frid[1,2,3], Divyashri Baraniya[4], Josefine Halbig[2,5], Veronika Rypdal[3,6], Nils Thomas Songstad[6], Annika Rosèn[7,8], Johanna Rykke Berstad[9], Berit Flatø[10,11], Fadhl Alakwaa[12], Elisabeth Grut Gil[7], Lena Cetrelli[13], Tsute Chen[14], Nezar Noor Al-Hebshi[4], Ellen Nordal[3,6] and Mohammed Al-Haroni[5]* on behalf of The NorJIA (Norwegian JIA Study — Imaging, oral health and quality of life in children with juvenile idiopathic arthritis).*

[1] Department of ENT, Division of Oral and Maxillofacial Surgery, University Hospital North Norway, Tromsø, Norway, [2] Public Dental Service Competence Centre of North Norway, Tromsø, Norway, [3] Department of Clinical Medicine, UiT the Arctic University of Norway, Tromsø, Norway, [4] Oral Microbiome Laboratory, Kornberg School of Dentistry, Temple University, Philadelphia, PA, United States, [5] Department of Clinical Dentistry, UiT the Arctic University of Norway, Tromsø, Norway, [6] Department of Pediatrics and Adolescence Medicine, University Hospital of North Norway, Tromsø, Norway, [7] Department of Clinical Dentistry, University of Bergen, Bergen, Norway, [8] Department of Oral and Maxillofacial Surgery, Haukeland University Hospital, Bergen, Norway, [9] Department of ENT, Division of Oral and Maxillofacial Surgery, Oslo University Hospital, Oslo, Norway, [10] Department of Rheumatology and Infectious Diseases, Institute of Clinical Medicine, University of Oslo, Oslo, Norway, [11] Department of Rheumatology, Oslo University Hospital, Oslo, Norway, [12] Department of Computational Medicine and Bioinformatics, University Michigan, Ann Arbor, MI, United States, [13] Center of Oral Health Services and Research (TkMidt), Trondheim, Norway, [14] Department of Microbiology, Forsyth Institute, Cambridge, MA, United States

OPEN ACCESS

Edited by:
Dongmei Deng,
VU University Amsterdam,
Netherlands

Reviewed by:
Andrea Santarelli,
Marche Polytechnic University, Italy
Yi Liu,
Sichuan University, China

***Correspondence:**
Mohammed Al-Haroni
Mohammed.Al-Haroni@uit.no
Nezar Noor Al-Hebsh
alhebshi@temple.edu

Specialty section:
This article was submitted to Microbiome in Health and Disease, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: 02 September 2020
Accepted: 09 October 2020
Published: 04 November 2020

Citation:
Frid P, Baraniya D, Halbig J, Rypdal V, Songstad NT, Rosèn A, Berstad JR, Flatø B, Alakwaa F, Gil EG, Cetrelli L, Chen T, Al-Hebshi NN, Nordal E and Al-Haroni M (2020) Salivary Oral Microbiome of Children With Juvenile Idiopathic Arthritis: A Norwegian Cross-Sectional Study. Front. Cell. Infect. Microbiol. 10:602239. doi: 10.3389/fcimb.2020.602239

Background: The oral microbiota has been connected to the pathogenesis of rheumatoid arthritis through activation of mucosal immunity. The objective of this study was to characterize the salivary oral microbiome associated with juvenile idiopathic arthritis (JIA), and correlate it with the disease activity including gingival inflammation.

Methods: Fifty-nine patients with JIA (mean age, 12.6 ± 2.7 years) and 34 healthy controls (HC; mean age 12.3 ± 3.0 years) were consecutively recruited in this Norwegian cross-sectional study. Information about demographics, disease activity, medication history, frequency of tooth brushing and a modified version of the gingival bleeding index (GBI) and the simplified oral hygiene index (OHI-S) was obtained. Microbiome profiling of saliva samples was performed by sequencing of the V1-V3 region of the 16S rRNA gene, coupled with a species-level taxonomy assignment algorithm; QIIME, LEfSe and R-package for Spearman correlation matrix were used for downstream analysis.

Results: There were no significant differences between JIA and HC in alpha- and beta-diversity. However, differential abundance analysis revealed several taxa to be associated with JIA: *TM7-G1*, *Solobacterium* and *Mogibacterium* at the genus level; and *Leptotrichia* oral taxon 417, *TM7-G1* oral taxon 352 and *Capnocytophaga* oral taxon 864 among others, at the species level. *Haemophilus* species, *Leptotrichia* oral taxon 223, and *Bacillus subtilis*, were associated with healthy controls. *Gemella morbillorum*, *Leptotrichia* sp. oral taxon 498 and *Alloprevotella* oral taxon 914 correlated positively with the

composite juvenile arthritis 10-joint disease activity score (JADAS10), while *Campylobacter* oral taxon 44 among others, correlated with the number of active joints. Of all microbial markers identified, only *Bacillus subtilis* and *Campylobacter* oral taxon 44 maintained false discovery rate (FDR) < 0.1.

Conclusions: In this exploratory study of salivary oral microbiome we found similar alpha- and beta-diversity among children with JIA and healthy. Several taxa associated with chronic inflammation were found to be associated with JIA and disease activity, which warrants further investigation.

Keywords: juvenile idiopathic arthritis, salivary microbiome, next generation sequencing (NGS), oral health, 16S rRNA

INTRODUCTION

Juvenile idiopathic arthritis (JIA) is the most common chronic rheumatic disease in children, with an annual incidence of 1 to 2 per 1000 children (Moe and Rygg, 1998; Berntson et al., 2003). The pathogenesis of JIA remains unknown, although environmental triggers of disease in genetically predisposed individuals have been suggested (Cho and Blaser, 2012). Infectious agents are suspected to be environmental triggers for JIA and one of several possible mechanisms is molecular mimicry between bacterial molecules and self-antigens. It is well-known that host-microbe interaction is important for recognition and development of the immune system (Rossi et al., 2013). The human microbiome includes the collective genomes of the microbiota, which is the term for all microbes in the body (Cho and Blaser, 2012). A dysbiotic imbalance in the host microbiota might contribute as a potential trigger in the development of immune-mediated diseases, including JIA (Scher and Abramson, 2011). The gut microbiota in children with JIA is reported to differ from healthy individuals (Stoll et al., 2014; Di Paola et al., 2016; Stoll et al., 2016; Tejesvi et al., 2016; Aggarwal et al., 2017; Stoll et al., 2018; De Filippo et al., 2019; Dong et al., 2019; Van Dijkhuizen et al., 2019), where a lower abundance of Firmicutes and a higher abundance of Bacteroidetes were found in the gut microbiota of patients with oligoarticular and polyarticular rheumatoid factor (RF) negative new-onset JIA (Tejesvi et al., 2016). Higher abundance of Bacteroides is also seen in enthesitis-related-arthritis (Aggarwal et al., 2017). In all studies on the gut microbiome in JIA, however, no single species has been identified and different studies show changes in different taxa (Stoll et al., 2014; Di Paola et al., 2016; Stoll et al., 2016; Tejesvi et al., 2016; Aggarwal et al., 2017; Stoll et al., 2018; Dong et al., 2019; Van Dijkhuizen et al., 2019). These studies suggest that dysbiosis in the microbiota with overabundance in pathogenic microbes, may result in dysregulation of the immune system through disruption of the integrity of mucosal barrier and altered interaction with gut immune cells (Majumder and Aggarwal, 2020). Aberrations in mucosal homeostasis may be associated with the increased bacterial urease activity, reportedly found in fecal samples of JIA patients as compared to healthy controls. The increase of urease activity is hypothesized to be the result of alterations in the anaerobic bacterial environment (Malin et al., 1996). It is still an open question whether the observed microbial dysbiosis is a cause or an effect of the disease.

The oral microbiome has been proposed to play a role in rheumatoid arthritis by contributing to systemic inflammation (Lorenzo et al., 2019), and may also play a similar role in JIA. Furthermore, dysbiosis and periodontitis have been found to be associated to increased severity of rheumatoid arthritis (Scher et al., 2012). Gingivitis, i.e. gingival bleeding, which is a reversible inflammation of the gingiva caused by dental biofilm accumulation is a prerequisite for progression to periodontitis (Murakami et al., 2018). Gingival inflammation is in some studies found to be higher in individuals with JIA compared to healthy (Welbury et al., 2003; Ahmed et al., 2004) It has been suggested that bacteria found in the saliva are representative for the oral bacteria associated with the different oral mucosal surfaces. So far, there have been no attempts to characterize the salivary oral microbiome associated with JIA with the next generation sequencing method (NGS). *The aim of this study was therefore to investigate the oral microbiome in saliva of children with JIA and relate it to the disease activity including gingival inflammation.*

MATERIALS AND METHODS

Study Design and Subject Recruitment

The present cross-sectional study is a project within NorJIA (Norwegian JIA Study – Imaging, oral health and quality of life in children with juvenile idiopathic arthritis (JIA)), a larger Norwegian prospective multicenter cohort study on JIA registered in Clinical Trials.gov (NCT03904459). The clinical and demographic data was collected between November 2015 and December 2018 at the Department of Pediatrics and Adolescence Medicine, University Hospital North Norway (UNN), Public Dental Service Competence Centre of North Norway (PCNN), Tromso, Haukeland University Hospital Bergen, and Oslo University Hospital, Rikshospitalet, Oslo. Informed consent was collected from all study participants and the study was approved by the Institutional Medical Research Ethics Committee (2015/318). Ninety-three children in total were recruited; Fifty-nine patients with JIA: patients with JIA

TABLE 1 | Demographic and disease activity characteristics among children with juvenile idiopathic arthritis (JIA) and healthy controls (HC).

	JIA (n = 59)	HC (n = 34)	Cut-off	P-value*
Demographic characteristics				
Female, number (%)	43 (73)	27 (79)		0.48[a]
Age at sampling, years	12.6 ± 2.7	12.3± 3.0		0.65[b]
Age at onset	6.0 (2.0–10.0)			–
Geographics, number (%)				
Troms county	34 (58)	34 (100)		–
Finnmark county	17 (29)			–
Nordland county	5 (9)			–
Eastcoast county	2 (3)			–
Westcoast county	1 (2)			–
Disease duration, years	5.0 (3.0–10.0)			–
JIA category, number (%)				
Persistent oligoarthritis	11 (19)			–
Extended oligoarthritis	13 (22)			–
Polyarthrtitis RF positive	3 (5)			–
Polyarthrtitis RF negative	15 (25)			–
Systemic arthritis	0 (0)			–
Psoriatic arthritis	3 (5)			–
Enthesitis related arthritis	7 (12)			–
Undifferentiated arthritis	7 (12)			–
GBI, % (IQR)	22 (6–44) (n = 44)	6 (0–11) (n = 25)	>10	0.00[b]
OHI-S (IQR)	0.5 (0.3–0.8) (n = 43)	0.3 (0.0–0.4) (n = 25)		0.00[b]
DI-S (IQR)	0.5 (0.3–0.8) (n = 43)	0.3 (0.0–0.3) n = 25)		0.00[b]
Disease activity variables**				
JADAS10	12.8 (7.6–18.0) n = 48			
Patients with active disease, number (%)	44 (74)			–
Patients with active joints, number (%)	23 (39)			–
Patients with TMJ arthritis, number (%)	15 (25)			–
Patients with IACs to the TMJ, number (%)	8 (13)			–
Numberof active joints	0.0 (0.0–1.0)			
MDgloVAS	2.5 (1.0–5.0) (n = 58)			
PRgloVAS	2.5 (0.5–4.0) (n = 49)			
HLA-B27 positive, number (%)	20 (36.4) (n = 55)			–
Rheumatoid factor positive, number (%)	1 (2.0) (n = 51)			–
Type of Medication				
No DMARDs, number (%)***	15 (25)			
Methotrexate, number (%)	20 (34)			–
Biologics combination, number (%)****	24 (41)			–

*Values are the median (IQR) unless indicated otherwise. [a]Chi-square test. [b]Wilcoxon-Mann-Whitney test. *P <0.05 for statistical significance. **Remission status according to the ACR provisional remission criteria (Wallace et al., 2011); ***NSAIDs and/or IACs; ****Current or previous use alone or in combination with other DMARDs; JIA, juvenile idiopathic arthritis; GBI, gingival bleeding index; OHI-S, simplified oral hygiene index; DI-S, simplified debris index; JADAS10, the composite juvenile arthritis10-joint disease activity score; TMJ, temporomandibular joint; MDgloVAS, medical doctor global evaluation of overall disease activity on a 10-cm visual analogue scale; IACs, intraarticular corticosteroid injections; DMARDs, disease modifying antirheumatic drugs.*

and newly diagnosed temporomandibular joint (TMJ) arthritis (n = 15) at the departments above, consecutive patients with JIA without TMJ arthritis (n = 44) from the outpatient clinic at UNN, and healthy controls (HC; n = 34) at PCNN matched for age and gender to 34 of the patients with JIA (**Table 1**). HC were recruited from the larger multicenter study NorJIA, and consisted of children attending the regular Norwegian community dental care. The clinical characteristics of the two groups are presented in **Table 1**. Inclusion criteria for patients with JIA were fulfillment of the JIA classification criteria defined by the International League of Associations for Rheumatology (ILAR) (Petty et al., 2004), and age at the study visit <18 years, with or without arthritis activity in one or both TMJs. TMJ arthritis was defined as clinical symptoms and findings in addition to signs of arthritis in magnetic resonance imaging (MRI). Patients on antibiotics prior to sampling were excluded.

Demographics and Assessment of JIA Disease Activity

Patient demographics, subtype of JIA, duration and onset of JIA, medication, general clinical examination, measures of disease activity and severity were collected by three experienced pediatric rheumatologists calibrated through regular meetings in the study period with clinical variables thoroughly discussed and defined in a common study protocol based on the Temporomandibular joint Juvenile Arthritis Working group (TMJaw) recommendations (Stoustrup et al., 2019). Number of active joints was defined according to the general definition of arthritis: swelling within a joint or limitation in the range of joint movement with joint pain or tenderness (Filocamo et al., 2011), while TMJ arthritis was based on both clinical signs and symptoms, and MRI imaging. Patient-reported global assessment of overall well-being (PRgloVAS) and patient-reported pain (PRpainVAS) within the last week on a 10-cm visual analogue scale (VAS) were collected. On this scale, 0 indicates

no activity/no pain/best global health, and 10 indicate the maximum activity/worst pain/poorest global health, respectively. A routine complete blood cell count, including rheumatoid factor (RF) and human leukocyte antigen B27 (HLA-B27) was registered. The composite juvenile arthritis disease activity score (JADAS10, range from 0 to 40) was calculated as the simple sum of the medical doctor global evaluation of overall disease activity on a 10-cm visual analogue scale (VAS), MDgloVAS (range 0–10), PRgloVAS (range 0–10), active joint count (up to maximum 10 joints), and the erythrocyte sedimentation rate (ESR) (normalized to 0–10) (Consolaro et al., 2009; Consolaro et al., 2011). Inactive disease was defined according to the ACR provisional criteria requiring all the following: 1) no active joints; 2) no fever, rash, serositis, splenomegaly or generalized lymphadenopathy attributable to JIA; 3) no active uveitis; 4) normal ESR or C-reactive protein (CRP); 5) MDgloVAS = 0; and 6) duration of morning stiffness of ≤15 minutes (Wallace et al., 2011).

Intraoral Examination and Collection of Saliva

A modified version of the Gingival bleeding index (GBI) (Ainamo and Bay, 1975) and the Simplified Oral Hygiene Index (OHI-S) (Greene and Vermillion, 1964) were used with 6 index teeth for two reasons: 1) The youngest children had transitional dentitions with premolars and canines not yet erupted. Central incisors and first molars being the first permanent teeth to be erupted were therefore chosen as index teeth. 2) For the youngest children it was considered too exhausting and time consuming to investigate the complete dental set during the oral examination. For GBI, a dental probe was carefully applied without any pain, in the upper part of the gingival sulcus, and then removed without doing a horizontal movement along the tooth surface. Bleeding on probing within 10 seconds was registered. Angulation of the dental probe of 60 degrees to the vertical axis of the tooth was applied if possible. The mesial, distal and central site of the buccal surface of the index teeth 16, 26, 11, and 31 were chosen together with the lingual surface of the index teeth 36 and 46. The number of bleeding tooth sites were divided to the total number of tooth sites examined and finally presented as the mean percentage (%), range 0% to 100% where a higher percentage represents a worse score in bleeding. Gingival inflammation was diagnosed according to a gingival bleeding index cut-off score ≥ 10% (Trombelli et al., 2018).

OHI-S is a sum score of simplified-debris index (DI-S) and simplified-calculus index (CI-S) and is presented as a mean score index, where a higher index represents a worse score in OHI-S. DI-S and CI-S are the buccal scores + the lingual scores divided by the total number of examined surfaces. OHI-S was not calculated in children with fixed orthodontic appliances and subgingival status was not evaluated. All children ≥12 years filled out an oral health related questionnaire and the parents/proxies filled out for children <12 years. One of the questions was; how often do you brush your teeth: 1) Never 2) Most days 3) Once daily 4) Twice daily or more.

Two calibrated specialists in oral and maxillofacial surgery and pediatric dentistry (PF, JH) collected before oral examination, unstimulated whole saliva for 6 minutes and paraffin chewing stimulated whole saliva for 3 minutes, according to a standardized protocol (i.e. restrictions to food and drinks 2 h prior to sampling). Furthermore, medications taken the same day or the day before sampling was recorded. Only SWS were used for microbial analyses. After collection, each saliva sample was aliquoted and placed in a −80°C freezer until further analyses.

DNA Extraction

Seven hundred and fifty microliters from each SWS sample was mixed with an equal amount of phosphate buffer saline (PBS) and spun down at 9600 g for 5 minutes, before the supernatant was carefully removed. The pellet was resuspended in 155 mL PBS and 25 mL MetaPolyzyme multilytic enzyme mix (Zigma-Aldrich, USA) and incubated on a 37°C heat block for 4 h, for digestion of the bacterial cell wall. The digests were then transferred to a QIAcube (Qiagen, Hilden, Germany) for DNA extraction using preprogramed protocol using the QIAamp DNA Mini Kit (Qiagen, Germany) with 100 µL elution volume. The quality of the isolated DNA (high molecular weight and non-fragmented DNA) was assessed by running extracted DNA samples on agarose gel (1%) with 1 kb ladder (Termo Fisher Scientific, Invirtogen, USA). The amount of yield DNA was then measured using Invitrogen Qubit 3.0 Fluorometer (Termo Fisher Scientific, Invirtogen, USA) according to the manufacturer's instructions.

16S rRNA Sequencing and Bioinformatic Analysis

16S rRNA gene library preparation and sequencing were done at the Australian Center for Ecogenomics (Brisbane, Australia) as detailed previously (Al-Hebshi et al., 2017a). Briefly, the V1-3 region was amplified using the degenerate primers 27FYM (Frank et al., 2008) and 519R (Lane et al., 1985) in standard PCR conditions. The amplicons (~ 520 bp) were purified and indexed with unique 8-base barcodes in a second PCR. The tagged libraries were then pooled together in equimolar concentrations and sequenced using MiSeq v3 2 × 300 bp chemistry (Illumina, USA). Preprocessing of data (merging of reads, primer trimming, quality-filtration, alignment and chimera removal) was performed as detailed elsewhere (Al-Hebshi et al., 2017b). The high quality, merged reads were assigned species-level taxonomies using our BLASTn-based algorithm (Al-Hebshi et al., 2015; Al-Hebshi et al., 2017b). The resultant microbial profiles were used as input to QIIME (Quantitative Insights Into Microbial Ecology) software package version 1.9.1 (Caporaso et al., 2010) for downstream analysis including subsampling, generation of taxonomy plots/tables and rarefaction curves, and calculation of species richness, coverage, alpha diversity indices and beta diversity distance matrices. Principal component analysis (PCoA) was used for clustering samples based on overall microbial similarity, while Linear discriminant analysis (LDA)

effect size (LEfSe) (Segata et al., 2011) was used to detect differentially abundant taxa between the groups.

Statistical and Bioinformatic Analyses

For description of clinical and demographic data, median (IQR), mean (standard deviation) and frequencies (percentage) were used. Different disease characteristics and associations between patients with JIA and HC were analyzed by chi-square test or Fisher's exact test for categorical variables and Student's t-test for continuous variables if reasonably normally distributed, otherwise Man-Whitney U test was used. A multivariable logistic regression analysis was performed to adjust for OHI-S, age and gender in the association between JIA and GBI. A *P*-value ≤0.05 was considered statistically significant for clinical parameters. For testing correlation between species and measures of disease activity (JADAS10, number of active joints), p-values were adjusted for multiplicity with Benjamini-Hockberg method (FDR ≤0.1). To assess the association between the bacterial profiles and disease activity (JADAS10 and number of active joints) a Spearman correlation matrix was computed with R package. Correlations with P-value ≤ 0.01 were considered significant.

RESULTS

Demographic and Disease Activity Parameters

Demographics and disease activity parameters for the group with JIA, and HC are given in **Table 1**. There was a female predominance in both groups, and RF negative polyarthritis was the most common category among children with JIA (25%). The simplified oral hygiene index (OHI-S) was significantly higher among children with JIA with median 0.5 (IQR 0.3–0.7) compared to median 0.3 (IQR 0.0–0.4) in HC. Also, the simplified debris-index (DI-S) was higher in JIA (**Table 1**). There was also significantly higher modified gingival bleeding index (GBI) in the group with JIA with median 22 (Interquartile range (IQR)) 6–40) % compared to HC with median 6.0 (IQR 0–11) % but no association was found between JIA and GBI when adjusting for OHI-S (**Supplementary Table 1**). Within the JIA group no significant differences in GBI were found between patients without DMARDs, on methotrexate or on biologics, with GBI median 10 (IQR 10–30) %, median 20 (IQR 10–40) %, and median 30 (IQR 20–40) %, respectively. There were no difference in frequency of tooth brushing between children with JIA and HC, in JIA 37 of 47 (79%) and in HC 23 of 32 (72%) brushed their teeth twice or more daily. No differences were seen in toothbrush frequency, OHI-S, DI-S, or GBI between children with and without TMJ-arthritis. Restrictions according to food and drinks 2 h prior to saliva sampling were taken in 41 of 93 patients (44%). In 10 of 93 (11%) no restrictions were taken, and in 42 of 93 (45%) information on restrictions were not available. Among the 93 participants 16 reported intake of oral medications other than disease modifying antirheumatic drugs (DMARDs) such as non-steroidal anti-inflammatory drugs (NSAIDs) and/or other medications. Three participants reported intake of oral or parenteral methotrexate, and 6 parenteral biologic agents, while the remaining 68 participants reported no intake of medication the same day or the day before saliva sampling.

Sequencing and Data Processing Statistics

The raw data has been deposited and is publicly available from SRA (# PRJNA605805). A total of ~8.2 million sequences were obtained (range of 30,113–526,987 reads per sample), of which about 85% were successfully merged; however, only 35% were retained after quality filters and 20% after chimera removal. Of the high-quality, non-chimeric reads, 88% could be assigned species-level taxonomy (mean of 15 360± 18 368 reads per sample). Details of the reads statistics before and after each quality control step are provided in **Supplementary dataset 1**.

Overall Microbial Profile

Using a minimum count of 100 reads per species as cutoff, a total of 216 bacterial species belonging to 58 genera and 8 phyla were identified across all samples; the relative abundances and detection frequencies of these taxa in each sample is provided in **Supplementary dataset 2**, **3** and **4** respectively. On average, 134 species (range, 90–186) and 45 genera (range, 32–58) were detected per subject. Fifty-seven species and 29 genera were identified in more than 90% of the samples and can be defined as core salivary taxa. The average relative abundances of all phyla, top genera and species (those present at an average abundance of ≥ 1% in the control group) in each of the study groups are shown in **Figure 1**. Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria, and Fusobacteria were the major phyla in order accounting for more than 98% of the reads. Thirteen genera accounted for 90% of the average microbiome, with *Prevotella*, *Streptococcus*, *Haemophilus, Actinomyces*, *Porphyromonas* and *Rothia* alone making up ~ 70%. At the species level, *Prevotella melaninogenica, Haemophilus parainfluenzae, Rothia mucilaginosa, Porphyromonas* sp. *oral taxon 279, Prevotella histicola, Actinomyces odontolyticus* were the most abundant species, constituting around 40% of the microbiome on average.

Bacterial Diversity and Differentially Abundant Species

There were no significant differences between children with JIA and the healthy group in the number of species (i.e. alpha diversity, species richness), or in PCoA (i.e. beta diversity, the ratio between the two groups) as shown in **Figure 2**. Differential abundance analysis revealed significant differences (**Figure 3**). At the phylum level, JIA was associated with enrichment of Spirochaetes and Saccharibacteria and depletion of Proteobacteria. Genera *TM7-G1*, *Solobacterium* and *Mogibacterium* were associated with JIA, while *Haemophilus* and *Bacillus* were associated with healthy subjects (**Figure 3B**). *Haemophilus parainfluenzae*, *Leptotrichia species oral taxon 223, Haemophilus pittmanae*, *Prevotella denticola* and *Bacillus subtilis* were key bacterial species associated with the healthy group,

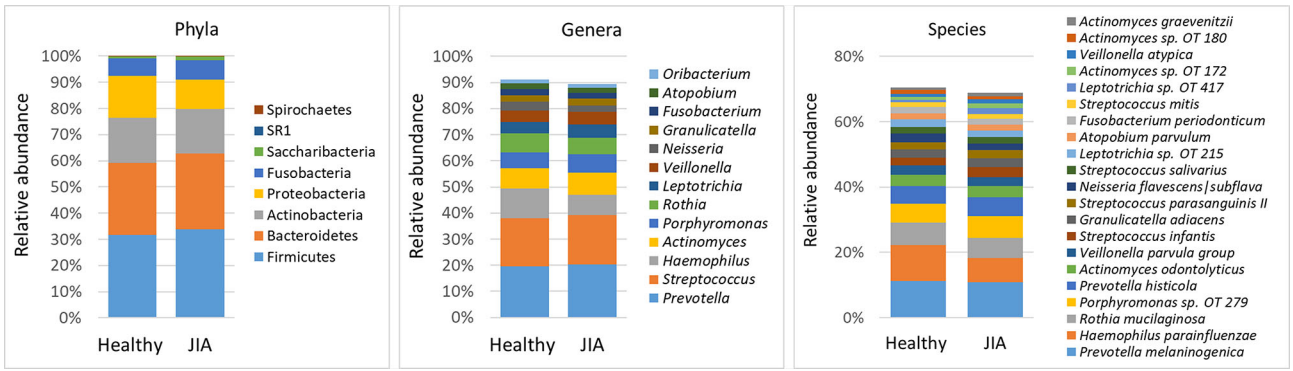


FIGURE 1 | Microbiological profiles. DNA extracted from saliva was sequenced for the V1-V3 region of the 16S rRNA gene using paired-end chemistry. The generated reads were merged, quality-filtered and classified to the species level using a BLASTn-based algorithm. The stacked bars show the average relative abundances of all phyla and top genera and species (those with relative abundance ≥ 1%) identified in the study groups. OT, oral taxon.

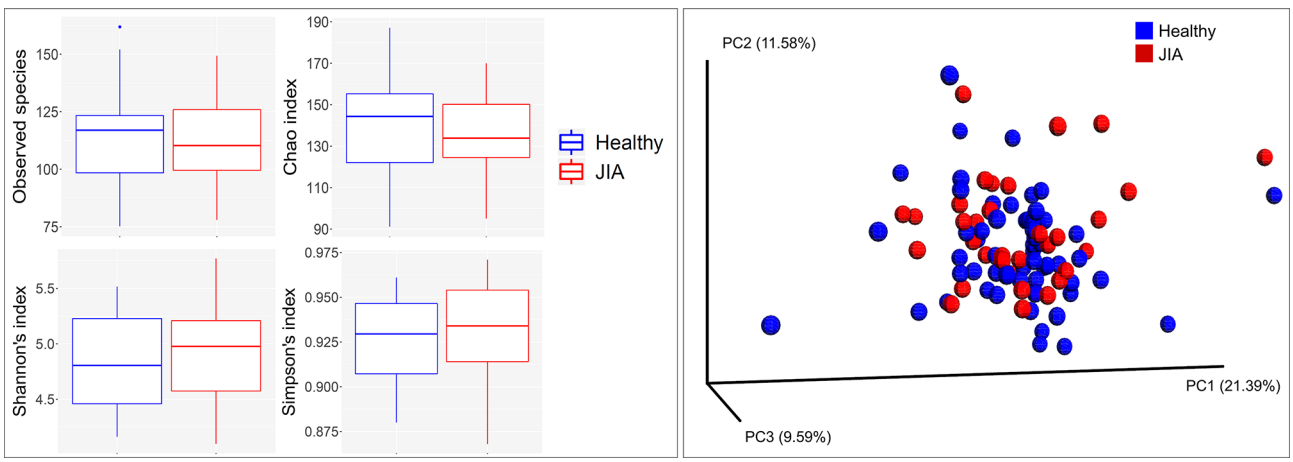


FIGURE 2 | Species richness and diversity. Taxonomic profiles were rarified and used to calculate observed richness, expected richness (alpha diversity index; Chao index), evenness measure (alpha diversity index; Shannon's and Simpson's) and distance matrices employing standard QIIME scripts. Left: Box and whisker plots of species richness and aloha diversity in each group. Differences were not significant by Mann–Whitney U test. Right: Clustering of samples with PCoA based on abundance Jaccard distance matrix. Plots were generated with QIIME and R Package.

whereas the JIA group showed higher abundance of 11 bacterial species of which *Leptotrichia species oral taxon 417, TM7 G1*, *Capnocytophaga species oral taxon 864*, *Veilonella atypica* and *Mogibacterium diversum* were most enriched (**Figure 3C**).

The relative abundances of the top six differentially abundant taxa (based on LDA score) are shown for individual samples in **Figure 4**. These microbial associations between the JIA and the healthy group were independent of the differences between the two groups in GBI. The microbial associations with the GBI are shown in **Supplementary Figure 1**. Notably, after adjustment of p-values for multiple comparisons with Benjamini-Hochberg method only genus *Bacillus* and *B. subtilis* maintained a false discovery rate (FDR) of ≤ 0.1 and the per sample relative abundance plot for *B. subtilis* is shown in **Supplementary Figure 2**.

Microbial Association With Disease Activity

Correlation analysis revealed significant association between a group of bacteria species and disease activity (**Figure 5**). *Gemella morbillorum*, *Leptotrichia* sp. oral taxon 498 and *Alloprevotella* oral taxon 914 correlated positively with the composite juvenile arthritis 10-joint disease activity score (JADAS10), primarily through their association with the medical doctor global evaluation of disease activity (MDgloVAS). *G. morbillorum* also correlated with patient reported global assessment of well-being (PRgloVAS). Several species correlated positively with the number of active joints but *Campylobacter* oral taxon 44 showed the strongest association and was the only species that maintained FDR ≤ 0.1 when *P*-values were adjusted for multiplicity with the Benjamini-Hockberg method.

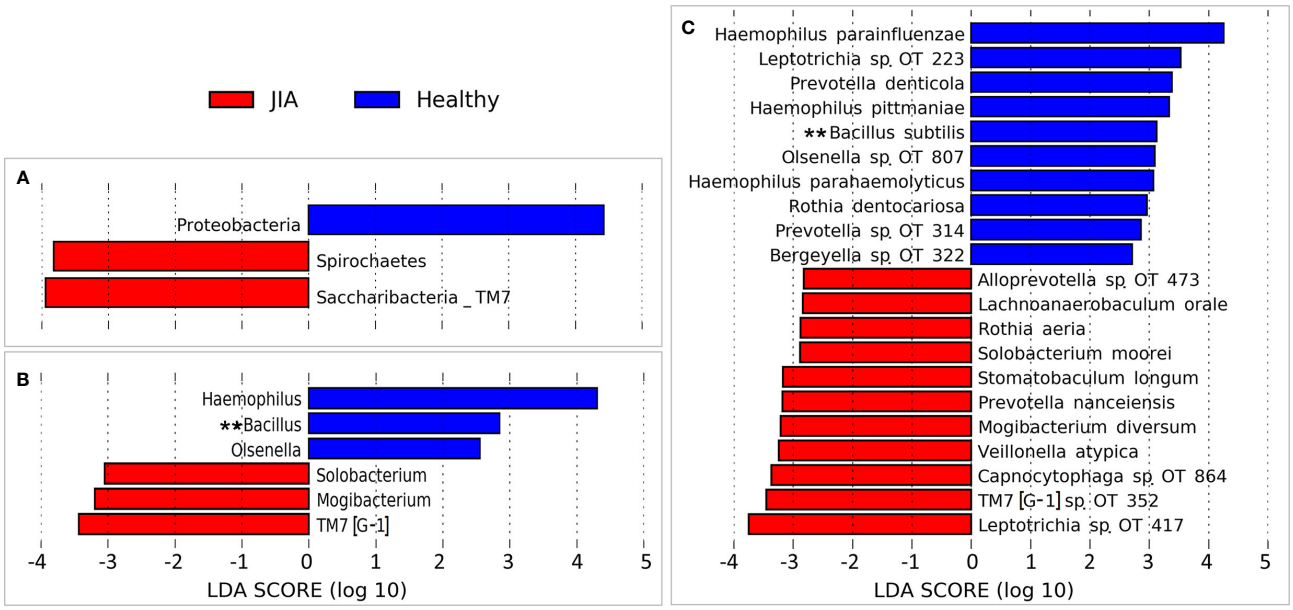


FIGURE 3 | Differentially abundant taxa. **(A)** Phyla, **(B)** Genera and **(C)** species that showed significant differences in relative abundance between the two study groups as identified by linear discriminant analysis (LDA) effect size analysis (LEfSe) – 2.5 LDA score cutoff. OT, oral taxon. **FDR ≤ 0.1 (Benjamini-Hochberg method).

Children with JIA and TMJ arthritis showed different microbial associations compared to JIA subjects without TMJ arthritis (**Figure 6**). *Haemophilus parainfluenzae, Prevotella pallens, actinomyces species oral taxon 180* were the top species differentially enriched in JIA with TMJ subjects. *Campylobacter* oral taxon 44 also showed an association with TMJ arthritis (**Supplementary Figure 3**), although opposed to its association with the number of active joints, it did not maintain FDR ≤ 0.1. *Rothia mucilaginosa, Atopobium parvulum* and *Oribacterium sinus* were significantly more abundant in children with JIA without TMJ arthritis.

DISCUSSION

This study is to our knowledge, the first to examine the salivary oral microbiome in JIA patients with the NGS method. Different methods are suggested for collecting saliva samples which might alter the NGS sequencing results but it seems that the overall microbial composition of saliva is not significantly affected (Lim et al., 2017; Mascitti et al., 2019). Salivary microbiome is a good representative of bacteria found in many oral mucosal sites such as cheeks, tongue, gingiva, especially when studying inflammatory diseases rather than subgingival plaque, which is rather a site-specific medium for the bacteria associated with the periodontium.

The oral environment is a complex environment that contains distinct microbial niches each with its distinct microbial inhabitant in mucosal membranes and subgingival plaque. Sampling of saliva has been used when studying dysbiosis of salivary microbiota in inflammatory bowel disease (Said et al., 2014), and sampling by tongue and buccal mucosal brushings have been used in ulcerative colitis (Docktor et al., 2012).

In our study, saliva samples were directly aliquoted and immediately frozen before further processing. Other collection methods for sampling of oral microbiota are available and appear to be stable concerning bacterial diversity at ambient temperature after 4 to 7 days, such as mouthwash sampling methods (Vogtmann et al., 2019). Some distinct differences in relative abundance of specific microbial taxa are seen between these methods compared to samples that are frozen immediately (Vogtmann et al., 2019), indicating that direct freezing of samples may still be the best choice.

A majority of the children with JIA were females diagnosed with either oligoarthritis persistent or RF negative polyarthritis in accordance with most population-based studies, pointing to representability of our study case (Nordal et al., 2011). Another strength of this study was that a majority of the participants had no intake of systemic medication the same day or the day before saliva sampling.

A limitation of the study is that information is available only in 41 of 93 children regarding food and drinks restrictions prior to sampling. Additionally, our findings must be evaluated in the context of a limited sample size and being a cross-sectional exploratory study.

Microbial Profile, Bacterial Diversity and Differentially Abundant Species

In our study between 90 and 186 different species were found in the salivary microbiome of the study subjects with fifty-seven species shared in more than 90% of the subjects. The core salivary microbiome in the study subjects comprised 22 species

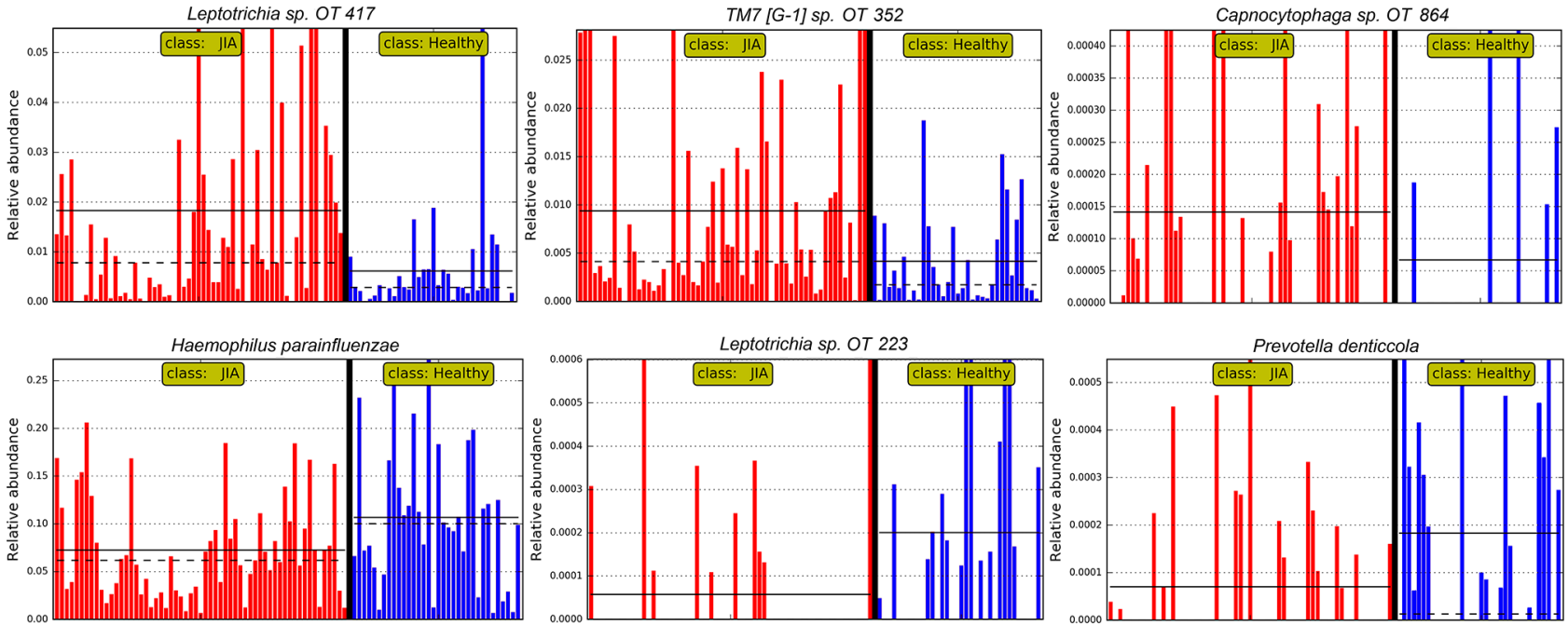


FIGURE 4 | Per sample abundance plots. Relative abundances of top six differentially abundant species (based on LDA score) in individual samples. OT, oral taxon. FDR ≤ 0.1 (Benjamini-Hochberg method).

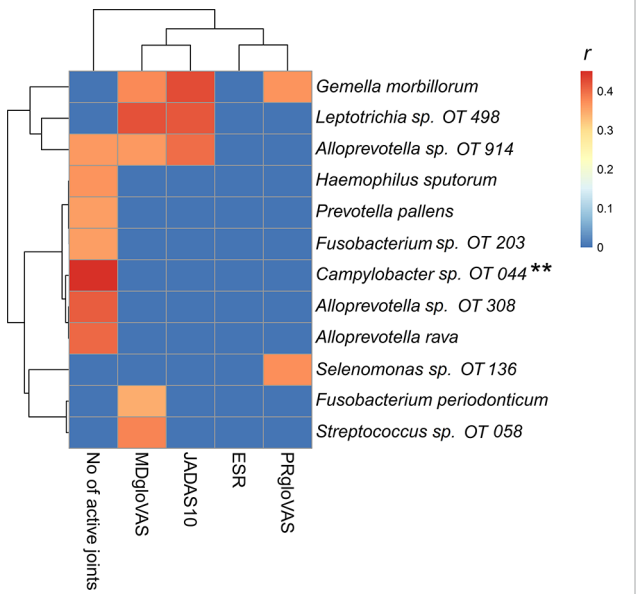


FIGURE 5 | Heatmap of the microbial association with disease activity. A Spearman correlation matrix was computed using R package. Correlations with P-value ≤ 0.01 were considered significant. The r-value for nonsignificant correlations was set to zero (blue on the heatmap). OT, oral taxon. **FDR ≤ 0.1 (Benjamini-Hochberg method). PRgloVAS; patient reported global assessment of well-being; ESR, erythrocyte sedimentation rate; JADAS10, the composite juvenile idiopathic arthritis 10-joint disease activity score; MDgloVAS, the medical doctor global evaluation of disease activity.

detected in all individuals. In other studies the core microbiome was found to be similar or even in a lower number (Hall et al., 2017; Hansen et al., 2018). We found no significant differences between children with JIA and HC in species richness or in PCoA, i.e. alpha or beta diversity. Differential abundance analysis, however, revealed that JIA was significantly associated with taxa associated with chronic inflammation, and that several of these species including *Campylobacter* oral taxon 44 correlated with disease activity in terms of increased number of active joints. The predominant bacterial genera found in the salivary microbiome in both JIA and healthy saliva in our study consisted of *Prevotella, Streptococcus, Haemophilus, Actinomyces*, *Porphyromonas* and *Rothia*, which is similar to other studies in both children with JIA and healthy (De Filippo et al., 2019). The microbial diversity and richness of the salivary oral microbiome between the JIA and healthy controls were comparable. This is in line with other studies that investigated the oral microbial diversity between patients with chronic inflammatory diseases, i.e. rheumatoid arthritis, and healthy controls (Scher et al., 2012; Tong et al., 2019). Differential abundance analysis revealed several taxa to be associated with or depleted in JIA. At phylum level, we found an overabundance of Spirochaetes and Saccharibacteria and depletion of Proteobacteria. This is in line with Xu et al. showing Proteobacteria (*Neisseria*) and Firmicutes (*Selenomonas*) as a healthy core salivary microbiome, together with the phyla Bacteroidetes (*Porphyromonas*). They also showed that the salivary microbial composition shifts with aging in children, and a strength of our study is age-matching between children with JIA and healthy controls (Xu et al., 2018).

At the genus level, *Haemophilus*, *Bacillius* and *Olsenella* were depleted from JIA patients while there were significant overabundance of bacteria known to be associated with chronic inflammation, such as *TM7-G1*, *Solobacterium* and *Mogibacterium* (Moore et al., 1985; Casarin et al., 2012; Hiranmayi et al., 2017). The higher abundance of *Haemophilus parainfluenzae*, *Leptotrichia species oral taxon 223, Haemophilus pittmanae*, *Prevotella denticola* and *Bacillus subtilis* in the healthy group in our study could highlight the importance for health of sustaining a high proportion of these species in the oral microbiome. *Bacillus subtilis* species has been reported to have a role in limiting inflammatory response by down-regulation of the pro-inflammatory interleukin-8 production and up regulation of inducible nitric oxide synthase (iNOS) protein levels (Rhayat et al., 2019). The depletion of *Haemophilus* species in salivary microbiome has also been reported in patients with rheumatoid arthritis (Zhang et al., 2015). On the other hand *Leptotrichia species oral taxon 417, TM7 G1*, *Capnocytophaga species oral taxon 864*, *Veilonella atypica*, and *Mogibacterium diversum* were found to be highly abundant in children with JIA. The oral taxon TM7 has been associated with chronic inflammation (Demmer et al., 2017). In line with our study, Grevich et al. found depletion of genera Prevotella (phylum Bacteroidetes) in JIA, but they report an overabundance of the genera Haemophilus and Kingella (phylum proteobacteria) in JIA, which was not found in our study (Grevich et al., 2019).

In children with JIA the saliva in those with TMJ arthritis was enriched with certain bacteria compared to those without TMJ arthritis. This finding was not explained by differences in toothbrush frequency or differences in oral hygiene indices (i.e. DI-S, OHI-S, GBI), which was similar in both groups.

The role of oral and gut microbiota in many inflammatory diseases has been suggested, and there is evidence for the role of molecular mimicry in such diseases. It has for example, been reported that a molecular mimicry between a peptide from the von Willebrand factor type A from the oral microbe *Capnocytophaga ochracea* can be attributed to the activation of the Sjögrens syndrome antigen A/Ro60-Reactive T cells (Szymula et al., 2014).

Microbial Association to Disease Activity

Interestingly, in our study *Campylobacter* oral taxon 44 (proteobacteria at the phyla level) showed a moderate correlation with the number of joints affected in patients with JIA. *Campylobacter* has a well-known association with reactive arthritis and other inflammatory diseases in both children and adults (Lackner et al., 2019). Dong et al. found a negative correlation between disease activity and *Proteobacteria*, *Ruminococcaceae, Faecalibacterium*, or *Enterobacteriaceae* (Dong et al., 2019), in the gut microbiota in 32 patients with JIA. In line with our study, the same authors found a positive correlation between disease activity and increased abundance of phyla *Firmicutes (our study: G.morbillorum), Bacteroidetes (our study: Alloprevotella, prevotella pallens)*, and *Bacteroidaceae* at the phyla level (Dong et al., 2019).

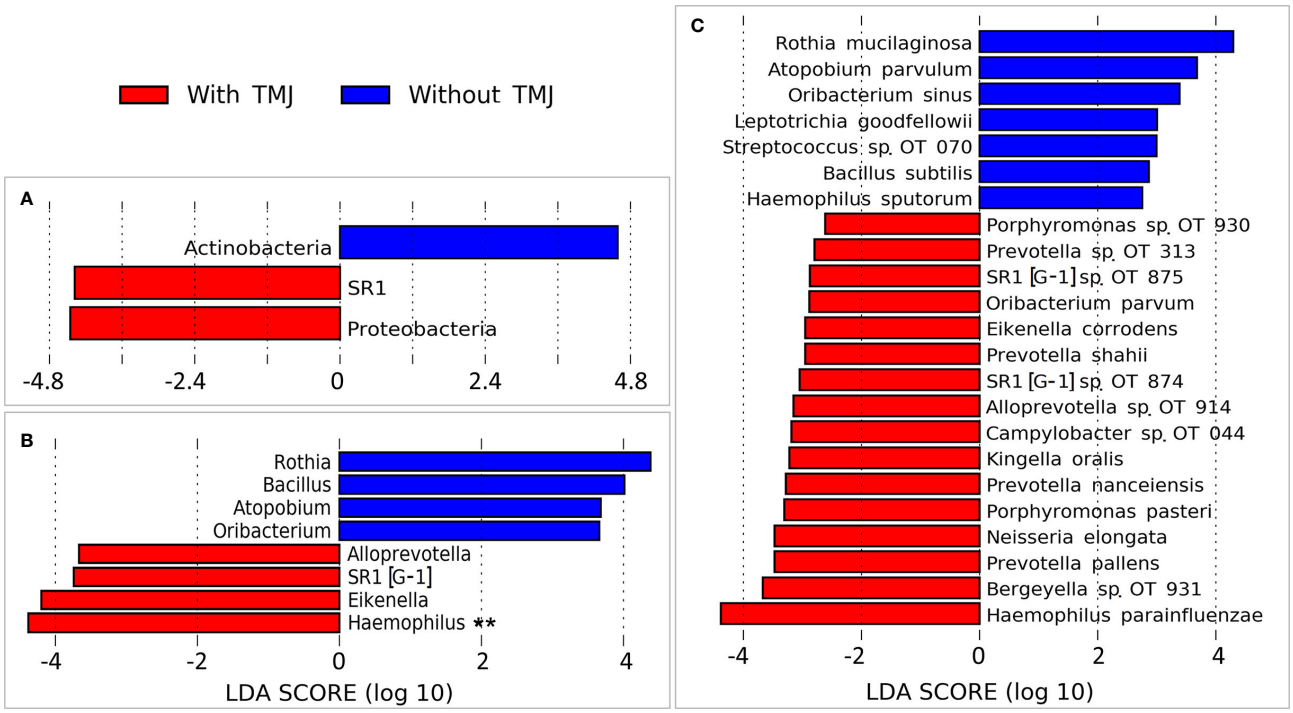


FIGURE 6 | Differentially abundant taxa by TMJ arthritis. **(A)** Phyla, **(B)** Genera and **(C)** species that showed significant differences in relative abundance between the JIA subjects with and without TMJ involvement, as identified by linear discriminant analysis (LDA) effect size analysis (LEfSe). 2.5 LDA score cutoff. OT, oral taxon. **FDR ≤ 0.1 (Benjamini-Hochberg method).

The species found enriched in our study on the salivary oral microbiome in JIA are not the same as found in studies on the gut microbiome in JIA (Stoll et al., 2014; Di Paola et al., 2016; Stoll et al., 2016; Tejesvi et al., 2016; Aggarwal et al., 2017; Stoll et al., 2018; Dong et al., 2019; Van Dijkhuizen et al., 2019). However, in all studies disruption of microbial ecology was found with overabundance of taxa associated to chronic inflammation in JIA, and this dysbiosis may have community effects on the host more powerful than the actions of just one single microbe (Chriswell and Kuhn, 2020). Dijkhuizen et al. report dysbiosis in gut microbiota in children newly diagnosed with JIA compared to healthy controls. They also found that age and geographic origin were connected to microbiota profiles (Van Dijkhuizen et al., 2019). In rheumatoid arthritis both oral and gut dysbiosis are well described and no consistent single bacterial species appears to be the causing agent (Bergot et al., 2020). Interestingly, periodonto-pathogenic bacteria such as *Porphyromonas gingivalis* have been suggested to contribute to generation of anti-citrullinated protein antibodies (ACPAs) in rheumatoid arthritis (Moen et al., 2006; Wegner et al., 2010a; Wegner et al., 2010b). Dysbiosis and periodontitis were also found to be associated to increased severity of rheumatoid arthritis (Scher et al., 2012). Other factors reported to be associated with dysbiosis are diet, lifestyle and drug use, especially the use of antibiotics (Turnbaugh et al., 2009; Arvonen et al., 2015). Early life antibiotic use is shown to increase the risk of developing JIA later in life, and may predispose due to a shift in microbiota composition (Arvonen et al., 2015; Horton et al., 2015). In our study none of the participants were on any antibiotics on the day of sampling, but previous history of antibiotic use were not recorded.

Gingival Inflammation and the Oral Microbiome

The significantly higher gingival inflammation found in patients with JIA compared to healthy controls is in line with many studies investigating JIA and oral health (Welbury et al., 2003; Ahmed et al., 2004; Leksell et al., 2008; Santos et al., 2015; Grevich et al., 2019). Other studies find no significant difference between JIA and healthy controls (Miranda et al., 2003; Savioli et al., 2004; Reichert et al., 2006; Feres De Melo et al., 2014; Pugliese et al., 2016; Kobus et al., 2017; Maspero et al., 2017), probably depending on different study design and measurement indices of gingival inflammation (Skeie et al., 2019). Despite higher GBI, no difference in frequency of tooth brushing was found between JIA and healthy controls in our study. However, we do not know how effective the tooth brushing was performed, some of the children with JIA have restricted wrist, finger or jaw movements that may reduce the quality of the tooth brushing. After adjusting for dental plaque and calculus (OHI-S), JIA was not found to be a predictor for gingival inflammation in terms of higher GBI. In line with other studies (Reichert et al., 2006) we found dental plaque (OHI-S) to be associated to gingival inflammation. There were no overlap

between overabundant bacteria associated with GBI and those associated with JIA. Overabundance of bacteria associated with chronic inflammation in JIA could be explained by a disruption of microbial hemostasis in JIA and not by gingival inflammation in JIA.

There are some indications that the biologic agent etanercept might reduce periodontal inflammation in children with JIA (Maspero et al., 2017). In our study children with JIA had more gingival bleeding compared to healthy controls, despite immune-modulating medication. Altogether 40% were on biologic treatment either alone or in combination with methotrexate, but we found no significant differences in GBI between the different medication groups.

CONCLUSION

Several taxa, including genera Solobacterium, Mogibacterium, and TM7-G1 known to be associated with chronic inflammation, were found enriched in the saliva of children with JIA and were associated with disease activity in our study. No significant difference was found in alpha- and beta-diversity compared to healthy. Prospective cohort-studies with treatment-naïve patients with new onset JIA are warranted to further elucidate the role of the oral microbiome in disease etiology and severity.

DATA AVAILABILITY STATEMENT

The raw data has been deposited and is publicly available from SRA (# PRJNA605805).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Medical Research Ethics Committee. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

Planning of the study, analysis of the data and interpretation of the results as well as writing of the manuscript: PF, MA-H, EN, NA-H. Collection clinical data: PF, EN, JB, BF, AR, VR, NS, JH, EG, and LC. MA-H supervised the laboratory work and DB, FA TC, and NA-H the DNA sequencing analysis. Critical review and editing of manuscript: PF, EN, JB, BF, AR, VR, NS, JH, EG, LC, NA-H, and MA-H. All authors contributed to the article and approved the submitted version.

FUNDING

Grethe Harbitz funds, Research funding from Troms County, Norsk Revmatikerforbund, Helse Nord Research funding and Tromsø Research Foundation and Department of Clinical Dentistry, UiT The Arctic University of Norway.

ACKNOWLEDGMENTS

The publication charges for this article have been funded by a grant from the publication fund of UiT The Arctic University of Norway. We thank the The NorJIA study group (Norwegian JIA Study – Imaging, oral health and quality of life in children with juvenile idiopathic arthritis) and the research nurse Lisbeth K. B. Aune for logistics, recruitment, and methods contributions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2020.602239/full#supplementary-material

SUPPLEMENTARY FIGURE 1 | Heatmap of the microbial association with GBI. A Spearman correlation matrix was computed using R package. Correlations with P-value ≤ 0.01 were considered significant.

SUPPLEMENTARY FIGURE 2 | Per sample abundance plots for *Bacillus subtilis*. Relative abundances of *Bacillus subtilis* in individual samples by disease status (left) and TMJ involvement status (right).

SUPPLEMENTARY FIGURE 3 | Per sample abundance plots for *Campylobacter* sp. oral taxon 44. Relative abundances of *Campylobacter* sp. oral taxon 44 in individual samples by TMJ involvement status.

SUPPLEMENTARY DATA SHEET 1 | Sequencing and data processing statistics.

SUPPLEMENTARY DATA SHEET 2 | Relative abundances and detection frequencies of phyla identified in the individual samples.

SUPPLEMENTARY DATA SHEET 3 | Relative abundances and detection frequencies of genera identified in the individual samples.

SUPPLEMENTARY DATA SHEET 4 | Relative abundances and detection frequencies of species identified in the individual samples.

REFERENCES

Aggarwal, A., Sarangi, A. N., Gaur, P., Shukla, A., and Aggarwal, R. (2017). Gut microbiome in children with enthesitis-related arthritis in a developing country and the effect of probiotic administration. *Clin. Exp. Immunol.* 187, 480–489. doi: 10.1111/cei.12900

Ahmed, N., Bloch-Zupan, A., Murray, K. J., Calvert, M., Roberts, G. J., and Lucas, V. S. (2004). Oral health of children with juvenile idiopathic arthritis. *J. Rheumatol.* 31, 1639–1643.

Ainamo, J., and Bay, I. (1975). Problems and proposals for recording gingivitis and plaque. *Int. Dent. J.* 25, 229–235.

Al-Hebshi, N. N., Nasher, A. T., Idris, A. M., and Chen, T. (2015). Robust species taxonomy assignment algorithm for 16S rRNA NGS reads: application to oral carcinoma samples. *J. Oral. Microbiol.* 7, 28934. doi: 10.3402/jom.v7.28934

Al-Hebshi, N. N., Alharbi, F. A., Mahri, M., and Chen, T. (2017a). Differences in the Bacteriome of Smokeless Tobacco Products with Different Oral Carcinogenicity: Compositional and Predicted Functional Analysis. *Genes (Basel)* 8 (4), 106. doi: 10.20944/preprints201703.0030.v1

Al-Hebshi, N. N., Nasher, A. T., Maryoud, M. Y., Homeida, H. E., Chen, T., Idris, A. M., et al. (2017b). Inflammatory bacteriome featuring Fusobacterium nucleatum and Pseudomonas aeruginosa identified in association with oral squamous cell carcinoma. *Sci. Rep.* 7, 1834. doi: 10.1038/s41598-017-02079-3

Arvonen, M., Virta, L. J., Pokka, T., Kroger, L., and Vahasalo, P. (2015). Repeated exposure to antibiotics in infancy: a predisposing factor for juvenile idiopathic arthritis or a sign of this group's greater susceptibility to infections? *J. Rheumatol.* 42, 521–526. doi: 10.3899/jrheum.140348

Bergot, A. S., Giri, R., and Thomas, R. (2020). The microbiome and rheumatoid arthritis. *Best Pract. Res. Clin. Rheumatol.* 33 (6), 101497. doi: 10.1016/j.berh.2020.101497

Berntson, L., Andersson Gare, B., Fasth, A., Herlin, T., Kristinsson, J., Lahdenne, P., et al. (2003). Incidence of juvenile idiopathic arthritis in the Nordic countries. A population based study with special reference to the validity of the ILAR and EULAR criteria. *J. Rheumatol.* 30, 2275–2282.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Casarin, R. C., Saito, D., Santos, V. R., Pimentel, S. P., Duarte, P. M., Casati, M. Z., et al. (2012). Detection of Mogibacterium timidum in subgingival biofilm of aggressive and non-diabetic and diabetic chronic periodontitis patients. *Braz. J. Microbiol.* 43, 931–937. doi: 10.1590/S1517-83822012000300012

Cho, I., and Blaser, M. J. (2012). The human microbiome: at the interface of health and disease. *Nat. Rev. Genet.* 13, 260–270. doi: 10.1038/nrg3182

Chriswell, M. E., and Kuhn, K. A. (2020). Microbiota-mediated mucosal inflammation in arthritis. *Best Pract. Res. Clin. Rheumatol.* 33 (6), 101497. doi: 10.1016/j.berh.2020.101492

Consolaro, A., Ruperto, N., Bazso, A., Pistorio, A., Magni-Manzoni, S., Filocamo, G., et al. (2009). Development and validation of a composite disease activity score for juvenile idiopathic arthritis. *Arthritis Rheum.* 61, 658–666. doi: 10.1002/art.24516

Consolaro, A., Ruperto, N., Pistorio, A., Lattanzi, B., Solari, N., Galasso, R., et al. (2011). Development and initial validation of composite parent- and child-centered disease assessment indices for juvenile idiopathic arthritis. *Arthritis Care Res. (Hoboken)* 63, 1262–1270. doi: 10.1002/acr.20509

De Filippo, C., Di Paola, M., Giani, T., Tirelli, F., and Cimaz, R. (2019). Gut microbiota in children and altered profiles in juvenile idiopathic arthritis. *J. Autoimmun.* 98, 1–12. doi: 10.1016/j.jaut.2019.01.001

Demmer, R. T., Breskin, A., Rosenbaum, M., Zuk, A., Leduc, C., Leibel, R., et al. (2017). The subgingival microbiome, systemic inflammation and insulin resistance: The Oral Infections, Glucose Intolerance and Insulin Resistance Study. *J. Clin. Periodontol.* 44, 255–265. doi: 10.1111/jcpe.12664

Di Paola, M., Cavalieri, D., Albanese, D., Sordo, M., Pindo, M., Donati, C., et al. (2016). Alteration of Fecal Microbiota Profiles in Juvenile Idiopathic Arthritis. Associations with HLA-B27 Allele and Disease Status. *Front. Microbiol.* 7, 1703. doi: 10.3389/fmicb.2016.01703

Docktor, M. J., Paster, B. J., Abramowicz, S., Ingram, J., Wang, Y. E., Correll, M., et al. (2012). Alterations in diversity of the oral microbiome in pediatric inflammatory bowel disease. *Inflammation Bowel Dis.* 18, 935–942. doi: 10.1002/ibd.21874

Dong, Y. Q., Wang, W., Li, J., Ma, M. S., Zhong, L. Q., Wei, Q. J., et al. (2019). Characterization of microbiota in systemic-onset juvenile idiopathic arthritis with different disease severities. *World J. Clin. cases* 7, 2734–2745. doi: 10.12998/wjcc.v7.i18.2734

Feres De Melo, A. R., Ferreira De Souza, A., De Oliveira Perestrelo, B., and Leite, M. F. (2014). Clinical oral and salivary parameters of children with juvenile idiopathic arthritis. *Oral. Surg. Oral. Med. Oral. Pathol. Oral. Radiol.* 117, 75–80. doi: 10.1016/j.oooo.2013.08.024

Filocamo, G., Consolaro, A., Schiappapietra, B., Dalpra, S., Lattanzi, B., Magni-Manzoni, S., et al. (2011). A new approach to clinical care of juvenile idiopathic arthritis: the Juvenile Arthritis Multidimensional Assessment Report. *J. Rheumatol.* 38, 938–953. doi: 10.3899/jrheum.100930

Frank, J. A., Reich, C. I., Sharma, S., Weisbaum, J. S., Wilson, B. A., and Olsen, G. J. (2008). Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl. Environ. Microbiol.* 74, 2461–2470. doi: 10.1128/AEM.02272-07

Greene, J. C., and Vermillion, J. R. (1964). The Simplified Oral Hygiene Index. *J. Am. Dent. Assoc.* 68, 7–13. doi: 10.14219/jada.archive.1964.0034

Grevich, S., Lee, P., Leroux, B., Ringold, S., Darveau, R., Henstorf, G., et al. (2019). Oral health and plaque microbial profile in juvenile idiopathic arthritis. *Pediatr. Rheumatol. Online J.* 17, 81. doi: 10.1186/s12969-019-0387-5

Hall, M. W., Singh, N., Ng, K. F., Lam, D. K., Goldberg, M. B., Tenenbaum, H. C., et al. (2017). Inter-personal diversity and temporal dynamics of dental, tongue, and salivary microbiota in the healthy oral cavity. *NPJ Biofilms Microbiomes* 3, 2. doi: 10.1038/s41522-016-0011-0

Hansen, T. H., Kern, T., Bak, E. G., Kashani, A., Allin, K. H., Nielsen, T., et al. (2018). Impact of a vegan diet on the human salivary microbiota. *Sci. Rep.* 8, 5847. doi: 10.1038/s41598-018-24207-3

Hiranmayi, K. V., Sirisha, K., Ramoji Rao, M. V., and Sudhakar, P. (2017). Novel Pathogens in Periodontal Microbiology. *J. Pharm. Bioallied Sci.* 9, 155–163. doi: 10.4103/jpbs.JPBS_288_16

Horton, D. B., Scott, F. I., Haynes, K., Putt, M. E., Rose, C. D., Lewis, J. D., et al. (2015). Antibiotic Exposure and Juvenile Idiopathic Arthritis: A Case-Control Study. *Pediatrics* 136, e333–e343. doi: 10.1542/peds.2015-0036

Kobus, A., Kierklo, A., Zalewska, A., Kuzmiuk, A., Szajda, S. D., Lawicki, S., et al. (2017). Unstimulated salivary flow, pH, proteins and oral health in patients with Juvenile Idiopathic Arthritis. *BMC Oral. Health* 17, 94. doi: 10.1186/s12903-017-0386-1

Lackner, J., Weiss, M., Muller-Graf, C., and Greiner, M. (2019). The disease burden associated with Campylobacter spp. in Germany 2014. *PloS One* 14, e0216867. doi: 10.1371/journal.pone.0216867

Lane, D. J., Pace, B., Olsen, G. J., Stahl, D. A., Sogin, M. L., and Pace, N. R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc. Natl. Acad. Sci. U.S.A.* 82, 6955–6959. doi: 10.1073/pnas.82.20.6955

Leksell, E., Ernberg, M., Magnusson, B., and Hedenberg-Magnusson, B. (2008). Intraoral condition in children with juvenile idiopathic arthritis compared to controls. *Int. J. Paediatr. Dent.* 18, 423–433. doi: 10.1111/j.1365-263X.2008.00931.x

Lim, Y., Totsika, M., Morrison, M., and Punyadeera, C. (2017). The saliva microbiome profiles are minimally affected by collection method or DNA extraction protocols. *Sci. Rep.* 7, 8523. doi: 10.1038/s41598-017-07885-3

Lorenzo, D., Gianvincenzo, Z., Carlo Luca, R., Karan, G., Jorge, V., Roberto, M., et al. (2019). Oral-Gut Microbiota and Arthritis: Is There an Evidence-Based Axis? *J. Clin. Med.* 8 (10). doi: 10.3390/jcm8101753

Majumder, S., and Aggarwal, A. (2020). Juvenile idiopathic arthritis and the gut microbiome: Where are we now? *Best Pract. Res. Clin. Rheumatol.* 33 (6), 101496. doi: 10.1016/j.berh.2020.101496

Malin, M., Verronen, P., Mykkanen, H., Salminen, S., and Isolauri, E. (1996). Increased bacterial urease activity in faeces in juvenile chronic arthritis: evidence of altered intestinal microflora? *Br. J. Rheumatol.* 35, 689–694. doi: 10.1093/rheumatology/35.7.689

Mascitti, M., Togni, L., Troiano, G., Caponio, V. C. A., Gissi, D. B., Montebugnoli, L., et al. (2019). Beyond Head and Neck Cancer: The Relationship Between Oral Microbiota and Tumour Development in Distant Organs. *Front. Cell Infect. Microbiol.* 9, 232. doi: 10.3389/fcimb.2019.00232

Maspero, C., Giannini, L., Galbiati, G., Prevedello, C., and Farronato, G. (2017). Periodontal conditions in juvenile idiopathic arthritis. *Minerva Stomatologica* 66, 43–50. doi: 10.23736/S0026-4970.17.03937-1

Miranda, L. A., Fischer, R. G., Sztajnbok, F. R., Figueredo, C. M., and Gustafsson, A. (2003). Periodontal conditions in patients with juvenile idiopathic arthritis. *J. Clin. Periodontol.* 30, 969–974. doi: 10.1034/j.1600-051X.2003.00406.x

Moe, N., and Rygg, M. (1998). Epidemiology of juvenile chronic arthritis in northern Norway: a ten-year retrospective study. *Clin. Exp. Rheumatol.* 16, 99–101.

Moen, K., Brun, J. G., Valen, M., Skartveit, L., Eribe, E. K., Olsen, I., et al. (2006). Synovial inflammation in active rheumatoid arthritis and psoriatic arthritis facilitates trapping of a variety of oral bacterial DNAs. *Clin. Exp. Rheumatol.* 24, 656–663.

Moore, W. E., Holdeman, L. V., Cato, E. P., Smibert, R. M., Burmeister, J. A., Palcanis, K. G., et al. (1985). Comparative bacteriology of juvenile periodontitis. *Infect. Immun.* 48, 507–519. doi: 10.1128/IAI.48.2.507-519.1985

Murakami, S., Mealey, B. L., Mariotti, A., and Chapple, I. L. C. (2018). Dental plaque-induced gingival conditions. *J. Periodontol.* 89 (Suppl 1), S17–S27. doi: 10.1111/jcpe.12937

Nordal, E., Zak, M., Aalto, K., Berntson, L., Fasth, A., Herlin, T., et al. (2011). Ongoing disease activity and changing categories in a long-term nordic cohort study of juvenile idiopathic arthritis. *Arthritis Rheum.* 63, 2809–2818. doi: 10.1002/art.30426

Petty, R. E., Southwood, T. R., Manners, P., Baum, J., Glass, D. N., Goldenberg, J., et al. (2004). International League of Associations for Rheumatology classification of juvenile idiopathic arthritis: second revision, Edmonton 2001. *J. Rheumatol.* 31, 390–392.

Pugliese, C., Van Der Vinne, R. T., Campos, L. M., Guardieiro, P. R., Saviolli, C., Bonfa, E., et al. (2016). Juvenile idiopathic arthritis activity and function ability: deleterious effects in periodontal disease? *Clin. Rheumatol.* 35, 81–91. doi: 10.1007/s10067-015-3125-5

Reichert, S., Machulla, H. K., Fuchs, C., John, V., Schaller, H. G., and Stein, J. (2006). Is there a relationship between juvenile idiopathic arthritis and periodontitis? *J. Clin. Periodontol.* 33, 317–323. doi: 10.1111/j.1600-051X.2006.00909.x

Rhayat, L., Maresca, M., Nicoletti, C., Perrier, J., Brinch, K. S., Christian, S., et al. (2019). Effect of Bacillus subtilis Strains on Intestinal Barrier Function and Inflammatory Response. *Front. Immunol.* 10, 564. doi: 10.3389/fimmu.2019.00564

Rossi, O., Van Baarlen, P., and Wells, J. M. (2013). Host-recognition of pathogens and commensals in the mammalian intestine. *Curr. Top. Microbiol. Immunol.* 358, 291–321. doi: 10.1007/82_2011_191

Said, H. S., Suda, W., Nakagome, S., Chinen, H., Oshima, K., Kim, S., et al. (2014). Dysbiosis of salivary microbiota in inflammatory bowel disease and its association with oral immunological biomarkers. *DNA Res.* 21, 15–25. doi: 10.1093/dnares/dst037

Santos, D., Silva, C., and Silva, M. (2015). Oral health and quality of life of children and adolescents with juvenile idiopathic arthritis according to their caregivers' perceptions. *Spec. Care Dentist* 35, 272–278. doi: 10.1111/scd.12129

Savioli, C., Silva, C. A., Ching, L. H., Campos, L. M., Prado, E. F., and Siqueira, J. T. (2004). Dental and facial characteristics of patients with juvenile idiopathic arthritis. *Rev. Hosp. Clin. Fac. Med. Sao Paulo* 59, 93–98. doi: 10.1590/S0041-87812004000300001

Scher, J. U., and Abramson, S. B. (2011). The microbiome and rheumatoid arthritis. *Nat. Rev. Rheumatol.* 7, 569–578. doi: 10.1038/nrrheum.2011.121

Scher, J. U., Ubeda, C., Equinda, M., Khanin, R., Buischi, Y., Viale, A., et al. (2012). Periodontal disease and the oral microbiota in new-onset rheumatoid arthritis. *Arthritis Rheum.* 64, 3083–3094. doi: 10.1002/art.34539

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., et al. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol.* 12 (6), R60. doi: 10.1186/gb-2011-12-6-r60

Skeie, M. S., Gil, E. G., Cetrelli, L., Rosen, A., Fischer, J., Astrom, A. N., et al. (2019). Oral health in children and adolescents with juvenile idiopathic arthritis - a systematic review and meta-analysis. *BMC Oral. Health* 19, 285. doi: 10.1186/s12903-019-0965-4

Stoll, M. L., Kumar, R., Morrow, C. D., Lefkowitz, E. J., Cui, X., Genin, A., et al. (2014). Altered microbiota associated with abnormal humoral immune responses to commensal organisms in enthesitis-related arthritis. *Arthritis Res. Ther.* 16, 486. doi: 10.1186/s13075-014-0486-0

Stoll, M. L., Kumar, R., Lefkowitz, E. J., Cron, R. Q., Morrow, C. D., and Barnes, S. (2016). Fecal metabolomics in pediatric spondyloarthritis implicate decreased metabolic diversity and altered tryptophan metabolism as pathogenic factors. *Genes Immun.* 17, 400–405. doi: 10.1038/gene.2016.38

Stoll, M. L., Weiss, P. F., Weiss, J. E., Nigrovic, P. A., Edelheit, B. S., Bridges, S. L.Jr., et al. (2018). Age and fecal microbial strain-specific differences in patients with spondyloarthritis. *Arthritis Res. Ther.* 20, 14. doi: 10.1186/s13075-018-1510-6

Stoustrup, P., Herlin, T., Spiegel, L., Rahimi, H., Koos, B., Pedersen, T. K., et al. (2019). Standardizing the Clinical Orofacial Examination in Juvenile Idiopathic Arthritis: An Interdisciplinary, Consensus-based, Short Screening Protocol. *J. Rheumatol.* 47 (9), 1397–1404. doi: 10.3899/jrheum.190661

Szymula, A., Rosenthal, J., Szczerba, B. M., Bagavant, H., Fu, S. M., and Deshmukh, U. S. (2014). T cell epitope mimicry between Sjogren's syndrome Antigen A (SSA)/Ro60 and oral, gut, skin and vaginal bacteria. *Clin. Immunol.* 152, 1–9. doi: 10.1016/j.clim.2014.02.004

Tejesvi, M. V., Arvonen, M., Kangas, S. M., Keskitalo, P. L., Pirttila, A. M., Karttunen, T. J., et al. (2016). Faecal microbiome in new-onset juvenile idiopathic arthritis. *Eur. J. Clin. Microbiol. Infect. Dis.* 35, 363–370. doi: 10.1007/s10096-015-2548-x

Tong, Y., Zheng, L., Qing, P., Zhao, H., Li, Y., Su, L., et al. (2019). Oral Microbiota Perturbations Are Linked to High Risk for Rheumatoid Arthritis. *Front. Cell Infect. Microbiol.* 9, 475. doi: 10.3389/fcimb.2019.00475

Trombelli, L., Farina, R., Silva, C. O., and Tatakis, D. N. (2018). Plaque-induced gingivitis: Case definition and diagnostic considerations. *J. Periodontol.* 89 (Suppl 1), S46–S73. doi: 10.1002/JPER.17-0576

Turnbaugh, P. J., Ridaura, V. K., Faith, J. J., Rey, F. E., Knight, R., and Gordon, J. I. (2009). The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci. Transl. Med.* 1, 6ra14. doi: 10.1126/scitranslmed.3000322

Van Dijkhuizen, E. H. P., Del Chierico, F., Malattia, C., Russo, A., Pires Marafon, D., Ter Haar, N. M., et al. (2019). Microbiome Analytics of the Gut Microbiota in Patients With Juvenile Idiopathic Arthritis: A Longitudinal Observational Cohort Study. *Arthritis Rheumatol.* 71, 1000–1010. doi: 10.1002/art.40827

Vogtmann, E., Chen, J., Kibriya, M. G., Amir, A., Shi, J., Chen, Y., et al. (2019). Comparison of Oral Collection Methods for Studies of Microbiota. *Cancer Epidemiol. Biomarkers Prev.* 28, 137–143. doi: 10.1158/1055-9965.EPI-18-0312

Wallace, C. A., Giannini, E. H., Huang, B., Itert, L., Ruperto, N.Childhood Arthritis Rheumatology Research, A., et al. (2011). American College of Rheumatology provisional criteria for defining clinical inactive disease in select categories of juvenile idiopathic arthritis. *Arthritis Care Res. (Hoboken)* 63, 929–936. doi: 10.1002/acr.20497

Wegner, N., Lundberg, K., Kinloch, A., Fisher, B., Malmstrom, V., Feldmann, M., et al. (2010a). Autoimmunity to specific citrullinated proteins gives the first clues to the etiology of rheumatoid arthritis. *Immunol. Rev.* 233, 34–54. doi: 10.1111/j.0105-2896.2009.00850.x

Wegner, N., Wait, R., Sroka, A., Eick, S., Nguyen, K. A., Lundberg, K., et al. (2010b). Peptidylarginine deiminase from Porphyromonas gingivalis citrullinates human fibrinogen and alpha-enolase: implications for autoimmunity in rheumatoid arthritis. *Arthritis Rheum.* 62, 2662–2672. doi: 10.1002/art.27552

Welbury, R. R., Thomason, J. M., Fitzgerald, J. L., Steen, I. N., Marshall, N. J., and Foster, H. E. (2003). Increased prevalence of dental caries and poor oral hygiene in juvenile idiopathic arthritis. *Rheumatol. (Oxford)* 42, 1445–1451. doi: 10.1093/rheumatology/keg395

Xu, Y., Jia, Y. H., Chen, L., Huang, W. M., and Yang, D. Q. (2018). Metagenomic analysis of oral microbiome in young children aged 6-8 years living in a rural isolated Chinese province. *Oral. Dis.* 24, 1115–1125. doi: 10.1111/odi.12871

Zhang, X., Zhang, D., Jia, H., Feng, Q., Wang, D., Liang, D., et al. (2015). The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nat. Med.* 21, 895–905. doi: 10.1038/nm.3914

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cellular and Infection Microbiology

ORIGINAL RESEARCH
published: 23 November 2020
doi: 10.3389/fcimb.2020.600884

Compositional Shift of Oral Microbiota Following Surgical Resection of Tongue Cancer

Shinya Kageyama[1†], Yuka Nagao[2†], Jiale Ma[1], Mikari Asakawa[1], Ryoji Yoshida[2], Toru Takeshita[1,3], Akiyuki Hirosue[2*], Yoshihisa Yamashita[1] and Hideki Nakayama[2]*

[1] Section of Preventive and Public Health Dentistry, Division of Oral Health, Growth and Development, Faculty of Dental Science, Kyushu University, Fukuoka, Japan, [2] Department of Oral and Maxillofacial Surgery, Faculty of Life Sciences, Kumamoto University, Kumamoto, Japan, [3] OBT Research Center, Faculty of Dental Science, Kyushu University, Fukuoka, Japan

OPEN ACCESS

Edited by:
Naile Dame-Teixeira,
University of Brasilia, Brazil

Reviewed by:
Eliete Neves Da Silva Guerra,
University of Brasilia, Brazil
J. Christopher Fenno,
University of Michigan, United States
Daniela Jorge Corralo,
The University of Passo Fundo, Brazil

***Correspondence:**
Toru Takeshita
taketooo@dent.kyushu-u.ac.jp
Akiyuki Hirosue
ahiro711@kumamoto-u.ac.jp

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Microbiome in Health and Disease, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: 31 August 2020
Accepted: 21 October 2020
Published: 23 November 2020

Citation:
Kageyama S, Nagao Y, Ma J, Asakawa M, Yoshida R, Takeshita T, Hirosue A, Yamashita Y and Nakayama H (2020) Compositional Shift of Oral Microbiota Following Surgical Resection of Tongue Cancer. Front. Cell. Infect. Microbiol. 10:600884. doi: 10.3389/fcimb.2020.600884

Salivary microbiota is considered a source of microorganisms for the respiratory and digestive tracts, and a trigger for diseases in these distant organs. Meanwhile, the microbiota on the tongue surface is thought to be a major source of salivary microbiota. Therefore, surgical resection of the tongue for definitive treatment of oral cancer could drastically change the salivary bacterial balance and virulence. Here, we investigated the shift of the salivary microbiota following surgical resection in patients with tongue cancer. The stimulated saliva samples were collected from 25 tongue cancer patients pre- and post-resection of the tongue, and bacterial density and composition was determined using quantitative PCR analysis and 16S ribosomal RNA (rRNA) gene sequencing, respectively. Although no significant difference in the total bacterial density in saliva pre- and post-surgery was observed, the bacterial composition significantly differed according to the analysis of similarity. Among predominant operational taxonomic units (OTUs) with ≥1% of relative abundance, the proportions of OTUs corresponding to *Streptococcus salivarius*, *Prevotella melaninogenica*, and *Prevotella histicola* were significantly decreased following the tongue resection. On the other hand, the proportions of OTUs corresponding to *Lautropia mirabilis*, *Neisseria flava*, *Streptococcus sanguinis*, and *Fusobacterium nucleatum*, known to be inhabitants of dental plaque, were significantly increased. These results suggest that surgical resection of the tongue causes a compositional shift of the salivary microbiota, characterized by an increase in bacterial species derived from dental plaque, including periodontal pathogens. These results suggest the necessity of more careful and frequent postoperative oral care after surgical resection of tongue cancer.

Keywords: saliva, microbiome, 16S ribosomal RNA, next-generation sequencing, quantitative real-time PCR, tongue neoplasms, oral surgery, glossectomy

INTRODUCTION

Saliva is secreted from the salivary glands into the oral cavity and contains various oral debris including desquamated epithelial cells, food residue, and dense oral bacteria. Since, saliva is swallowed constantly, the salivary microbiota is considered as a source of microorganisms to respiratory and digestive tracts and a trigger of diseases in these distant organs. Recent studies suggested the association of oral bacteria with diseases occurring in organs far from oral cavity such as pneumonia, colorectal cancer, and inflammatory bowel disease (Gevers et al., 2014; Segal et al., 2016; Huffnagle et al., 2017; Kageyama et al., 2018; Kageyama et al., 2019; Schmidt et al., 2019; Yachida et al., 2019). The salivary microbiota is a mixture of bacteria shed from various oral niches, such as tongue dorsum, tooth surface, gingival crevice, and buccal mucosa. Among them, tongue microbiota is thought of a major source of salivary microbiota as the bacterial composition in saliva resembles that on tongue dorsum (Mager et al., 2003; Segata et al., 2012; Kageyama et al., 2017).

Cancer is a serious global health problem with a high mortality risk (Siegel et al., 2019). In Japan, the cancer-related deaths account for a quarter of all causes and are the leading causes of mortality for both sexes (Ministry of Health, Labour and Welfare, 2019). Oral cancer, predominantly oral squamous cell carcinoma, generally appears on the tongue, gingiva, floor of the mouth, and palate. Approximately 10,000 Japanese are diagnosed with oral cancer annually (Cancer Information Service, National Cancer Center Japan, 2020). Among them, tongue cancer is the most common malignancy and accounts for approximately 50% of the cases (Cancer Information Service, National Cancer Center Japan, 2020). A large number of epidemiological studies have demonstrated that lifestyle factors, such as smoking, alcohol intake, underweight, and low consumption of vegetables and fruits, are associated with oral cancer (Bosetti et al., 2000; Huang et al., 2003; Varela-Lema et al., 2010; Radoï et al., 2015). Regarding treatment, although internal radiotherapy is performed for T1–2 or superficial T3 cancers or concurrent chemoradiotherapy for advanced cancer, surgical resection of tumors is the most well-established definitive approach to oral cancer (Shah and Gil, 2009; Nibu et al., 2017; Colevas et al., 2018). Surgery reduces the area of the tongue surface with complex papillary structures, and the defect is reconstructed by foreign tissue in some cases. Therefore, the salivary bacterial balance and virulence could drastically change following surgical resection of tongue cancer.

In this study, we examined the salivary microbiota collected from patients with tongue cancer pre- and post-definitive surgery. We compared their bacterial density and composition using quantitative PCR analysis and 16S ribosomal RNA (16S rRNA) gene amplicon sequencing and confirmed how bacterial density and composition of salivary microbiota changed following tongue resection. This study aimed to characterize the bacterial shift before and after surgery and identify bacterial species that show drastic changes.

MATERIALS AND METHODS

Study Subjects and Sample Collection

Study subjects of this study were patients with tongue cancer who visited Kumamoto University Hospital, Japan. A total of 53 tongue cancer patients were enrolled at the preoperative hospitalization from November 2017 to April 2019. They were diagnosed based on the histological and radiological findings, including computed tomography (CT), magnetic resonance imaging, ultrasonography, and positron emission tomography-computed tomography (PET-CT) findings. The staging of their tumors was performed according to TNM classification of the AJCC eighth edition (Amin et al., 2017). During the preoperative hospitalization, stimulated saliva samples were collected from the subjects (8.3 ± 9.4 days before cancer treatment). We instructed the subjects to chew gums and spew the whole saliva into sterile plastic tubes. After finishing the postoperative nasogastric tube feeding and starting to consume a diet orally, a similar procedure was followed to collect stimulated saliva (21.2 ± 10.9 days after cancer treatment). All post-treatment samples were collected prior to postoperative radiotherapy and chemotherapy. The samples were stored at −80°C until further analysis. After excluding 28 subjects who did not receive surgical resection (n=3), did not have both pre- and post-treatment samples (n=22), and had less than 7 teeth (n=3), 50 samples from 25 subjects were finally examined. Written informed consent was obtained from all participants. The ethics committee of Kumamoto University approved this study with the informed consent procedure (approval number 1427, 1928, and 2389).

Quantitative PCR Analysis of Total Bacterial Density in Saliva

The collected saliva samples were subjected to quantitative PCR analysis of total bacterial density. DNA was extracted from each sample using the bead-beating method (Yamanaka et al., 2012), and quantitative PCR was performed using a QuantiFast SYBR Green PCR Kit (QIAGEN, Hilden, Germany) in QuantStudio 3 (Thermo Fisher Scientific, MA, USA) according to the manufacturer's instructions. The primers 806F (5'-TTA GAT ACC CYG GTA GTC C-3') and 926R (5'-CCG TCA ATT YCT TTG AGT TT-3'), with target V5 regions of the 16S ribosomal RNA (rRNA) gene, were used for the quantification of the total bacterial density (Asakawa et al., 2018).

16S Ribosomal RNA Gene Amplicon Sequencing of Saliva

The V1–V2 regions of 16S rRNA gene were amplified using the following primers: 8F (5′-AGA GTT TGA TYM TGG CTC AG-3′) with the Ion Torrent adapter A and the sample-specific 8-base tag sequence and 338R (5′-TGC TGC CTC CCG TAG GAG T-3′) with the Ion Torrent trP1 adapter sequence. PCR amplification, purification, and quantification of each PCR amplicon was performed as described previously (Takeshita et al., 2016). The purified PCR amplicons were pooled, and gel-purification was performed using Wizard SV Gel and PCR Clean-Up System (Promega, WI, USA). The DNA concentration was determined

using a KAPA Library Quantification Kit (KAPA Biosystems, MA, USA) and the DNA was diluted for use as the template DNA in emulsion PCR. Emulsion PCR and enrichment of template-positive particles were performed using Ion PGM Template Hi-Q View OT2 Kit (Thermo Fisher Scientific) in Ion One Touch 2 System (Thermo Fisher Scientific). The enriched particle was loaded onto an Ion 318 v2 chip (Thermo Fisher Scientific) and sequencing was performed on the Ion PGM (Thermo Fisher Scientific) using Ion PGM Hi-Q View Sequencing Kit (Thermo Fisher Scientific).

Data Analysis and Taxonomy Assignment

The quality filtering of raw sequence reads using a script written in R (version 3.6.2) was carried out. The reads were excluded from the analysis when they exhibited ≤200 bases, or had an average quality score ≤25, or did not include the correct forward, or the correct reverse primer sequence (one mismatch was allowed) or had a homopolymer of >6 nucleotides. The quality-checked reads were demultiplexed by examining the eight-base tag sequence, and then forward and reverse primer sequences were trimmed. Operational taxonomic units (OTUs) were constructed by clustering quality-checked reads, excluding singleton reads, with a minimum pairwise identity of 97% using UPARSE (Edgar, 2013) as described previously (Takeshita et al., 2016). All quality-checked reads were mapped to each OTU with ≥97% identity using UPARSE (Takeshita et al., 2016). Chimeras were identified using ChimeraSlayer and removed from analysis (Haas et al., 2011). The taxonomy of representative sequences was determined using BLAST against 889 oral bacterial 16S rRNA gene sequences (HOMD 16S rRNA RefSeq version 14.51) in the Human Oral Microbiome Database (Chen et al., 2010). Nearest-neighbor species with ≥98.5% identity was selected as candidates for each representative OTU. The taxonomy of sequences without hits were further determined using RDP classifier with a minimum support threshold of 80% (Wang et al., 2007). The number of OTUs and UniFrac distance were calculated following rarefaction to 5000 reads/sample using R. The sequence data have been deposited in DDBJ Sequence Read Archive under accession number DRA010919.

Statistical Analysis

The bacterial characteristics of subjects pre- and post-tongue resection were compared. The total bacterial densities and diversities were compared using Wilcoxon signed-rank test for comparison of paired samples. The UniFrac metric was used to determine the dissimilarity between bacterial compositions (Lozupone and Knight, 2005). The dissimilarity between groups was evaluated using the analysis of similarities (ANOSIM) with 999 permutations based on the weighted UniFrac distance. Relative abundances of predominant genera were compared using Wilcoxon signed-rank test and obtained P-values were adjusted using a Benjamini-Hochberg false discovery rate (FDR) correction for multiple testing. The detection of discriminant bacterial species was also performed using Wilcoxon signed-rank test and FDR correction. Two-sided $P < 0.05$ indicated statistical significance. All statistical analyses were performed using R.

RESULTS

The Characteristics of Subjects and Salivary Microbiota Sequence

A total of 25 patients with tongue cancer (16 males and 9 females, age 24–93 years old) were enrolled in the present study. The detailed characteristics of study subjects are presented in **Table 1**. Histologically, all their cancers were squamous cell carcinoma. Most of their clinical tumor (cT) stages were cT1–2 (84%) and clinical nodal (cN) stages were cN0 (92%). Distant metastasis was clinically not observed. Of the 25 subjects, 21 were administered antibiotics for biopsy (mainly amoxicillin, a beta-lactam antibiotic, n=19), and all subjects were administered antibiotics for surgery (mainly cefmetazole, a second-generation cephalosporin, n=23). Pre- and post-treatment sampling was performed on an average of 26.5 ± 9.9 and 18.6 ± 9.3 days after antibiotics exposure, respectively. Analysis of 50 stimulated saliva samples by 16S rRNA gene amplicon analysis was carried out, and 461,830 high-quality reads (9,237 ± 1,651 reads per sample) were obtained to determine their bacterial diversity and composition.

Shift of Total Bacterial Density in Saliva Following the Surgical Resection

A quantitative PCR analysis was performed to evaluate the effect of the surgical resection on the total bacterial density of salivary microbiota. As shown in **Table 2**, there was no significant difference in pre- and post-treatment samples ($P = 0.35$).

TABLE 1 | The clinical characteristics of study subjects.

Clinical characteristics		
Age (years), mean ± SD		64.6 ± 18.2
Sex, n (%)	Male	16 (64.0)
	Female	9 (36.0)
Number of teeth, mean ± SD		20.6 ± 7.2
Smoking habit, n (%)	Current	8 (32.0)
	Non-current	17 (68.0)
Alcohol consumption, n (%)	Everyday	10 (40.0)
	Non-everyday	15 (60.0)
cT-stage, n (%)	1	11 (44.0)
	2	10 (40.0)
	3	4 (16.0)
cN-stage, n (%)	0	23 (92.0)
	2b	2 (8.0)
Antibiotics for biopsy, n (%)	None	3 (12.0)
	AMPC	19 (76.0)
	CDTR-PI	1 (4.0)
	CFPN-PI	1 (4.0)
Antibiotics for surgery, n (%)	CMZ	23 (92.0)
	CEZ and CTRX	1 (4.0)
	CMZ and SBT/ABPC	1 (4.0)
Reconstructive surgery, n (%)	Yes	6 (24.0)
	No	19 (76.0)

AMPC, amoxicillin; CDTR-PI, cefditoren pivoxil; CFPN-PI, cefcapene pivoxil; CMZ, cefmetazole; CEZ, cefazolin; CTRX, ceftriaxone; SBT/ABPC, sulbactam/ampicillin; cT-stage, clinical tumor stage; cN-stage, clinical nodal stage; SD, standard deviation.

Shift of Bacterial Diversity and Bacterial Composition in Saliva Following the Surgical Resection

The collected saliva samples were examined using 16S rRNA gene sequencing to evaluate the effect of the surgical resection on the bacterial balance of salivary microbiota. The post-treatment saliva exhibited significantly lower bacterial diversity than the pre-treatment saliva according to observed number of OTUs (P = 0.01, **Table 2**). **Figure 1** presents a principal coordinate analysis (PCoA) plot based on the weighted UniFrac distances. According to ANOSIM, there was a significant difference in bacterial composition of salivary microbiota pre- and post-surgery (P = 0.001). On confirmation of the bacterial composition of their salivary microbiota at genus level, 18 predominant genera with ≥1% of the relative abundance accounted for 93.4 ± 4.4 and 93.3 ± 5.3% in pre- and post-treatment samples, respectively. Among them, post-surgical resection, *Streptococcus*, *Prevotella*, *Gemella*, and *Leptotrichia* were significantly decreased while, *Neisseria*, *Fusobacterium*, and *Lautropia* were significantly increased.

TABLE 2 | Bacterial density and diversity in pre- and post-treatment samples.

	Pre-treatment (n=25)	Post-treatment (n=25)	P value
Bacterial density (log copies/ml), mean ± SD	9.85 ± 0.41	9.96 ± 0.33	0.35
Bacterial diversity			
Number of OTU, mean ± SD	147.5 ± 31.1	132.6 ± 22.9	0.015
Shannon index, mean ± SD	3.4 ± 0.35	3.4 ± 0.30	0.58

SD, standard deviation.

Bacterial Species Showing Drastic Compositional Shift Following the Surgical Resection

To find bacterial species that were drastically increased or decreased following the tongue resection, discriminant OTUs in pre- and post-treatment samples were identified using the Wilcoxon signed-rank test. Among predominant OTUs with ≥1% of the relative abundance, the analysis revealed nine OTUs were differentially abundant in pre- and post-treatment samples (**Figure 2**). Following the surgical resection, OTUs corresponding to *Streptococcus salivarius* HOT-755, *Prevotella melaninogenica* HOT-469, *Prevotella histicola* HOT-298, *Gemella morbillorum* HOT-046, and *Actinomyces* species HOT172 were significantly decreased,

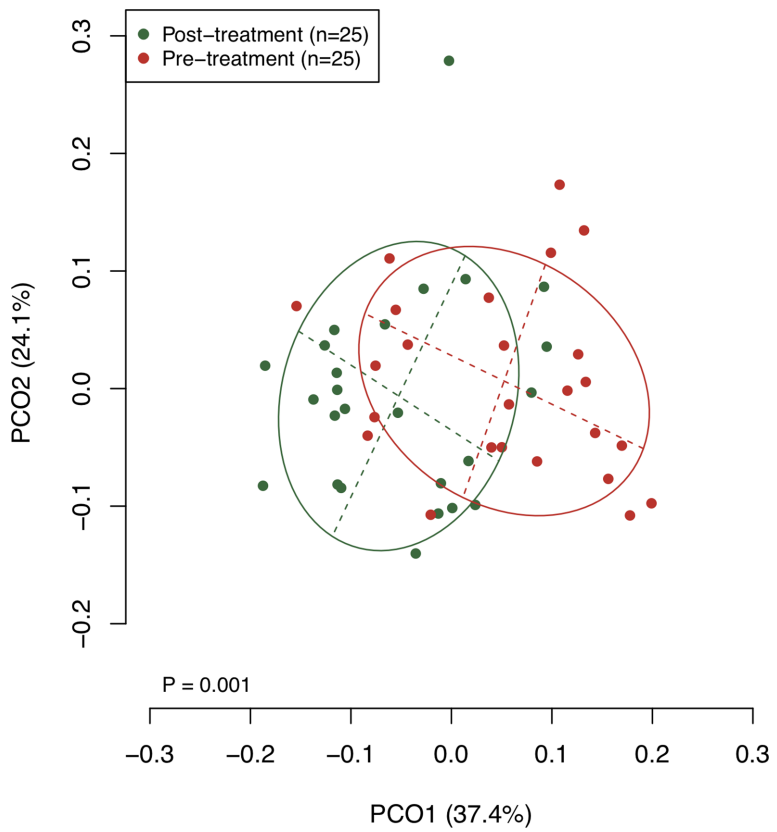


FIGURE 1 | A principal coordinate analysis (PCoA) based on weighted UniFrac distance. The bacterial composition of pre- and post-treatment samples are depicted using different colors. These two components explain the 61.5% variance. The intersection of the broken lines indicates the center of gravity for each group. The ellipse covers 67% of the samples belonging to each group.

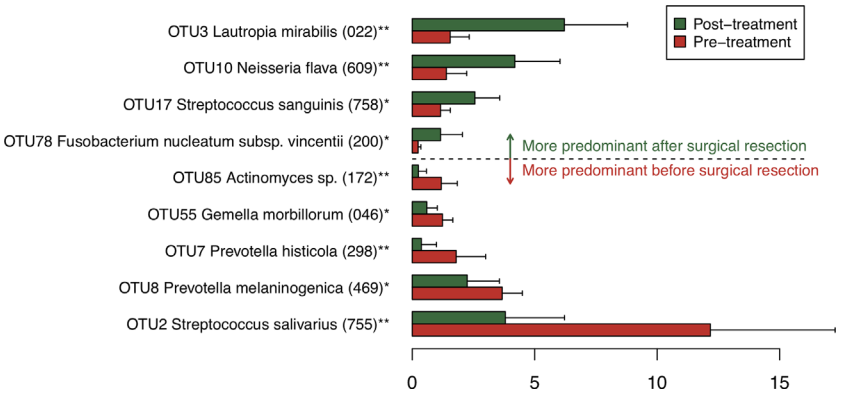


FIGURE 2 | Bacterial species corresponding to the differentially abundant operational taxonomic units (OTUs) between pre- and post-treatment samples. Bar plots show mean relative abundances of differentially abundant OTUs. Only nine OTUs with ≥1% of the relative abundance and significant difference were shown. The bar plots of pre- and post-treatment samples are depicted using different colors. Error bars indicate 95% confidence intervals. Oral taxon IDs are given in parentheses following bacterial names. **P < 0.01. *P < 0.05.

whereas OTUs corresponding to *Lautropia mirabilis* HOT-022, *Neisseria flava* HOT-609, *Streptococcus sanguinis* HOT-758, and *Fusobacterium nucleatum* HOT-200 were significantly increased.

DISCUSSION

The present study demonstrated that following surgical resection of tongue cancer, the bacterial diversity decreased, and the bacterial balance of predominant bacteria in salivary microbiota shifted. In post-treatment samples, *L. mirabilis*, *N. flava*, *S. sanguinis*, and *F. nucleatum* were significantly increased compared to that in the pretreatment samples, and they were all oral indigenous bacteria. This result suggests that surgical resection causes a balance shift of salivary microbiota, but not drastic and terrible oral dysbiosis, such as overgrowth of non-indigenous pathogenic bacteria. On the other hand, *F. nucleatum*, a periodontal pathogen, increased in salivary microbiota following tongue resection. *F. nucleatum* generally inhabits subgingival plaque and has periodontopathogenic properties, such as activation of inflammatory cytokines that lead to periodontal attachment and tissue damage (Baqui et al., 1998). In addition, *F. nucleatum* is considered an opportunistic pathogen implicated in the carcinogenesis of oral cancer, not only periodontitis (Gholizadeh et al., 2016). Interestingly, although it generally inhabits the oral cavity, several studies suggest its association with gastrointestinal diseases, such as inflammatory bowel disease and colorectal cancer by inducing inflammation and downregulating host immunity (Strauss et al., 2011; Mima et al., 2015; Nosho et al., 2016; Wu et al., 2019). In fact, *F. nucleatum* is frequently and abundantly detected in colorectal tissue from patients with these diseases (Strauss et al., 2011; Mima et al., 2015; Nosho et al., 2016; Yachida et al., 2019). As swallowing function often declines after tongue resection, assessment of salivary microbiota in postoperative patients might help in preventing the development of associated gastrointestinal diseases as well as oral diseases.

In the discriminant analysis, the bacterial species, *L. mirabilis*, *N. flava*, *S. sanguinis*, and *F. nucleatum*, which mainly inhabit dental plaque, were increased after resection of the tongue (Takeshita et al., 2015; Kageyama et al., 2017; Ihara et al., 2019). On the other hand, *S. salivarius*, *P. melaninogenica*, *P. histicola*, and *Actinomyces* species, known as predominant species in tongue dorsum, were identified as decreasing species (Kageyama et al., 2017; Asakawa et al., 2018). Although the preferred habitat of *G. morbillorum* remains unclear, all other discriminant bacteria demonstrated a common tendency: an increase in dental plaque bacteria and a decrease in tongue bacteria. These results suggest that oral environment post-surgical resection alters by reduction in surface area of tongue, and dental plaque-derived bacteria become dominant in salivary microbiota. These findings are in accordance to the previously proposed concept of tongue microbiota being a major source of salivary microbiota.

Of the 25 subjects who underwent surgical resection of tongue cancer, reconstructive surgery was performed in six subjects (4 men and 2 women), and their surgical defects were reconstructed using a cervical island skin flap (n=3), pectoralis major myocutaneous flap (n=2), and skin graft (n=1). The tongue dorsum is lined by stratified squamous epithelium with numerous tongue papillae, which provide an anaerobic environment for harboring diverse anaerobic bacteria. The complex papillary structure also retains blood serum components, infusion from the gingival crevice, epithelial cells, and food residue on the tongue surface as tongue coating and provide a nutrient-rich environment for tongue bacteria. On the other hand, the skin graft and flap are lined by the stratified squamous epithelium with thick keratinized layer, and the surface structure is relatively smooth. Thus, it is reasonable to assume that this difference changes the bacterial species on the tongue surface and consequently changes the bacterial composition of the salivary microbiota. However, there was no significant difference in the bacterial composition of post-treatment saliva with or without reconstructive surgery (data not shown). In this study, saliva samples were collected at 21.5 ± 10.5 days post-surgery to identify the effect of tongue resection. The

formation of tongue coating and tongue microbiota in the long term may differ on the original tongue surface or the grafted surface, and the results would be fundamental for the development of a novel approach to control the virulence of oral microbiota. Further studies are required to elucidate the long-term effects of the skin graft and flap on salivary and tongue microbiota.

This study has several potential limitations. Firstly, most subjects of the present study were administered antibiotics for biopsy and surgery prior to pre-treatment and post-treatment sampling, respectively. Of the 25 subjects, 21 subjects were administered antibiotics for biopsy and all subjects were administered antibiotics for surgery. However, previous report indicated that bacterial composition of the salivary microbiota was stable against antibiotic treatment compared to feces microbiota, and impacts of antibiotics including ciprofloxacin, amoxicillin, and minocycline on the salivary microbiota composition were lost by 1 week to 1 month post-exposure (Zaura et al., 2015). Another report also suggested the impact of amoxicillin on salivary bacteria peaked out 4 h after exposure (Larsson Wexell et al., 2016). In this study, the pre- and post-treatment sampling were performed on an average of 26.5 ± 9.9 and 18.6 ± 9.3 days after antibiotics exposure, respectively. Although most samples were collected within a month after administration of antibiotics, it was unlikely that the effect of antibiotics remained robust, especially for a selective increase of dental plaque-derived bacteria. Secondly, although oral conditions such as number of teeth, periodontal conditions, and dental caries status affect the bacterial diversity and composition of salivary microbiota, however, the data about dental examination including periodontal pocket depth, bleeding on probing, decayed teeth, and oral hygiene status was not collected, except number of teeth from panoramic radiographs. This limited the thorough understanding of the effect of tongue resection on the salivary microbiota. However, considering dental caries and periodontitis gradually progress, there was probably no drastic change in periodontal pocket depth and dental caries status between samplings. It is also speculated that oral hygiene status was unchanged because of the stable bacterial density in saliva before and after the surgery (**Table 2**). Lastly, the sample size of the present study, especially those who underwent reconstructive surgery, was small. Although smoking is known as a lifestyle factor affecting the oral microbiota (Takeshita et al., 2016), a significant difference by smoking was not observed in the bacterial composition of pretreatment samples, post-treatment samples, and a bacterial composition shift following the surgery. Similar results were obtained for alcohol consumption. In addition, although *F. nucleatum* increased in the salivary microbiota following tongue resection, the shift in the overall virulence of salivary microbiota remains unclear. Further studies with larger sample sizes are required to identify the effects of surgical resection and reconstructive surgery on the bacterial composition and the virulence of oral microbiota as well as the effects of lifestyle factors on oral microbiota in the perioperative period.

In conclusion, surgical resection of tongue cancer causes a shift in the bacterial diversity and composition of salivary microbiota characterized by an increase in bacterial species derived from dental plaque. In particular, *F. nucleatum* is a periodontal pathogen and is suspected to be associated with oral and gastrointestinal diseases. These results might suggest the necessity of more careful and frequent postoperative oral care after surgical resection of tongue cancer.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ncbi.nlm.nih.gov/sra/, accession no: DRA010919.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The ethics committee of Kumamoto University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SK wrote the first draft of the manuscript. SK, YN, RY, AH, TT, YY, and HN critically revised the manuscript. YN, RY, AH, and HN collected the clinical data and sample. YN, JM, and MA performed the molecular analysis. SK, JM, MA, and TT performed the bioinformatics, and statistical analysis. SK, YN, AH, TT, YY, and HN contributed to the conception and design of the study. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by JSPS KAKENHI Grant Numbers JP20K18808, JP20H03901, JP19K22722, JP19H03863, JP18H03005, and JP18K09727.

REFERENCES

Amin, M. B., Edge, S., Greene, F., Byrd, D. R., Brookland, R. K., Washington, M. K., et al. (2017). *AJCC Cancer Staging Manual 8th edition* (New York: Springer).

Asakawa, M., Takeshita, T., Furuta, M., Kageyama, S., Takeuchi, K., Hata, J., et al. (2018). Tongue Microbiota and Oral Health Status in Community-Dwelling Elderly Adults. *mSphere* 3, e00332–e00318. doi: 10.1128/mSphere.00332-18

Baqui, A. A., Meiller, T. F., Chon, J. J., Turng, B. F., and Falkler, W. A. (1998). Interleukin-6 production by human monocytes treated with granulocyte-macrophage colony-stimulating factor in the presence of lipopolysaccharide of oral microorganisms. *Oral. Microbiol. Immunol.* 13, 173–180. doi: 10.1111/j.1399-302X.1998.tb00729.x

Bosetti, C., Negri, E., Franceschi, S., Conti, E., Levi, F., Tomei, F., et al. (2000). Risk factors for oral and pharyngeal cancer in women: a study from Italy and Switzerland. *Br. J. Cancer* 82, 204–207. doi: 10.1054/bjoc.1999.0900

Cancer Information Service and National Cancer Center Japan (2020). *Cancer Registry and Statistics* (Tokyo: National Cancer Center Japan).

Chen, T., Yu, W. H., Izard, J., Baranova, O. V., Lakshmanan, A., and Dewhirst, F. E. (2010). The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database* 2010, baq013. doi: 10.1093/database/baq013

Colevas, A. D., Yom, S. S., Pfister, D. G., Spencer, S., Adelstein, D., Adkins, D., et al. (2018). NCCN guidelines insights: Head and neck cancers, version 1.2018 featured updates to the NCCN guidelines. *J. Natl. Compr. Canc. Netw.* 16, 479–490. doi: 10.6004/jnccn.2018.0026

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* 10, 996–998. doi: 10.1038/nmeth.2604

Gevers, D., Kugathasan, S., Denson, L. A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., et al. (2014). The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* 15, 382–392. doi: 10.1016/j.chom.2014.02.005

Gholizadeh, P., Eslami, H., Yousefi, M., Asgharzadeh, M., Aghazadeh, M., and Kafil, H. S. (2016). Role of oral microbiome on oral cancers, a review. *Biomed. Pharmacother.* 84, 552–558. doi: 10.1016/j.biopha.2016.09.082

Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., et al. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.* 21, 494–504. doi: 10.1101/gr.112730.110

Huang, W. Y., Winn, D. M., Brown, L. M., Gridley, G., Bravo-Otero, E., Diehl, S. R., et al. (2003). Alcohol Concentration and Risk of Oral Cancer in Puerto Rico. *Am. J. Epidemiol.* 157, 881–887. doi: 10.1093/aje/kwg055

Huffnagle, G. B., Dickson, R. P., and Lukacs, N. W. (2017). The respiratory tract microbiome and lung inflammation: A two-way street. *Mucosal Immunol.* 10, 299–306. doi: 10.1038/mi.2016.108

Ihara, Y., Takeshita, T., Kageyama, S., Matsumi, R., Asakawa, M., Shibata, Y., et al. (2019). Identification of Initial Colonizing Bacteria in Dental Plaques from Young Adults Using Full-Length 16S rRNA Gene Sequencing. *mSystems* 4, e00360–e00319. doi: 10.1128/mSystems.00360-19

Kageyama, S., Takeshita, T., Asakawa, M., Shibata, Y., Takeuchi, K., Yamanaka, W., et al. (2017). Relative abundance of total subgingival plaque-specific bacteria in salivary microbiota reflects the overall periodontal condition in patients with periodontitis. *PloS One* 12, e0174782. doi: 10.1371/journal.pone.0174782

Kageyama, S., Takeshita, T., Furuta, M., Tomioka, M., Asakawa, M., Suma, S., et al. (2018). Relationships of Variations in the Tongue Microbiota and Pneumonia Mortality in Nursing Home Residents. *J. Gerontol. A Biol. Sci. Med. Sci.* 73, 1097–1102. doi: 10.1093/gerona/glx205

Kageyama, S., Takeshita, T., Takeuchi, K., Asakawa, M., Matsumi, R., Furuta, M., et al. (2019). Characteristics of the Salivary Microbiota in Patients With Various Digestive Tract Cancers. *Front. Microbiol.* 10, 1780. doi: 10.3389/fmicb.2019.01780

Larsson Wexell, C., Ryberg, H., Sjöberg Andersson, W. A., Blomqvist, S., Colin, P., Van Bocxlaer, J., et al. (2016). Antimicrobial Effect of a Single Dose of Amoxicillin on the Oral Microbiota. *Clin. Implant Dent. Relat. Res.* 18, 699–706. doi: 10.1111/cid.12357

Lozupone, C., and Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71, 8228–8235. doi: 10.1128/AEM.71.12.8228-8235.2005

Mager, D. L., Ximenez-Fyvie, L. A., Haffajee, A. D., and Socransky, S. S. (2003). Distribution of selected bacterial species on intraoral surfaces. *J. Clin. Periodontol.* 30, 644–654. doi: 10.1034/j.1600-051x.2003.00376.x

Mima, K., Sukawa, Y., Nishihara, R., Qian, Z. R., Yamauchi, M., Inamura, K., et al. (2015). Fusobacterium nucleatum and T Cells in Colorectal Carcinoma. *JAMA Oncol.* 1, 653–661. doi: 10.1001/jamaoncol.2015.1377

Ministry of Health, Labour and Welfare (2019). *Vital Statistics Japan* (Tokyo: Ministry of Health, Labour and Welfare).

Nibu, K., Hayashi, R., Asakage, T., Ojiri, H., Kimata, Y., Kodaira, T., et al. (2017). Japanese Clinical Practice Guideline for Head and Neck Cancer. *Auris Nasus Larynx* 44, 375–380. doi: 10.1016/j.anl.2017.02.004

Nosho, K., Sukawa, Y., Adachi, Y., Ito, M., Mitsuhashi, K., Kurihara, H., et al. (2016). Association of Fusobacterium nucleatum with immunity and molecular alterations in colorectal cancer. *World J. Gastroenterol.* 22, 557–566. doi: 10.3748/wjg.v22.i2.557

Radoï, L., Menvielle, G., Cyr, D., Lapôtre-Ledoux, B., Stücker, I., and Luce, D. (2015). Population attributable risks of oral cavity cancer to behavioral and medical risk factors in France: results of a large population-based case–control study, the ICARE study. *BMC Cancer* 15, 827. doi: 10.1186/s12885-015-1841-5

Schmidt, T. S. B., Hayward, M. R., Coelho, L. P., Li, S. S., Costea, P. I., Voigt, A. Y., et al. (2019). Extensive transmission of microbes along the gastrointestinal tract. *eLife* 8, e42693. doi: 10.7554/eLife.42693

Segal, L. N., Clemente, J. C., Tsay, J. C., Koralov, S. B., Keller, C., Wu, B. G., et al. (2016). Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a Th17 phenotype. *Nat. Microbiol.* 1, 1–24. doi: 10.1038/nmicrobiol.2016.31

Segata, N., Haake, S. K., Mannon, P., Lemon, K. P., Waldron, L., Gevers, D., et al. (2012). Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol.* 13, R42. doi: 10.1186/gb-2012-13-6-r42

Shah, J. P., and Gil, Z. (2009). Current concepts in management of oral cancer - Surgery. *Oral. Oncol.* 45, 394–401. doi: 10.1016/j.oraloncology.2008.05.017

Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer statistic. *CA Cancer J. Clin.* 69, 7–34. doi: 10.3322/caac.21551

Strauss, J., Kaplan, G. G., Beck, P. L., Rioux, K., Panaccione, R., Devinney, R., et al. (2011). Invasive potential of gut mucosa-derived fusobacterium nucleatum positively correlates with IBD status of the host. *Inflamm. Bowel Dis.* 17, 1971–1978. doi: 10.1002/ibd.21606

Takeshita, T., Yasui, M., Shibata, Y., Furuta, M., Saeki, Y., Eshima, N., et al. (2015). Dental plaque development on a hydroxyapatite disk in young adults observed by using a barcoded pyrosequencing approach. *Sci. Rep.* 5, 8136. doi: 10.1038/srep08136

Takeshita, T., Kageyama, S., Furuta, M., Tsuboi, H., Takeuchi, K., Shibata, Y., et al. (2016). Bacterial diversity in saliva and oral health-related conditions: the Hisayama Study. *Sci. Rep.* 6, 22164. doi: 10.1038/srep22164

Varela-Lema, L., Ruano-Ravina, A., Juiz Crespo, M. A., and Barros-Dios, J. M. (2010). Tobacco consumption and oral and pharyngeal cancer in a Spanish male population. *Cancer Lett.* 288, 28–35. doi: 10.1016/J.CANLET.2009.06.015

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi: 10.1128/AEM.00062-07

Wu, J., Li, Q., and Fu, X. (2019). Fusobacterium nucleatum Contributes to the Carcinogenesis of Colorectal Cancer by Inducing Inflammation and Suppressing Host Immunity. *Transl. Oncol.* 12, 846–851. doi: 10.1016/j.tranon.2019.03.003

Yachida, S., Mizutani, S., Shiroma, H., Shiba, S., Nakajima, T., Sakamoto, T., et al. (2019). Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat. Med.* 25, 968–976. doi: 10.1038/s41591-019-0458-7

Yamanaka, W., Takeshita, T., Shibata, Y., Matsuo, K., Eshima, N., Yokoyama, T., et al. (2012). Compositional stability of a salivary bacterial population against supragingival microbiota shift following periodontal therapy. *PloS One* 7, e42806. doi: 10.1371/journal.pone.0042806

Zaura, E., Brandt, B. W., de Mattos, M. J. T., Buijs, M. J., Caspers, M. P. M., Rashid, M. U., et al. (2015). Same Exposure but two radically different responses to antibiotics: Resilience of the salivary microbiome versus long-term microbial shifts in feces. *mBio* 6, e01693–e01615. doi: 10.1128/mBio.01693-15

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cellular and Infection Microbiology

ORIGINAL RESEARCH
published: 15 December 2020
doi: 10.3389/fcimb.2020.581888

Shifts in the Bacterial Community of Supragingival Plaque Associated With Metabolic-Associated Fatty Liver Disease

Fen Zhao[1,2,3†], Ting Dong[1,2,3†], Ke-Yong Yuan[1,2,3], Ning-Jian Wang[4], Fang-Zhen Xia[4], Di Liu[5,6], Zhi-Min Wang[7], Rui Ma[1,2,3], Ying-Li Lu[4*] and Zheng-Wei Huang[1,2,3*]*

[1] Department of Endodontics, Shanghai Ninth People's Hospital, College of Stomatology, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [2] National Clinical Research Center for Oral Diseases, Shanghai, China, [3] Shanghai Key Laboratory of Stomatology & Shanghai Research Institute of Stomatology, Shanghai, China, [4] Institute and Department of Endocrinology and Metabolism, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [5] Computational Virology Group, Center for Bacteria and Viruses Resources and Bioinformation, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan, China, [6] University of Chinese Academy of Sciences, Beijing, China, [7] Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center, Shanghai, China

OPEN ACCESS

Edited by:
Dongmei Deng,
VU University Amsterdam,
Netherlands

Reviewed by:
Mingyun Li,
Sichuan University, China
Sayaka Katagiri,
Tokyo Medical and Dental University,
Japan

***Correspondence:**
Zheng-Wei Huang
huangzhengwei@shsmu.edu.cn
Ying-Li Lu
luyingli2008@126.com
Rui Ma
marui1723@sina.com

[†]These authors share first authorship

Specialty section:
This article was submitted to Microbiome in Health and Disease, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: 01 August 2020
Accepted: 05 November 2020
Published: 15 December 2020

Citation:
Zhao F, Dong T, Yuan K-Y, Wang N-J, Xia F-Z, Liu D, Wang Z-M, Ma R, Lu Y-L and Huang Z-W (2020) Shifts in the Bacterial Community of Supragingival Plaque Associated With Metabolic-Associated Fatty Liver Disease. Front. Cell. Infect. Microbiol. 10:581888. doi: 10.3389/fcimb.2020.581888

Metabolic-associated fatty liver disease (MAFLD), also known as the hepatic manifestation of metabolic disorders, has become one of the most common chronic liver diseases worldwide. The associations between some oral resident microbes and MAFLD have been described. However, changes to the oral microbial community in patients with MAFLD remain unknown. In this study, variations to the supragingival microbiota of MAFLD patients were identified. The microbial genetic profile of supragingival plaque samples from 24 MAFLD patients and 22 healthy participants were analyzed by 16S rDNA sequencing and bioinformatics analysis. Clinical variables, including indicators of insulin resistance, obesity, blood lipids, and hepatocellular damage, were evaluated with laboratory tests and physical examinations. The results showed that the diversity of the supragingival microbiota in MAFLD patients was significantly higher than that in healthy individuals. Weighted UniFrac principal coordinates analysis and partial least squares discriminant analysis showed that the samples from the MAFLD and control groups formed separate clusters (Adonis, P = 0.0120). There were 27 taxa with differential distributions (linear discriminant analysis, LDA>2.0) between two groups, among which *Actinomyces* spp. and *Prevotella 2* spp. were over-represented in the MAFLD group with highest LDA score, while *Neisseria* spp. and *Bergeyella* spp. were more abundant in the control group. Co-occurrence networks of the top 50 abundant genera in the two groups suggested that the inter-genera relationships were also altered in the supragingival plaque of MAFLD patients. In addition, in genus level, as risk factors for the development of MAFLD, insulin resistance was positively correlated with the abundances of *Granulicatella*, *Veillonella*, *Streptococcus*, and *Scardovia*, while obesity was positively correlated to the abundances of *Streptococcus*, *Oslenella*, *Scardovia*, and *Selenomonas*. Metagenomic

predictions based on Phylogenetic Investigation of Communities by Reconstruction of Unobserved States revealed that pathways related to sugar (mainly free sugar) metabolism were enriched in the supragingival plaque of the MAFLD group. In conclusion, as compared to healthy individuals, component and interactional dysbioses were observed in the supragingival microbiota of the MAFLD group.

Keywords: metabolic-associated fatty liver disease, 16S rDNA sequencing, microbial community dysbiosis, insulin resistance, obesity, supragingival plaque

INTRODUCTION

Metabolic-associated fatty liver disease (MAFLD), formerly known as non-alcoholic fatty liver disease (NAFLD), refers to a wide spectrum of liver diseases characterized by the presence of hepatic steatosis in the absence of secondary causes, which include simple hepatic steatosis and steatohepatitis (Eslam et al., 2020). Progressive hepatic fibrosis is a hallmark of advanced steatohepatitis and can lead to cirrhosis, liver failure, and hepatocellular carcinoma (Sheka et al., 2020). MAFLD is an emerging public health concern worldwide, with a global pooled prevalence, by imaging, of 25.24% (Chalasani et al., 2018). The etiology of MAFLD is described as a complex hepatic manifestation of metabolic disorders, and the "multiple-hit" theory has been widely accepted as a potential mechanism involving insulin resistance (IR), obesity, chronic low grade inflammation, a sedentary lifestyle, regular consumption of a high fat diet, adipose tissue dysfunction, genetic factors, and gut microbial dysbiosis (Tiniakos et al., 2010; Buzzetti et al., 2016). With the increasing acknowledgement of the oral microbiome as a source of systemic inflammation, dysbiosis in the oral microbiome had been closely associated with metabolic diseases, including MAFLD (Acharya et al., 2017). A previous study found that the frequency of *Porphyromonas gingivalis* in the oral cavity is significantly higher in MAFLD patients as compared with health subjects (Yoneda et al., 2012). In another study, intravenous injection of sonicated *P. gingivalis* caused impaired glucose tolerance, IR, and liver steatosis in C57BL/6J mice fed a high-fat diet (Sasaki et al., 2018). However, because relatively few oral resident microbes have been the focus of previous studies, changes to the oral microbial community in patients with MAFLD remain unknown.

As a common site of oral sampling, the microbial profile of supragingival plaque is thought to reflect the health status of the host, such as gestation (Lin et al., 2018), type 2 diabetes (Artese et al., 2015), and other conditions (Espinoza et al., 2018). As compared with subgingival plaque, the supragingival microbiota can be acquired more easily and with less discomfort to the host (Lin et al., 2018). Therefore, sampling of the supragingival microbiota presents a promising method to assess oral and systematic health, especially for a large-scale health census. To the best of our knowledge, the present study is the first to assess the association between supragingival microbiota and MAFLD.

In the present study, supragingival plaques were obtained from 24 patients with MAFLD and 22 healthy controls. Illumina MiSeq PE300 sequencing and bioinformatics analysis were employed to identify changes to the microbial profiles and inter-taxa relationships of the supragingival plaque of MAFLD patients. Specific microbial genera correlated with MAFLD and related clinical indexes were also identified. Furthermore, potential functional alterations to the supragingival microbiome were predicted based on the sequencing results. These findings would provide a deeper understanding of the oral ecological dysbiosis associated with MAFLD.

MATERIALS AND METHODS

Study Population

The study protocol was approved by the Ethics Committee of Shanghai Ninth People's Hospital affiliated with Shanghai Jiao Tong University, School of Medicine (Shanghai, China) (approval no. SH9H-2019-T295-1) and conducted in accordance with the tenets of the Declaration of Helsinki. Written informed consent was obtained from all participants prior to enrollment.

The study participants were recruited from a health census and assigned to one of two groups: the MAFLD group, consisting of 24 persons diagnosed with MAFLD *via* upper abdomen ultrasonography and other clinical examinations (Eslam et al., 2020), or a control group, consisting of 22 persons with normal findings by upper abdomen ultrasonography. Age and sex in the participants of the two groups were matched. For each participant, upper abdomen ultrasonography was performed successively by two experienced sonographers and those with the same diagnosis were enrolled.

The exclusion criteria were as follows: (i) the presence of other liver diseases (e.g., viral hepatitis, autoimmune hepatitis, and hepatolenticular degeneration); (ii) drug-induced hepatic steatosis (e.g., tamoxifen, amiodarone, valproate, methotrexate, and glucocorticoids); (iii) other factors that may cause hepatic steatosis (e.g., long-term total parenteral nutrition, inflammatory bowel disease, celiac disease, hypothyroidism, Cushing's syndrome, lipoprotein deficiency, lipid-atrophic diabetes, etc.); (iv) use of lipid-lowering drugs within the past 6 months; (v) type 1 or type 2 diabetes; (vi) current oral disease (e.g., untreated oral abscess, precancerous lesions, oral cancer, oral fungal infection, missing more than eight teeth, etc.); and (vii) other conditions (e.g., pregnant or lactating women, long-term heavy smoking, use of antibiotics for more than 5 days within the past 6 months,

severe acute episode of a systematic disease, abnormal thyroid function, familial hyperlipidermia, etc.).

Acquisition of Clinical Variables

All participant demographics were retrieved from self-reported questionnaires and fasting blood samples were collected to detect clinical levels of total cholesterol (TC), total triglycerides (TG), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma glutamyl transpeptidase (GGT), fasting plasma glucose (FPG), and fasting serum insulin (FSI). As an approximation of IR, the Homeostatic Model Assessment for Insulin Resistance (HOMA-IR) equation was calculated as HOMA-IR = FPG (mmol/L) × FSI (mU/L)/22.5 (Matthews et al., 1985). The waist circumference, weight, height, and body mass index (BMI) of each participant were acquired *via* physical examination. The unpaired Student's t-test was applied for analysis of all clinical variables with an exception of "sex," which was analyzed using the Yates' continuity corrected chi-squared test.

Dental Examination and Supragingival Plaque Collection

An abbreviated dental exam was performed on all the participants by the same dentist using a periodontal probe. At least 9 teeth present and >30% of the probed sites with attachment loss ≥1 mm were diagnosed as periodontitis (Armitage, 1999). The prevalence of periodontitis between the two groups was analyzed using the Yates' continuity corrected chi-squared test. Supragingival plaque was collected before eating in the morning in accordance with the methods described in the Manual of Procedures for the Human Microbiome Project (https://www.hmpdacc.org/hmp/doc/HMP_MOP_ Version12_0_072910.pdf) with minor modifications. The index teeth (#3, #9, #12, #19, #25, and #28) were isolated with cotton rolls and dried under a gentle stream of air. A sterile sickle scaler was used to collect the supragingival plaque from the buccal surfaces of the index teeth. Then, the scaler tips were immersed in 300 μl of sterile normal saline contained in a sterile Eppendorf tube for 5–10 s with slight shaking and the surface of the scaler was wiped off on the inside edge of the tube. When the supragingival plaques of all index teeth were obtained, the Eppendorf tubes were sealed, marked, and kept frozen in liquid nitrogen until DNA extraction.

DNA Extraction, Amplification, and High-Throughput Sequencing

Total bacterial genomic DNA was extracted from the collected supragingival samples using the QIAamp DNA Mini Kit (Qiagen, Valencia, CA, USA) in accordance with manufacturer's protocols. The concentration and purification of the extracted DNA were determined using a NanoDrop 2000 UV-vis spectrophotometer (Thermo Scientific, Wilmington, DE, USA), while DNA quality was checked by 1% agarose gel electrophoresis.

The 16S rDNA hypervariable V3–V4 region was PCR-amplified with the forward primer 338F (5′-ACTCCTAC GGGAGGCAGCAG-3′) and the reverse primer 806R (5′-GGACTACHVGGGTWTCT AAT-3′) using a GeneAmp™ PCR System 9700 (Applied Biosystems, Carlsbad, CA, USA). The parameters of the PCR reactions have been described in a previous study (Tong et al., 2019). Each PCR reaction was performed in triplicate and all resulting PCR products were extracted from 2% agarose gels and then further purified using the AxyPrep DNA Gel Extraction Kit (Axygen Biosciences, Union City, CA, USA) and quantified using a QuantiFluor® Single-Tube Fluorometer (Promega Corporation, Madison, WI, USA) in accordance with the manufacturer's instructions. Purified amplicons from different samples were pooled in equimolar concentrations and paired-end sequenced on an Illumina MiSeq PE300 sequencing platform (Illumina, Inc., San Diego, CA, USA).

Data Processing and Bioinformatics Analysis

Raw fastq files were quality-filtered using the Trimmomatic read trimming tool (https://kbase.us/) and merged using FLASH software (version 1.2.11; https://ccb.jhu.edu/software/FLASH/index.shtml) in accordance with the criteria described in a previous report (Tong et al., 2019). After trimming, operational taxonomic units (OTUs) were clustered at a similarity cutoff value of 97% using the UPARSE algorithm (version 7.1; http://drive5.com/uparse/). The taxonomy of each 16S rRNA gene sequence was analyzed with RDP Classifier algorithm (http://rdp.cme.msu.edu/) against the Silva 16S rRNA database (Release132 http://www.arb-silva.de), the confidence threshold was set to 70%.

Alpha diversity indexes were calculated using MOTHUR software for describing and comparing microbial communities (version 1.30.2; https://www.mothur.org/wiki/Download_mothur) and rarefaction curves were constructed at an inter-sequence similarity value of 97% using the QIIME bioinformatics pipeline (version 1.9.1; http://qiime.org/install/index.html). Bar plots were generated to visualize the species composition of all samples at the phylum and genus levels, and heat maps at the genus level were constructed using the R platform (version 3.6.1). Weighted UniFrac principal coordinates analysis (PCoA), nonparametric multivariate analysis of variance (Adonis), and partial least squares discriminant analysis (PLS-DA) were performed to identify differences in species composition between the MAFLD and control groups using QIIME. The linear discriminant analysis (LDA) effect size (LEfSe; http://huttenhower.sph.harvard.edu/galaxy) was applied to identify the most discriminatory taxa between the groups at the phylum to genus levels. Taxa with logarithmic LDA scores of >2.0 were regarded as discriminative species. Co-occurrence networks of the 50 most abundant genera of each group were demonstrated using NetworkX (version 1.9.1). Spearman's correlation coefficients were calculated and those with a $|\rho|$ value of >0.5 and a probability (P) value of < 0.05 were visualized. Spearman's correlation coefficients among the clinical variables and the top 50 abundant genera of the supragingival microbiome were calculated, and the results were

visualized as heat maps *via* the R platform. Furthermore, the bioinformatics software package Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt2, version 1.1.0; http://picrust.github.io/picrust/) was used to predict the functional pathways of each group according to the Kyoto Encyclopedia of Genes and Genomes (KEGG). Statistical differences in top 30 abundant KEGG level 2 pathways and top 50 abundant KEGG level 3 pathways between two groups were determined by Wilcoxon rank-sum test with a Benjamini-Hochberg false discovery rate (FDR) correction to adjust P values for multiple testing.

RESULTS

Subject Characteristics

The demographic and clinical characteristics of the participants are summarized in **Table 1**. There were no significant differences in age, sex, blood pressure, heart rate and prevalence of periodontitis between the two groups. Subjects in the MAFLD group had relatively higher TG levels and lower HDL-C levels, and thus had higher TG/HDL-C ratios than the control group. Moreover, ALT and GGT levels were relatively higher in the MAFLD group than the control group. Because subjects with diabetes were excluded, there was no significant difference in FPG levels between the two groups, but FSI and HOMA-IR were significantly increased in the MAFLD group. Waist circumference was significantly higher in the MAFLD group, illustrating visceral adiposity was ubiquitous in MAFLD patients.

TABLE 1 | Demographic and clinical characteristics of the study participants.

Parameter	Health(n=22)	MAFLD(n=24)	*P*
Age (years)	52.91 ± 4.25	52.75 ± 4.39	0.9013
Sex (M: F)	11: 11	13: 11	0.9898
TC (mmol/L)	5.77± 0.78	5.47 ± 0.88	0.2254
TG (mmol/L)	1.55 ± 1.04	2.51 ± 1.28	0.0083*
LDL-C (mmol/L)	3.41± 0.52	3.28± 0.66	0.4735
HDL-C (mmol/L)	1.40 ± 0.36	1.10± 0.21	0.0011*
ALT (U/L)	22.68 ± 12.99	40.25 ± 27.55	0.0092*
AST (U/L)	31.5 ± 18.13	38.21 ± 19.09	0.2292
AST/ALT	1.52 ± 0.66	1.11 ± 0.43	0.0142*
GGT (U/L)	24.50 ± 21.16	57.42 ± 52.11	0.0084*
FPG (mmol/L)	5.01 ± 0.46	5.18+0.56	0.3196
FSI(mU/L)	4.02 ± 2.68	6.84 ± 2.68	0.0015*
HOMA-IR	0.94 ± 0.59	1.58 ± 0.66	0.0013*
Weight (kg)	62.20 ± 11.75	68.23 ± 9.64	0.0631
BMI	24.13 ± 3.00	25.79 ± 2.61	0.0502
WC(cm)	81.84 ± 9.45	89.79 ± 6.88	0.0020*
SBP (mm Hg)	139.00 ± 26.88	136.46 ± 19.65	0.7144
DBP (mm Hg)	82.77 ± 14.93	82.79 ± 12.17	0.9962
HR	76.09 ± 8.38	73.71 ± 10.25	0.3954
Periodontitis (%)	81.82	79.17	0.8843

*Data are presented as the mean ± standard deviation. *P < 0.05. All clinical and demographic data were analyzed using the unpaired t-test with an exception of "sex" and "periodontitis", which were analyzed with the Yates' continuity corrected $\chi 2$ test. Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; DBP, diastolic blood pressure; FPG, fasting plasma glucose; FSI, fasting serum insulin; GGT, gamma glutamyl trans-peptidase; HDL-C, high-density lipoprotein cholesterol; HOMA-IR, homeostatic model assessment of insulin resistance [HOMA-IR = FPG (mmol/L) × FSI (mU/L)/22.5]; HR, heart rate; LDL-C, low-density lipoprotein cholesterol; SBP, systolic blood pressure; TC, total cholesterol; TG, total triglycerides; WC, waist circumference.*

Bacterial Diversity and Community Structure of Supragingival Microbiota

The concentration and purification of the extracted DNA met the requirements for further experiments (**Table S1**), results of 2% agarose gel electrophoresis revealed that the PCR products of the extracted DNA was qualified to further sequencing and analysis (**Figure S1**). A total of 2,318,404 high quality sequences were produced, with an average of 50,507 ± 11,212 sequences per sample. In total, 23 phyla, 37 classes, 92 orders, 163 families, 339 genera, 628 species, and 1,021 OTUs were identified after taxonomic assignment of the sequences. All 16S rRNA gene sequences were submitted to the NCBI Sequence Read Archive (SRA) under bioproject accession PRJNA645880 (http://www.ncbi.nlm.nih.gov/sra).

As shown in **Table 2**, the MAFLD group had a higher Shannon index and lower Simpson index, suggesting the diversity of supragingival microbiota was higher in the MAFLD group than the control group (**Table 2**). Good's coverage indexes, which were close to 1 (**Table 2**), and rarefaction curves based on OTU levels (**Figure S2**) reached saturation plateaus, indicating that the sequencing depths were sufficient to represent the majority of the microbiota in both groups.

The taxa abundance of the supragingival microbiome from the two groups at the phylum and genus levels are depicted in **Figures 1A, B**, respectively. In general, the dominant taxa of the two communities were similar and consistent with the core species of the supragingival microbiome. The core phyla of all samples from both groups consisted of *Bacteroidetes, Proteobacteria, Actinobacteria, Firmicutes, Fusobacteria, Patescibacteria, Epsilonbacteraeota*, and *Spirochaetes*. Of the top 10 abundant genera, the MAFLD group had higher proportions of *Capnocytophaga, Leptotrichia, Corynebacterium, Actinomyces, Streptococcus, Fusobacterium, Prevotella*, and *Veillonella*; while *Neisseria* and *Comamonas* were more prevalent in the control group. The abundances of the top 50 abundant genera in each sample are displayed in the heat map presented in **Figure S3**.

Weighted UniFrac PCoA at the OTU level was employed to evaluate the similarity of the bacterial communities between the two groups. The results indicated that although samples from the two groups partly overlapped, there was a tendency of separation along the PC1 axis (**Figure 1C**). The results of

TABLE 2 | α-diversity of supragingival microbiota in the MAFLD and Control groups.

	MAFLD	Health	*P*
Ace	318.96 ± 76.36	285.74 ± 82.16	0.1561
Chao 1	324.99 ± 81.69	287.67 ± 77.17	0.0969
Shannon	4.02 ± 0.27	3.80 ± 0.39	0.0303*
Simpson	0.04 ± 0.01	0.05 ± 0.02	0.0192*
Coverage	0.9983 ± 0.0006	0.9987 ± 0.0006	0.1319

*Results are presented as the mean ± standard deviation. *P < 0.05. All α-diversity estimators were analyzed using the Wilcoxon rank-sum test.*

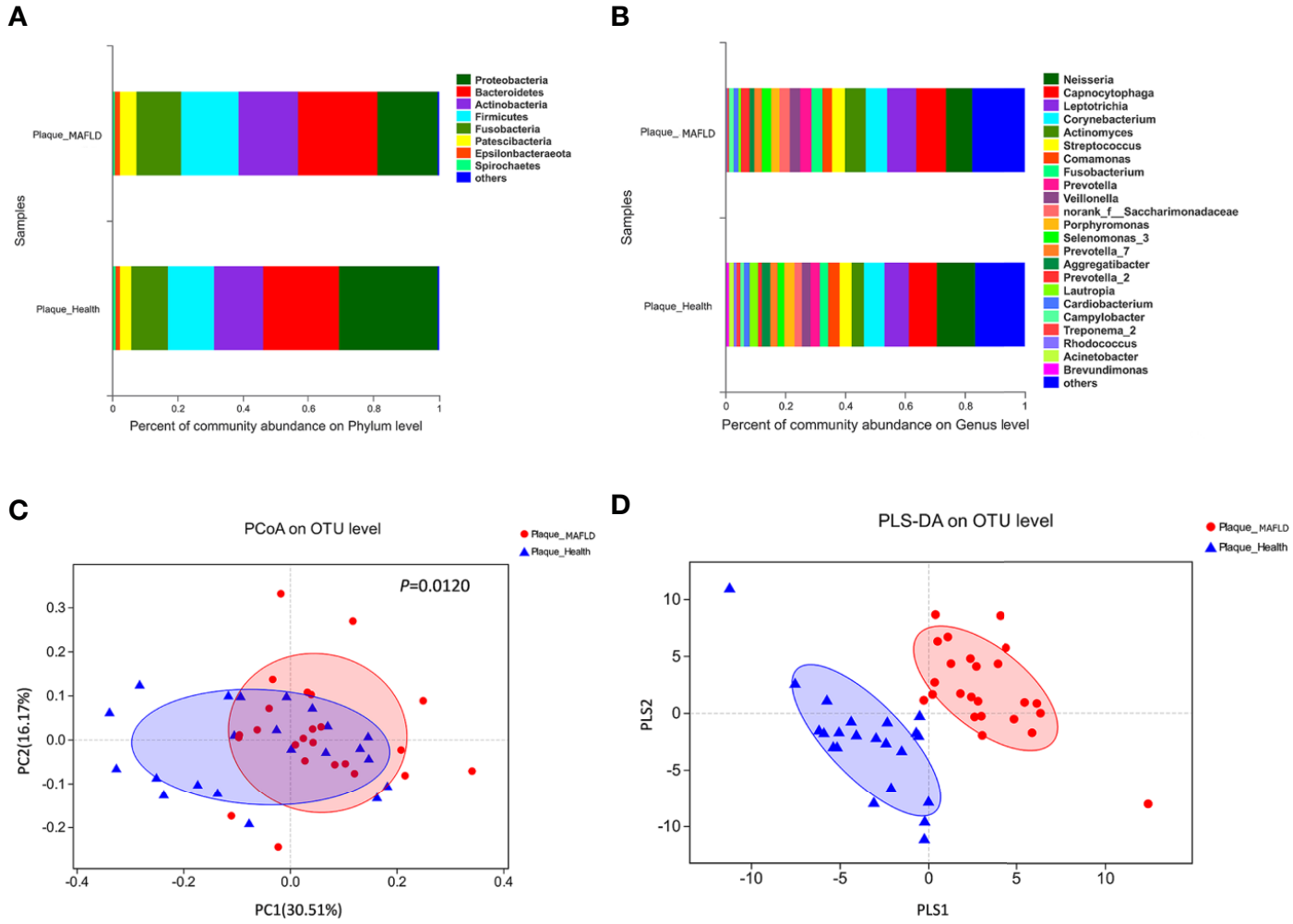


FIGURE 1 | Comparison of supragingival microbiota structures in the metabolic-associated fatty liver disease (MAFLD) and control groups. **(A)** Community structures at the phylum level. **(B)** Community structures at the genus level. **(C)** Weighted UniFrac principal coordinates analysis (PCoA) at the operational taxonomic unit (OTU) level. **(D)** Partial least squares discriminant analysis (PLS-DA) at the OTU level.

Adonis ($P = 0.0120$) based on weighted UniFrac distances further verified the existence of significant differences in the overall structures of the supragingival microbiota of the two groups. PLS-DA, a supervised analysis suitable for high dimensional data, showed separate clustering of the samples from the MAFLD and control groups (**Figure 1D**), further demonstrating remarkable differences in the supragingival microbiota between the two groups.

Alterations of the Supragingival Microbial Phylotypes/Inter-Genera Relationship Associated With MAFLD

A circular cladogram based on the LEfSe results demonstrated differentially abundant taxa between the two groups (**Figure 2A**). Genera with logarithmic LDA scores of >2.0 ($P < 0.05$) are plotted in **Figure 2B** and the selected taxa in other taxonomic level (from phylum to family) were shown in **Figure S4**. Briefly, there were 27 taxa with differential distributions between the two groups. At the genus level, *Actinomyces, Prevotella 2, Scardovia, Megasphaera*, and *Alysiella* were more abundant in the MAFLD group, while *Neisseria, Bergeyella, Sphingomonas*, and *H1* were over-represented in the control group. Co-occurrence networks of the top 50 abundant genera of the two groups were depicted in **Figures 3A, B**. In general, the taxa within the main bacterial cluster (>5 nodes and connected with intense lines) had stronger and more complex interrelationships in the MAFLD group.

Correlations Between Clinical Variables and Supragingival Microbiota

The heat map presented in **Figure 4** depicts the correlations between the 50 most abundant bacterial genera and single clinical variables based on the Spearman's correlation coefficients. Of the top 50 abundant genera, HOMA-IR showed positive correlations with *Streptococcus, Scardovia, Granulicatella*, and *Veillonella*, but a negative correlation with *Neisseria*. *Alloprevotella*, and *Peptostreptococcus* were extremely significantly negatively correlated with TC and LDL-C levels. *Aggregatibacter* was negatively correlated with TG levels, but positively correlated with HDL-C levels. *Streptococcus, Oslenella, Scardovia*, and *Selenomonas* were significantly positively correlated with BMI and waist circumference. AST/ALT and GGT, two indicators of hepatocellular damage, were negatively correlated with *Capnocytophaga*.

Predictive Metagenome Functional Profiling of the Supragingival Microbiomes of the MAFLD and Control Groups

To detect functional differences in the supragingival microbiomes of the MAFLD and control groups, PICRUSt 2 was employed to predict the metagenome functional contents based on the 16S rRNA datasets. Statistically significant differences ($P < 0.05$) in KEGG pathways were calculated with the Wilcoxon rank-sum test with FDR correction. As shown in **Figure 5A**, among the top 30 abundant KEGG level 2 pathways, Carbohydrate metabolism, Translation, Cellular community–prokaryotes and Membrane transport were significantly increased in the MAFLD group. At KEGG pathway level 3, starch and sucrose metabolism, fructose and mannose metabolism, and galactose metabolism were enriched in the MAFLD group. Meanwhile, a total of 9 pathways among the top 50 abundant KEGG level 3 pathways, including carbon metabolism, pyruvate metabolism, and the citrate cycle, were significantly decreased in the MAFLD group (**Figure 5B**).

DISCUSSION

The oral microbiota is attracting increased attention because of probable associations with metabolic disorders (Si et al., 2017; Chen et al., 2020; Wei et al., 2020). The associations between the oral microbiota and metabolic disorders can be explained in the following two aspects. On the one hand, microbial dysbiosis in the oral cavity is a source of systemic inflammation, which could lead to chronic low-grade inflammation and adversely affects the metabolic health of the host (Acharya et al., 2017). On the other hand, the oral microbiota can influence the composition of the gut microbiome, which plays important roles in metabolic health (Segata et al., 2012; Tremaroli and Bäckhed, 2012; Arimatsu et al., 2014; Li et al., 2019). Due to the anatomical position, about 10^{11} bacteria are swallowed from the oral cavity to the stomach every day (Segata et al., 2012), cultivation and sequencing techniques have also substantiated the association between the oral and gut microbiomes: Arimatsu et al. reported that oral administration of *P. gingivalis* significantly altered the *Firmicutes/Bacteroidetes* ratio, a significant index to evaluate the health status of the gut microbiome (Arimatsu et al., 2014); Li et al. found that the oral microbiota could overcome physical barriers and colonize the gut in gnotobiotic mice (Li et al., 2019). These findings acknowledged that the oral microbiota plays an important role in the development of metabolic diseases *via* "oral-gut axis". For MAFLD, although some oral resident microbes have been associated with the development of it (Yoneda et al., 2012; Sasaki et al., 2018), there has been no microbiome-wide association study of the association between the development of MAFLD and oral microbial ecology. Shaped by the health status of the host, supragingival plaque has been related to various metabolic disorders. For example, Hintao et al. reported significant differences in the microbial profiles of supragingival plaque between subjects with and without diabetes (Hintao et al., 2007); La Monte et al. found that metabolic syndrome was significantly associated with supragingival plaque (odds ratio = 1.74; 95% confidence interval = 1.22–2.50) (LaMonte et al., 2014). Considering supragingival plaque can be obtained with minimal discomfort and risk (Lin et al., 2018), it was collected in this study to explore the ecological shifts of oral microbiota in MAFLD patients. By screening with strict inclusion/exclusion criteria and matching of confounding factors, the differences among the participants were minimized as much as possible in order to focus on compositional and structural differences of the supragingival microbiota in MAFLD patients.

The diversity of the supragingival microbiota of each group was determined using alpha diversity estimators. It is generally

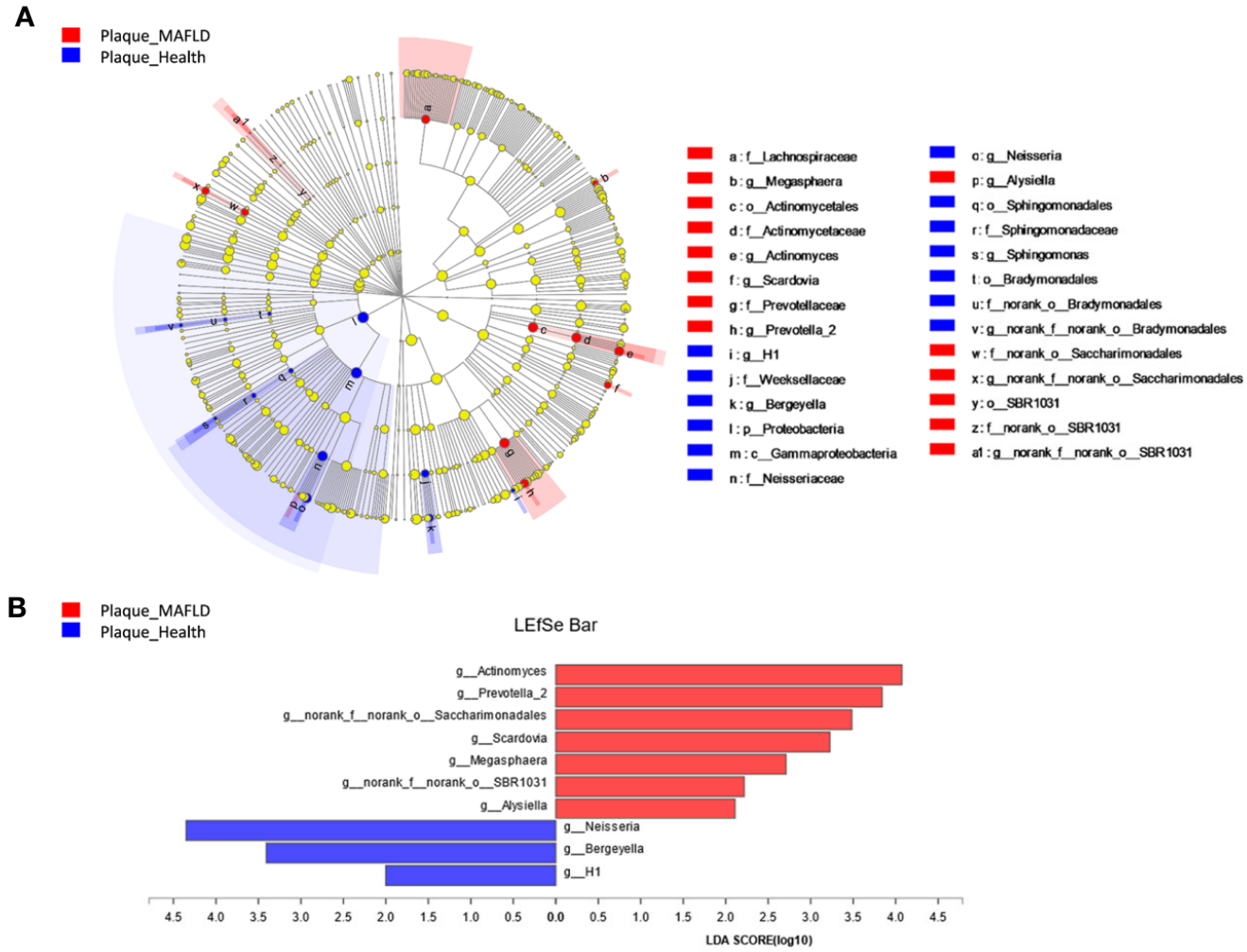


FIGURE 2 | Bacterial phylotypes with altered abundances associated with MAFLD. **(A)** A cladogram of taxonomic representation based on LEfSe. Red indicates enrichment in samples from MAFLD patients and blue indicates the taxa enriched in samples from healthy controls. **(B)** A histogram of the logarithmic linear discriminant analysis (LDA) scores were calculated for the selected genera (logarithmic LDA>2.0, $P<0.05$).

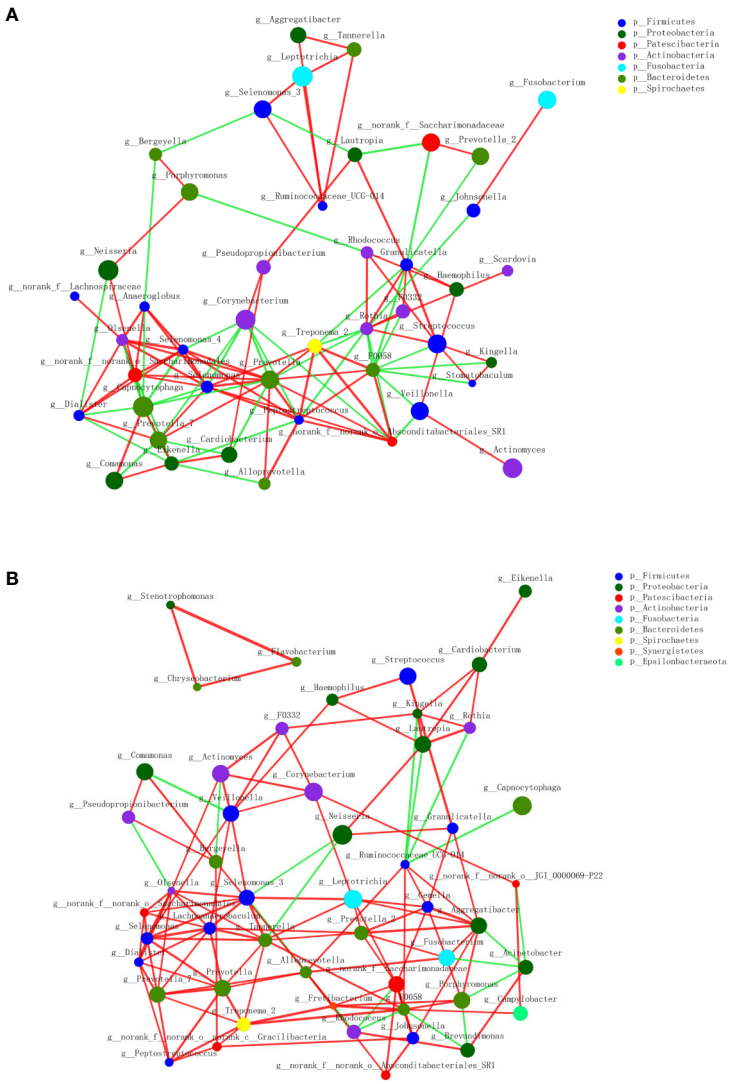


FIGURE 3 | Co-occurrence networks of the top 50 abundant genera in supragingival microbiota. The MAFLD group is shown in **(A)** and Health group in **(B)**. The size of the node indicates the mean relative abundance of the corresponding genus. The same color represents genera belonging to the same phylum. The thickness of the connecting lines corresponds to the coefficient values ($P < 0.05$). The red and green lines indicate positive and negative correlations, respectively.

acknowledged that microbial diversity reflects the health status of the host. For example, decreased diversity of gut microbiota indicates functional or metabolic disorders in the host (Kelly et al., 2016), while increased diversity of oral microbiota is reported to imply poor oral (Camelo-Castillo et al., 2015; Takeshita et al., 2016) and holistic health (Armingohar et al., 2014; Si et al., 2017), because in a state of poor oral health, gingival bleeding provides a richer nutrient source (Segata et al., 2012). In the present study, increased diversity (lower Simpson index and higher Shannon index) of the supragingival microbiota in the MAFLD group was observed, suggesting possible alterations to the nutritional status of supragingival plaque in MAFLD patients. Consistent with previous studies (Utter et al., 2016; Lin et al., 2018), the core phyla identified in the present study included *Proteobacteria*, *Bacteroidetes*, *Firmicutes*, *Actinobacteria*, and *Fusobacteria*, which accounted for 93.83% and 92.37% of the supragingival microbiomes of the control and MAFLD groups (**Figure 1A**), respectively. Similarly, although the proportions differed, the majority of the observed genera (including *Capnocytophaga*, *Leptotrichia*, *Corynebacterium*, *Actinomyces*, *Streptococcus*, *Fusobacterium*, *Prevotella*, *Veillonella*, *Neisseria*, and *Comamonas*) existed in both groups, thereby also supporting the core genera of the supragingival microbiota (Lin et al., 2018). A lower *Firmicutes/Bacteroidetes* ratio is considered as a healthy trait in both the oral cavity and gut (Chen et al., 2020). In the present study, the *Firmicutes/Bacteroidetes* ratio was lower in the supragingival plaque of the control group as compared to the MAFLD group (61.41% vs. 72.38%, calculated from **Figure 1A** respectively), indicating dysbiosis of the supragingival microbiota of the MAFLD group.

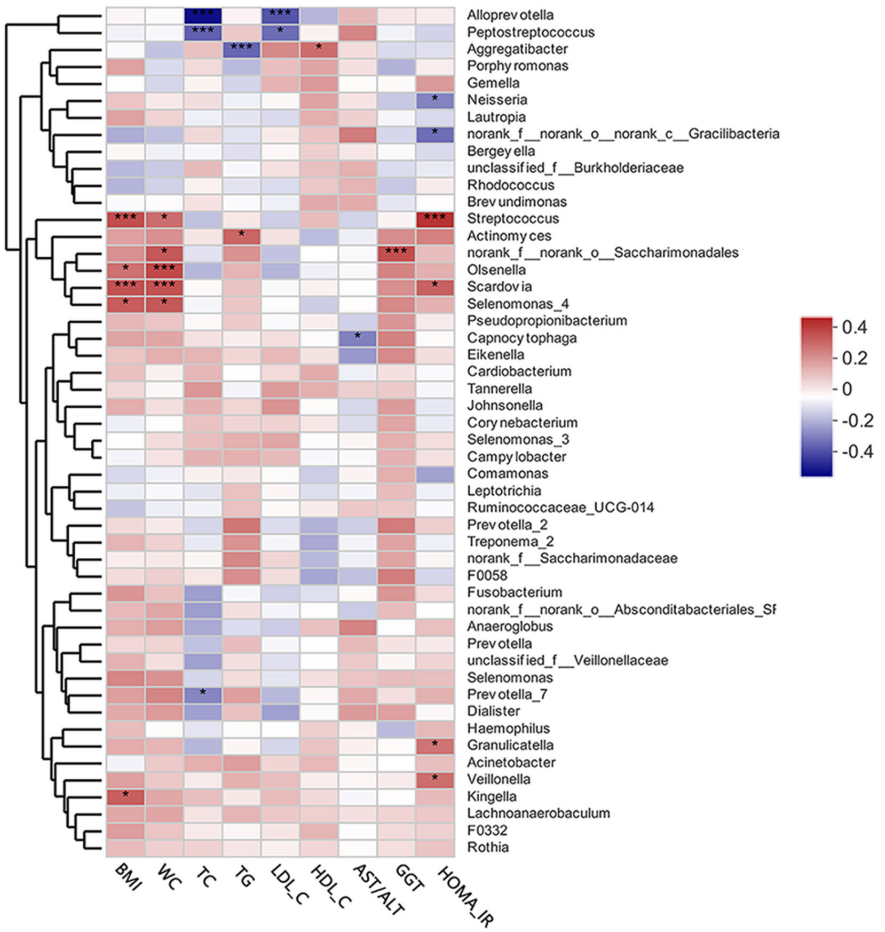


FIGURE 4 | A heat map of Spearman's correlation analysis of the top 50 abundant supragingival microbiota and clinical variables. The right side of the legend shows the color range of different R values. Species clustering trees are presented on the left side of the heat map. *$P < 0.05$; ***$P < 0.001$.

The PCoA and PLS-DA results demonstrated differences in the community compositions between the two groups (Adonis, $P = 0.0120$). The discriminatory taxa between two groups were identified using LEfSe. At the genus level, *Actinomyces* and *Prevotella 2* had the highest LDA scores in the MAFLD group. *Actinomyces* spp. is a normal resident bacteria of the oral cavity, which exerts important roles in biofilm formation (Polak et al., 2019). *Actinomyces* spp. has been associated with the severity of chronic periodontitis (Cao et al., 2018). *Prevotella 2* is a genus of Gram-negative, anaerobic bacteria that exists in the gut and are relevant to multiple disease states, including an increased lifetime risk of cardiovascular disease (Kelly et al., 2016), ankylosing spondylitis (Chen et al., 2019), and increased levels of C-reactive protein (Sun et al., 2019). Considering the consistency between the oral and gut microbiotas (Segata et al., 2012; Arimatsu et al., 2014; Li et al., 2019), the prevalence of *Prevotella 2* in the oral cavity is proposed as a potential marker of systematic diseases including MAFLD. In healthy participants, the genera *Neisseria* and *Bergeyella* had the highest LDA scores. *Neisseria* spp. is among the most abundant taxa in the oral cavity (Dong et al., 2018). A predominance of *Neisseria* spp. in the oral cavity indicates healthy conditions of the oral cavity (Meuric et al., 2017; Yamashita and Takeshita, 2017; Perez-Chaparro et al., 2018). *Bergeyella* spp. is a Gram-negative, aerobic bacteria (Muramatsu et al., 2019). In the present study, *Bergeyella* spp. was more prevalent in the control group, suggesting a negative correlation to MAFLD. Co-occurrence networks were used to predict inter-genera correlations of supragingival plaque between the two groups. As shown in **Figure 3**, there were significant differences in the interaction patterns of the two groups. In the MAFLD group, there were stronger and more complex interactions within the main cluster, but weaker correlations among the genera outside of the main cluster. Reportedly, an increase in interaction strength among taxa not only excludes other taxa, but decreases the stability of the microbial community (Ratzke et al., 2020). Therefore, it could be speculated that the supragingival microbial community of the MAFLD group was more unstable.

Inhibition of hepatic glucose production, increased accumulation of lipids in the liver, and IR are vital to the development of MAFLD (Bessone et al., 2019). It is currently

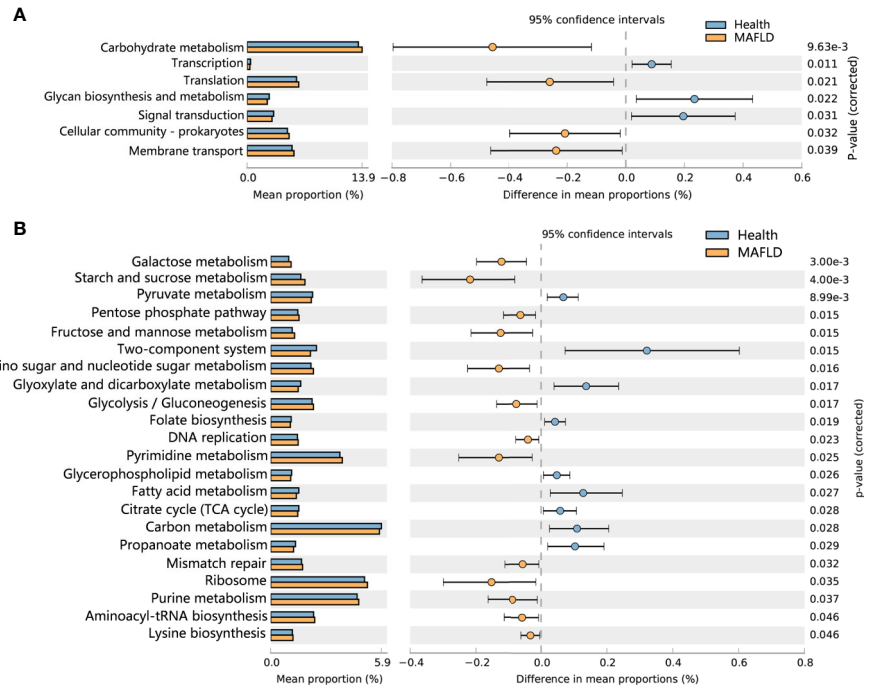


FIGURE 5 | Predictive metagenome functional profiling of supragingival microbiome of the MAFLD and control groups. **(A)** Relative abundance of predictive metagenome functional profiling of the top 20 abundant KEGG level 2 pathways (pathways with $P_{(FDR)}<0.05$ are shown). **(B)** Relative abundance of predictive metagenome functional profiling of the top 50 abundant KEGG level 3 pathways (pathways with $P_{(FDR)}<0.05$ are shown).

believed that IR is an independent risk factor for the severity of MAFLD (Bessone et al., 2019). As first proposed by Matthews et al. in 1985, HOMA-IR is both practical and highly efficient for the evaluation of IR in both clinical and scientific studies (Matthews et al., 1985; Tang et al., 2015). In the present study, HOMA-IR was beyond the normal range (normal range ≤1) in the MAFLD group and significantly higher than that in the control group ($P = 0.0013$) suggesting that IR is prevalent in patients with MAFLD. It was believed that chronic low-grade inflammation resulting from dysbiosis of the oral microbiota can reportedly aggravate IR (Bui et al., 2019). In this study, Spearman's correlation analysis revealed that the presence of *Granulicatella* spp., *Veillonella* spp., *Streptococcus* spp., and *Scardovia* spp. was positively correlated with HOMA-IR. It was reported that *Granulicatella* spp. has been positively correlated to periodontitis (Belstrøm et al., 2014), as well as infections outside of the oral cavity, such as infective endocarditis (Sasaki et al., 2020). *Veillonella* is a genus of Gram-negative anaerobic bacteria mainly found in the oral and gastrointestinal tracts. The presence of *Veillonella* spp. in the oral cavity has been correlated to increased production of pro-inflammatory cytokines (Bui et al., 2019; Li et al., 2020) and periodontal infections (Yamashita and Takeshita, 2017). *Streptococcus* spp. and *Scardovia* spp. are resident bacteria of the oral cavity that are closely related to caries formation (Kressirer et al., 2017). Although relatively few studies have investigated the relationship between caries-related bacteria and IR, patients with IR tend to have more decayed teeth (Loyola-Rodriguez et al., 2011).

In a state of chronic low-grade inflammation (Mervish et al., 2019), obesity is a contributor to various metabolic dysfunctions, such as MAFLD and type 2 diabetes (Canfora et al., 2019). As compared to BMI, visceral adiposity, as measured by waist circumference, has been closely linked to the severity of lipid deposition in the liver (Mervish et al., 2019), which is consistent with the results of the present study, which found an increase in waist circumference in MAFLD patients ($P = 0.0020$). In addition, multiple studies have verified the influence of obesity on the microbial profile of the oral cavity (Tam et al., 2018; Chattopadhyay et al., 2019). In this study, genera positively correlated with obesity mainly included *Streptococcus*, *Oslenella*, *Scardovia*, and *Selenomonas*. *Streptococcus* spp. and *Scardovia* spp. were also positively correlated to IR, supporting the positive association between obesity and IR (Mervish et al., 2019). The involvement of *Oslenella* spp. in endodontic infections (Vieira Colombo et al., 2016) and periodontal inflammation (Chen et al., 2020) have been well documented. In the *Veillonellaceae* family, *Selenomonas* is a genus of Gram-negative anaerobic bacteria. Members of *Veillonellaceae* family are considered to act as pro-inflammatory mediators (Tanner, 2015) and putative periodontal pathogens (Rocas and Siqueira, 2005). These results support the presumption that obesity is positively correlated to the abundance of bacteria associated with infectious diseases of the oral cavity (Maciel et al., 2016).

Dyslipidemia is a common clinical manifestation of MAFLD, especially hypertriglyceridemia and low serum HDL-C (Pacifico et al., 2014; Fukuda et al., 2016; Fan et al., 2019), which

were also verified in this study ($P = 0.0083$ for TG; $P = 0.0011$ for HDL-C). Reportedly, oral infectious diseases and dyslipidemia could have a two-way relationship without a clear cause-and-effect relationship (Jaramillo et al., 2013). *Actinomyces* spp. has been positively correlated to TG levels as a potential indicator of MAFLD-related metabolic dysfunction. A surprising result was that the presence of *Aggregatibacter* spp. was negatively correlated with TG levels, but positively correlated with HDL-C levels, which might indicate good health, challenging the mainstream concept that the presence of *Aggregatibacter* spp. (especially *A. actinomycetemcomitans*) is related to dyslipidemia and other metabolic diseases (Jaramillo et al., 2013). Sampling sites may explain this discrepancy because *Aggregatibacter* spp. is a anaerobic bacteria with growth behaviors that may change in response to aerobic conditions (supragingival habitats). However, the exact reasons for this paradox remain unclear.

Known as indicators of hepatocellular damage, elevated serum levels of transaminases and transpeptidases are also main clinical manifestations of MAFLD (Sheka et al., 2020). Moreover, a decreased AST/ALT ratio is regarded as biomarker of progressive MAFLD (Sheka et al., 2020). In this study, a decreased AST/ALT ratio ($P = 0.0142$) as well as elevated GGT ($P = 0.0084$) were prevalent in MAFLD patients, suggesting the enrolled MAFLD patients had different degrees of hepatocellular damage. *Capnocytophaga* is a genus of Gram-negative anaerobic bacteria reportedly associated with periodontitis (Cao et al., 2018) and hyperglycemia (Graves et al., 2019). In this study, an abundance of *Capnocytophaga* spp. was negatively correlated to the AST/ALT ratio, suggesting it could be a potential biomarker of MAFLD progression.

Metagenomic predictions based on PICRUSt2 revealed that functional changes between the control and MAFLD groups mainly involved metabolism (**Figure 5B**). Among the KEGG pathways level 3, metabolism of sugars (mainly free sugars, including starch and sucrose, fructose and mannose, and galactose) were more prevalent in subjects with MAFLD, revealing that supragingival plaque in MAFLD patients can easily obtain nutrients, which could explain the increased microbial diversity observed in the supragingival plaque of the MAFLD group (**Table 2**). Pathways related to aerobic respiration (including pyruvate metabolism, and the citrate cycle) were more abundant in the supragingival plaque of the control group, suggesting that the proportion of aerobic bacteria in the supragingival plaque is higher in healthy people. However, a deficiency of predicting functions based on taxa composition is that bacterial functions can change with the health status of the host (Ikeda et al., 2020). Consequently, metatranscriptomics and metabolomics of the microbiota may provide more realistic functional profiles.

As this is a pilot study with matching of confounding factors, some intriguing findings surfaced, but still need to be verified in future studies with larger sample sizes. In addition, with the increasing attention to the functions of the oral microbial community, it is essential to identify changes to the actual functional profiles of the supragingival microbiota in MAFLD by metatranscriptomics and metabolomics.

CONCLUSIONS

In the present study, dysbiosis of supragingival microbiota associated with MAFLD was characterized. Briefly, the results revealed increases in the abundances of bacteria associated with oral infections, decreases in the abundances of potential beneficial aerobic bacteria, and changes in the interactions of the core microflora with the supragingival microbiota in patients with MAFLD. Moreover, as risk factors for the development of MAFLD, IR was positively correlated to the abundances of *Granulicatella* spp., *Veillonella* spp., *Streptococcus* spp., and *Scardovia* spp., while obesity was positively correlated with the abundances of *Streptococcus* spp., *Oslenella* spp., *Scardovia* spp., and *Selenomonas* spp. The increased free sugar metabolic pathways suggested that supragingival bacteria related to the metabolism of free sugars were associated with MAFLD. These findings provide a deeper understanding of the association between the oral microbiome and MAFLD, although further studies are needed to explore potential causal relationships.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA645880.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Shanghai Ninth People's Hospital affiliated with Shanghai Jiao Tong University, School of Medicine. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FZ and TD performed the study design, data analysis, drafting, revising, final approval, and handled the accountability of all aspects of the work. K-YY and RM performed the bioinformatics analysis and data acquisition. N-JW and F-ZX performed the data analysis and acquisition. RM, Y-LL, DL, Z-MW, and Z-WH performed the study design, clinical sample collection, data analysis, revising, final approval and were also involved in the accountability of all aspects of the work. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grants from the Shanghai Top Priority Clinical Medicine Center (no. 2017ZZ01011), the Shanghai Municipal Key Clinical Specialty (no. shslczdzk1601), the Shanghai Clinical Research Center for Oral Diseases

(no. 19MC1910600), the National Natural Science Foundation of China (no. 82071104/81570964/81371143), and partly supported by the Shanghai Ninth People's Hospital affiliated with Shanghai Jiao Tong University, School of Medicine (no. JYJC201806).

ACKNOWLEDGMENTS

We thank our coworkers for their assistance and all the study participants. We also thank Shanghai Majorbio Biopharm Technology Co., Ltd. for providing the free online Majorbio Cloud Platform for the support provided for the bioinformatics analysis. This manuscript has been released as a pre-print at [ResearchGate] (Zhao et al., 2020).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2020.581888/full#supplementary-material

REFERENCES

Acharya, C., Sahingur, S. E., and Bajaj, J. S. (2017). Microbiota, cirrhosis, and the emergingoral-gut-liver axis. *JCI Insight* 2 (19), e94416. doi: 10.1172/jci.insight.94416

Arimatsu, K., Yamada, H., Miyazawa, H., Minagawa, T., Nakajima, M., Ryder, M. I., et al. (2014). Oral pathobiont induces systemic inflammationand metabolic changes associated with alteration of gut microbiota. *Sci. Rep.* 4, 4828. doi: 10.1038/srep04828

Armingohar, Z., Jorgensen, J. J., Kristoffersen, A. K., Abesha-Belay, E., and Olsen, I. (2014). Bacteria and bacterial DNA in atherosclerotic plaque and aneurysmal wall biopsies from patients with and without periodontitis. *J. Oral. Microbiol.* 6, 23408. doi: 10.3402/jom.v6.23408

Armitage, G. C. (1999). Development of a classification system for periodontal diseases and conditions. *Ann. Periodontol.* 4, 1–16. doi: 10.1902/annals.1999.4.1.1

Artese, H. P., Longo, P. L., Gomes, G. H., Mayer, M. P., and Romito, G. A. (2015). Supragingival biofilm control and systemicinflammation in patients with type 2 diabetes mellitus. *Braz. Oral. Res.* 29 (1), 1–7. doi: 10.1590/1807-3107BOR-2015.vol29.0071

Belstrøm, D., Fiehn, N. E., Nielsen, C. H., Kirkby, N., Twetman, S., Klepac-Ceraj, V., et al. (2014). Differences in bacterial saliva profile between periodontitis patients and a control cohort. *J. Clin. Periodontol.* 41 (2), 104–112. doi: 10.1111/jcpe.12190

Bessone, F., Razori, M. V., and Roma, M. G. (2019). Molecular pathways of nonalcoholic fatty liver disease development and progression. *Cell Mol. Life Sci.* 76 (1), 99–128. doi: 10.1007/s00018-018-2947-0

Bui, F. Q., Almeida-da-Silva, C. L. C., Huynh, B., Trinh, A., Liu, J., Woodward, J., et al. (2019). Association between periodontal pathogens and systemic disease. *BioMed. J.* 42 (1), 27–35. doi: 10.1016/j.bj.2018.12.001

Buzzetti, E., Pinzani, M., and Tsochatzis, E. A. (2016). The multiple-hit pathogenesis of non-alcoholic fatty liver disease (NAFLD). *Metabolism* 65 (8), 1038–1048. doi: 10.1016/j.metabol.2015.12.012

Camelo-Castillo, A. J., Mira, A., Pico, A., Nibali, L., Henderson, B., Donos, N., et al. (2015). Subgingival microbiota in health compared toperiodontitis and the influence of smoking. *Front. Microbiol.* 6, 119. doi: 10.3389/fmicb.2015.00119

Canfora, E. E., Meex, R. C. R., Venema, K., and Blaak, E. E. (2019). Gut microbial metabolites in obesity, NAFLD and T2DM. *Nat. Rev. Endocrinol.* 15 (5), 261–273. doi: 10.1038/s41574-019-0156-z

Cao, Y., Qiao, M., Tian, Z., Yu, Y., Xu, B., Lao, W., et al. (2018). Comparative Analyses of Subgingival Microbiome in Chronic Periodontitis Patients with and Without IgA Nephropathy by High Throughput 16S rRNA Sequencing. *Cell Physiol. Biochem.* 47 (2), 774–783. doi: 10.1159/000490029

Chalasani, N., Younossi, Z., Lavine, J. E., Charlton, M., Cusi, K., Rinella, M., et al. (2018). The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. *Hepatology* 67 (1), 328–357. doi: 10.1002/hep.29367

Chattopadhyay, I., Verma, M., and Panda, M. (2019). Role of Oral Microbiome Signatures in Diagnosisand Prognosis of Oral Cancer. *Technol. Cancer Res. Treat* 18, 1533033819867354. doi: 10.1177/1533033819867354

Chen, Z., Qi, J., Wei, Q., Zheng, X., Wu, X., Li, X., et al. (2019). Variations in gut microbial profiles in ankylosing spondylitis: disease phenotype-related dysbiosis. *Ann. Transl. Med.* 7 (20), 571. doi: 10.21037/atm.2019.09.41

Chen, B., Wang, Z., Wang, J., Su, X., Yang, J., Zhang, Q., et al. (2020). The oral microbiome profile and biomarker inChinese type 2 diabetes mellitus patients. *Endocrine* 68 (3), 564–572. doi: 10.1007/s12020-020-02269-6

Dong, L., Yin, J., Zhao, J., Ma, S. R., Wang, H. R., Wang, M., et al. (2018). Microbial Similarity and Preference forSpecific Sites in Healthy Oral Cavity and Esophagus. *Front. Microbiol.* 9, 1603. doi: 10.3389/fmicb.2018.01603

Eslam, M., Newsome, P. N., Sarin, S. K., Anstee, Q. M., Targher, G., Romero-Gomez, M., et al. (2020). A new definition for metabolic dysfunction-associated fatty liver disease: An international expert consensus statement. *J. Hepatol.* 73 (1), 202–209. doi: 10.1016/j.jhep.2020.03.039

Espinoza, J. L., Harkins, D. M., Torralba, M., Gomez, A., Highlander, S. K., Jones, M. B., et al. (2018). Supragingival Plaque Microbiome Ecology andFunctional Potential in the Context of Health and Disease. *mBio*9 (6), e01631–18. doi: 10.1128/mBio.01631-18

Fan, N., Peng, L., Xia, Z., Zhang, L., Song, Z., Wang, Y., et al. (2019). Triglycerides to high-density lipoprotein cholesterol ratio as a surrogate for nonalcoholic fatty liver disease: a cross-sectional study. *Lipids Health Dis.* 18 (1), 39. doi: 10.1111/liv.12977

Fukuda, Y., Hashimoto, Y., Hamaguchi, M., Fukuda, T., Nakamura, N., Ohbora, A., et al. (2016). Triglycerides to high-density lipoproteincholesterol ratio is an independent predictor of incident fatty liver; a population-based cohort study. *Liver Int.* 36 (5), 713–720. doi: 10.1111/liv.12977

Graves, D. T., Correa, J. D., and Silva, T. A. (2019). The Oral Microbiota Is Modified by Systemic Diseases. *J. Dent. Res.* 98 (2), 148–156. doi: 10.1177/0022034518805739

Hintao, J., Teanpaisan, R., Chongsuvivatwong, V., Ratarasan, C., and Dahlen, G. (2007). The microbiological profiles of saliva, supragingival and subgingival plaque and dental caries in adults with and without type 2 diabetes mellitus. *Oral. Microbiol. Immunol.* 22 (3), 175–181. doi: 10.1111/j.1399-302X.2007.00341.x

Ikeda, E., Shiba, T., Ikeda, Y., Suda, W., Nakasato, A., Takeuchi, Y., et al. (2020). Japanese subgingival microbiota in health vs disease and their roles in predicted functions associated with periodontitis. *Odontology* 108 (2), 280–291. doi: 10.1007/s10266-019-00452-4

Jaramillo, A., Lafaurie, G. I., Millán, L. V., Ardila, C. M., Duque, A., Novoa, C., et al. (2013). Association between periodontal disease and plasma levels of cholesterol and triglycerides. *Colomb. Med. (Cali)* 44 (2), 80–86. doi: 10.25100/cm.v44i2.1123

Kelly, T. N., Bazzano, L. A., Ajami, N. J., He, H., Zhao, J., Petrosino, J. F., et al. (2016). Gut Microbiome Associates With Lifetime Cardiovascular Disease Risk Profile Among Bogalusa Heart Study Participants. *Circ. Res.* 119 (8), 956–964. doi: 10.1161/circresaha.116.309219

Kressirer, C. A., Smith, D. J., King, W. F., Dobeck, J. M., Starr, J. R., and Tanner, A. C. R. (2017). Scardovia wiggsiae and its potential role as a caries pathogen. *J. Oral. Biosci.* 59 (3), 135–141. doi: 10.1016/j.job.2017.05.002

LaMonte, M. J., Williams, A. M., Genco, R. J., Andrews, C. A., Hovey, K. M., Millen, A. E., et al. (2014). Association between metabolic syndrome and periodontal disease measures in postmenopausal women: the Buffalo OsteoPerio study. *J. Periodontol.* 85 (11), 1489–1501. doi: 10.1902/jop.2014.140185

Li, B., Ge, Y., Cheng, L., Zeng, B., Yu, J., Peng, X., et al. (2019). Oral bacteria colonize and compete with gutmicrobiota in gnotobiotic mice. *Int. J. Oral. Sci.* 11 (1), 10. doi: 10.1038/s41368-018-0043-9

Li, B. Z., Zhou, H. Y., Guo, B., Chen, W. J., Tao, J. H., Cao, N. W., et al. (2020). Dysbiosis of oral microbiota is associated withsystemic lupus erythematosus. *Arch. Oral. Biol.* 113, 104708. doi: 10.1016/j.archoralbio.2020.104708

Lin, W., Jiang, W., Hu, X., Gao, L., Ai, D., Pan, H., et al. (2018). Ecological Shifts of Supragingival Microbiotain Association with Pregnancy. *Front. Cell. Infect. Microbiol.* 8, 24. doi: 10.3389/fcimb.2018.00024

Loyola-Rodriguez, J. P., Villa-Chavez, C., Patiñio-Marin, N., Aradillas-Garcia, C., Gonzalez, C., and de la Cruz-Mendoza, E. (2011). Association between caries, obesity and insulin resistance in Mexican adolescents. *J. Clin. Pediatr. Dent.* 36 (1), 49–53. doi: 10.17796/jcpd.36.1.e25411r576362262

Maciel, S. S., Feres, M., Goncalves, T. E., Zimmermann, G. S., da Silva, H. D., Figueiredo, L. C., et al. (2016). Does obesity influence the subgingival microbiota composition in periodontal health and disease? *J. Clin. Periodontol.* 43 (12), 1003–1012. doi: 10.1111/jcpe.12634

Matthews, D. R., Hosker, J. P., Rudenski, A. S., Naylor, B. A., Treacher, D. F., and Turner, R. C. (1985). Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* 28 (7), 412–419. doi: 10.1007/bf00280883

Mervish, N. A., Hu, J., Hagan, L. A., Arora, M., Frau, C., Choi, J., et al. (2019). Associations of the Oral Microbiota withObesity and Menarche in Inner City Girls. *J. Child Obes.* 4 (1), 2. doi: 10.21767/2572-5394.100068

Meuric, V., Le Gall-David, S., Boyer, E., Acuna-Amador, L., Martin, B., Fong, S. B., et al. (2017). Signature of Microbial Dysbiosis inPeriodontitis. *Appl. Environ. Microbiol.* 83 (14), e00462-17. doi: 10.1128/aem.00462-17

Muramatsu, Y., Haraya, N., Horie, K., Uchida, L., Kooriyama, T., Suzuki, A., et al. (2019). Bergeyella zoohelcum isolated from oral cavities of therapy dogs. *Zoonoses Public Health* 66 (8), 936–942. doi: 10.1111/zph.12644

Pacifico, L., Bonci, E., Andreoli, G., Romaggioli, S., Di Miscio, R., Lombardo, C. V., et al. (2014). Association of serum triglyceride-to-HDL cholesterol ratio with carotid artery intima-media thickness, insulin resistance and nonalcoholic fatty liver disease in children and adolescents. *Nutr. Metab. Cardiovasc. Dis.* 24 (7), 737–743. doi: 10.1016/j.numecd.2014.01.010

Perez-Chaparro, P. J., McCulloch, J. A., Mamizuka, E. M., Moraes, A., Faveri, M., Figueiredo, L. C., et al. (2018). Do different probing depths exhibit striking differences in microbial profiles? *J. Clin. Periodontol.* 45 (1), 26–37. doi: 10.1111/jcpe.12811

Polak, D., Shany-Kdoshim, S., Zaydel, L., Feuerstein, O., and Houri-Haddad, Y. (2019). High-resolution novel method for tracking bacteria in a multi-species biofilm. *Arch. Microbiol.* 201 (2), 259–266. doi: 10.1007/s00203-018-1614-z

Ratzke, C., Barrere, J., and Gore, J. (2020). Strength of species interactions determines biodiversity and stability in microbial communities. *Nat. Ecol. Evol.* 4 (3), 376–383. doi: 10.1038/s41559-020-1099-4

Rocas, I. N., and Siqueira, J. F. Jr. (2005). Species-directed 16S rRNA gene nested PCR detection of Olsenella species in association with endodontic diseases. *Lett. Appl. Microbiol.* 41 (1), 12–16. doi: 10.1111/j.1472-765X.2005.01723.x

Sasaki, N., Katagiri, S., Komazaki, R., Watanabe, K., Maekawa, S., Shiba, T., et al. (2018). Endotoxemia by Porphyromonas gingivalisInjection Aggravates Non-alcoholic Fatty Liver Disease, Disrupts Glucose/Lipid Metabolism, andAlters Gut Microbiota in Mice. *Front. Microbiol.* 9, 2470. doi: 10.3389/fmicb.2018.02470

Sasaki, M., Shimoyama, Y., Ishikawa, T., Kodama, Y., Tajika, S., and Kimura, S. (2020). Contribution of different adherent properties of Granulicatella adiacens and Abiotrophia defectiva to their associations with oral colonization and the risk of infective endocarditis. *J. Oral. Sci.* 62 (1), 36–39. doi: 10.2334/josnusd.19-0021

Segata, N., Haake, S. K., Mannon, P., Lemon, K. P., Waldron, L., Gevers, D., et al. (2012). Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol.* 13 (6), R42. doi: 10.1186/gb-2012-13-6-r42

Sheka, A. C., Adeyi, O., Thompson, J., Hameed, B., Crawford, P. A., and Ikramuddin, S. (2020). Nonalcoholic Steatohepatitis: A Review. *JAMA* 323 (12), 1175–1183. doi: 10.1001/jama.2020.2298

Si, J., Lee, C., and Ko, G. (2017). Oral Microbiota: Microbial Biomarkers of Metabolic Syndrome Independent of Host Genetic Factors. *Front. Cell. Infect. Microbiol.* 7, 516. doi: 10.3389/fcimb.2017.00516

Sun, Y., Chen, Q., Lin, P., Xu, R., He, D., Ji, W., et al. (2019). Characteristics of Gut Microbiota in PatientsWith Rheumatoid Arthritis in Shanghai, China. *Front. Cell. Infect. Microbiol.* 9, 369. doi: 10.3389/fcimb.2019.00369

Takeshita, T., Kageyama, S., Furuta, M., Tsuboi, H., Takeuchi, K., Shibata, Y., et al. (2016). Bacterial diversity in saliva and oralhealth-related conditions: the Hisayama Study. *Sci. Rep.* 6, 22164. doi: 10.1038/srep22164

Tam, J., Hoffmann, T., Fischer, S., Bornstein, S., Grassler, J., and Noack, B. (2018). Obesity alters composition and diversity of the oral microbiota in patients with type 2 diabetes mellitus independently of glycemic control. *PloS One* 13 (10), e0204724. doi: 10.1371/journal.pone.0204724

Tang, Q., Li, X., Song, P., and Xu, L. (2015). Optimal cut-off values for the homeostasis model assessment of insulin resistance (HOMA-IR) and pre-diabetes screening: Developments in research and prospects for the future. *Drug Discov. Ther.* 9 (6), 380–385. doi: 10.5582/ddt.2015.01207

Tanner, A. C. (2015). Anaerobic culture to detect periodontal and caries pathogens. *J. Oral. Biosci.* 57 (1), 18–26. doi: 10.1016/j.job.2014.08.001

Tiniakos, D. G., Vos, M. B., and Brunt, E. M. (2010). Nonalcoholic fatty liver disease: pathology and pathogenesis. *Annu. Rev. Pathol.* 5, 145–171. doi: 10.1146/annurev-pathol-121808-102132

Tong, Y., Zheng, L., Qing, P., Zhao, H., Li, Y., Su, L., et al. (2019). Oral Microbiota Perturbations Are Linked toHigh Risk for Rheumatoid Arthritis. *Front. Cell. Infect. Microbiol.* 9, 475. doi: 10.3389/fcimb.2019.00475

Tremaroli, V., and Bäckhed, F. (2012). Functional interactions between the gut microbiota and host metabolism. *Nature* 489 (7415), 242–249. doi: 10.1038/nature11552

Utter, D. R., Mark Welch, J. L., and Borisy, G. G. (2016). Individuality, Stability, and Variability of thePlaque Microbiome. *Front. Microbiol.* 7, 564. doi: 10.3389/fmicb.2016.00564

Vieira Colombo, A. P., Magalhaes, C. B., Hartenbach, F. A., Martins do Souto, R., and Maciel da Silva-Boghossian, C. (2016). Periodontal-disease-associated biofilm: A reservoir for pathogens of medical importance. *Microb. Pathog.* 94, 27–34. doi: 10.1016/j.micpath.2015.09.009

Wei, Y. S., Hsiao, Y. C., Su, G. W., Chang, Y. R., Lin, H. P., Wang, Y. S., et al. (2020). Identification of hyperglycemia-associatedmicrobiota alterations in saliva and gingival sulcus. *Arch. Biochem. Biophys.* 682, 108278. doi: 10.1016/j.abb.2020.108278

Yamashita, Y., and Takeshita, T. (2017). The oral microbiome and human health. *J. Oral. Sci.* 59 (2), 201–206. doi: 10.2334/josnusd.16-0856

Yoneda, M., Naka, S., Nakano, K., Wada, K., Endo, H., Mawatari, H., et al. (2012). Involvement of a periodontal pathogen, Porphyromonas gingivalis on the pathogenesis of non-alcoholic fatty liver disease. *BMC Gastroenterol.* 12, 16. doi: 10.1186/1471-230x-12-16

Zhao, F., Dong, T., Yuan, K.-Y., Xia, F.-Z., Liu, D., Wang, Z.-M., et al. (2020). Shifts in the bacterial community of supragingival plaque related to metabolic associated fatty liver disease. *Preprint*. doi: 10.21203/rs.3.rs-45886/v1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cellular and Infection Microbiology

ORIGINAL RESEARCH
published: 27 January 2021
doi: 10.3389/fcimb.2020.616814

Tonsil Mycobiome in PFAPA (Periodic Fever, Aphthous Stomatitis, Pharyngitis, Adenitis) Syndrome: A Case-Control Study

Mysore V. Tejesvi[1,2,3], Terhi Tapiainen[2,4,5], Petri Vänni[3,4], Matti Uhari[2,4], Marko Suokas[1], Ulla Lantto[4,6], Petri Koivunen[4,6] and Marjo Renko[4,7]*

[1] Ecology and Genetics, Faculty of Science, University of Oulu, Oulu, Finland, [2] Biocenter Oulu, University of Oulu, Oulu, Finland, [3] Genobiomics LLC, Oulu, Finland, [4] PEDEGO Research Unit, University of Oulu, Oulu, Finland, [5] Department of Paediatrics and Adolescent Medicine, Oulu University Hospital, Oulu, Finland, [6] Department of Otorhinolaryngology, Oulu University Hospital, Oulu, Finland, [7] Department of Paediatrics, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland

OPEN ACCESS

Edited by:
Thuy Do, University of Leeds, United Kingdom

Reviewed by:
Donato Rigante, Catholic University of the Sacred Heart, Italy
Ozgur Kasapcopur, Istanbul University-Cerrahpasa, Turkey
Natasa Toplak, Univerzitetnega Kliničnega Centra Ljubljana, Slovenia
Kalpana Manthiram, National Institutes of Health (NIH), United States

***Correspondence:**
Mysore V. Tejesvi
mysore.tejesvi@oulu.fi;
mvtejesvi@gmail.com

Specialty section:
This article was submitted to Microbiome in Health and Disease, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: *13 October 2020*
Accepted: *08 December 2020*
Published: *27 January 2021*

Citation:
Tejesvi MV, Tapiainen T, Vänni P, Uhari M, Suokas M, Lantto U, Koivunen P and Renko M (2021) Tonsil Mycobiome in PFAPA (Periodic Fever, Aphthous Stomatitis, Pharyngitis, Adenitis) Syndrome: A Case-Control Study. Front. Cell. Infect. Microbiol. 10:616814. doi: 10.3389/fcimb.2020.616814

Periodic fever, aphthous stomatitis, pharyngitis and adenitis syndrome (PFAPA) is the most common periodic fever syndrome in children with unknown etiology, effectively treated with tonsillectomy. Earlier we have shown that tonsil microbiome is different in patients with PFAPA as compared to that in controls. Recently, fungal microbiome, mycobiome, has been linked to the pathogenesis of inflammatory diseases. We now investigated the role of mycobiome of tonsils in PFAPA. Random forest classification, a machine learning approach, was used for the analysis of mycobiome data. We examined tonsils from 30 children with PFAPA and 22 control children undergoing tonsillectomy for non-infectious reasons. We identified 103 amplicon sequence variants, mainly from two fungal phyla, Ascomycota and Basidiomycota. The mean relative abundance of *Candida albicans* in the tonsil mycobiome was 11% (95% CI: 19 to 27%) in cases and 3.4 % (95% CI: -0.8% to 8%) in controls, p =0.104. Mycobiome data showed no statistical difference in differentiating between PFAPA cases and controls compared to a random chance classifier (area under the curve (AUC) = 0.47, SD = 0.05, p = 0.809). In conclusion, in this controlled study, tonsillar mycobiome in children with PFAPA syndrome did not differ from that of the controls.

Keywords: mycobiome, tonsil, PFAPA, machine learining, next generation sequencing

INTRODUCTION

Periodic fever, aphthous stomatitis, pharyngitis and adenitis (PFAPA) is a childhood febrile syndrome of unknown origin in which fever flares occur in regular 3- to 5-week cycles. Between febrile episodes, patients are asymptomatic (Marshall et al., 1989; Thomas et al., 1999). Although PFAPA syndrome has been suggested to be an autoinflammatory disorder due to dysregulated cytokine production in inflammasomes (Brown et al., 2011; Stojanov et al., 2011; Kolly et al., 2013), its etiology remains unknown. PFAPA is likely a polygenic or complex genetic disease and more

recognized in adult patients (Cantarini et al., 2012; Adrovic et al., 2019; Manthiram et al., 2020). Even though randomized controlled studies have shown that tonsillectomy (TE) is a curative treatment for PFAPA syndrome (Renko et al., 2007; Garavello et al., 2011), the mechanism of this effect remains unclear. Earlier we have shown that tonsil microbiome is different in patients with PFAPA as compared to that in controls (Tejesvi et al., 2016).

In healthy hosts, the host and commensal microbiomes are characterized by interaction and homeostasis (Levy et al., 2015). Earlier, microbiome research has mostly focused on the impact of the bacterial microbiome, referred to as bacteriome, on health (Iliev et al., 2012; Tejesvi et al., 2016; Sokol et al., 2017). However, recent research has drawn attention to the importance of host-associated fungi, the mycobiome, in the inflammatory processes of the human body (Ward et al., 2017). Changes in the mycobiome have been associated with the modulation of autoinflammatory immune responses and disease progression, for example, in Crohn's disease (El Mouzan et al., 2018). Furthermore, the mycobiome may be involved in the host immune response and constitute a risk factor for immunological disorders (Kumamoto, 2016). Finally, it may function as a reservoir of opportunistic pathogens, such as *Candida albicans*, in immunocompromised patients (Polvi et al., 2015; Huseyin et al., 2017).

The human gut mycobiome has been associated with chronic inflammatory diseases of the gut, with studies focusing on intestinal fungi (Iliev et al., 2012; Sokol et al., 2017). However, data on oral mycobiome is scarce. In this controlled study, our main objective was to investigate the role of the mycobiome in tonsils as a potential trigger of inflammatory responses in PFAPA syndrome using next-generation fungal microbiome sequencing technology. Furthermore, random forest classification, a machine learning approach, was used for mycobiome data in classifying PFAPA cases and controls.

MATERIALS AND METHODS

Recruitment of the Patients and Controls

Between March 2006 and April 2010, we recruited 30 consecutive patients (median age: 3.4 years; **Table 1**) who underwent TE due to PFAPA. The diagnostic criteria; i.e. at least five episodes of high fever of unknown origin recurring with a typical, regular pattern and asymptomatic intervals of 2 to 5 weeks, were the same as in our previous randomized controlled study on TE in PFAPA (Renko et al., 2007). During the same period, 22 children (median age: 5.6 years) undergoing TE due to hypertrophied tonsils were recruited as controls.

Demographics of the Patients and Controls

Data on the patients' symptoms were collected before surgery using a questionnaire. The median age of the PFAPA patients at the onset of the fever periods was 2.3 years. The average duration of PFAPA symptoms before TE was 12 months. The mean maximum fever was 39.7°C, and the mean duration of the febrile episodes was 3.9 days. The mean time interval between two subsequent febrile episodes (start to start) was 26 days. None of the patients received steroids for PFAPA prior to TE: The median age at the time of surgery was 3.4 years in the PFAPA group and 5.6 years in the control group. In all PFAPA patients, the symptoms were resolved after TE.

We obtained data on the children's use of antimicrobials in the 12 months before tonsillectomy from the Finnish National Drug Purchase Register, maintained by the Social Insurance Institution of Finland (Kela). Exposure to antimicrobials in the 12 months before tonsillectomy was greater in the PFAPA group (mean number of antimicrobial courses: 2.5) than in the control group (1.3). However, the difference was not statistically significant (95% confidence interval: −0.1, 2.4; p = 0.07; **Table 2**).

The parents of all patients provided written informed consent. This study's protocol was approved by the Ethics Committee of the Northern Ostrobothnia Hospital District, Oulu, Finland. All methods were carried out following relevant guidelines and regulations.

The Samples and DNA Extraction

A total of 30 tonsil samples from PFAPA patients and 22 control samples were stored at −80°C and later used for mycobiome and machine learning analyses. The microbiology of these samples, as well as the details of the patients and the operations, have been described previously (Lantto et al., 2015; Tejesvi et al., 2016). All microbiological analyses were performed blinded to indications for TE.

For the mycobiome analyses, we isolated DNA from the 54 tonsil samples, from 30 cases and 24 controls. DNA was extracted from the tonsil samples using a Quick-DNA Fungal/Bacterial Miniprep Kit (Zymo Research, USA) according to the manufacturer's protocol. The quantity and quality of DNA were determined using a NanoDrop 1000 spectrophotometer (Thermo Scientific, Waltham, MA, USA).

TABLE 1 | Demographic characteristics of pharyngitis and adenitis syndrome (PFAPA) cases and controls.

	PFAPA (N = 30)	Controls (N = 22)
Age when symptoms began, median (range), years	2.3 (0.1–16.5)	
Age at the time of surgery, median (range), years	3.4 (1.7–18.2)	5.6 (2.7–15.2)
Gender, boys, N (%)	18 (60%)	8 (36%)
Antimicrobial courses mean (SD)		
Within 12 months prior to TE	2.5 (2.6)	1.3 (1.6)

Data on age at onset of symptoms, time of surgery, and antimicrobial course are presented.

TABLE 2 | The mean relative abundances of most abundant fungal species in pharyngitis and adenitis syndrome (PFAPA) children and controls.

Fungi	PFAPA (%) (N = 30)	Control (%) (N = 22)	p-value[a]
Malassezia spp	11.38	14.37	0.789
Candida_albicans	10.90	3.43	0.190
Malassezia_restricta	9.78	4.16	0.411
Helotiales	6.08	1.66	0.319
Rhexocercosporidium_panacis	3.24	4.43	0.768
Herpotrichiellaceae	2.82	3.20	0.122
Malassezia_globosa	2.07	6.67	0.792
Tomentella_sublilacina	1.69	5.12	0.105
Cladosporium	1.03	1.79	0.871
Suillus_bovinus	0.09	3.10	0.347
Dermateaceae	0.04	4.15	0.076
Masticobasidum	0.00	0.90	0.039

The mean relative abundances are shown as percentages.
[a]*p-values calculated using Mann-Whitney U test.*

Amplification of the Fungal Internal Transcribed Region

The fungal internal transcribed spacer 2 (ITS2) region was amplified using ITS7b and ITS4 primers including an Ion Torrent pyrosequencing adapter with a 10-bp barcode sequence to the ITS4 primer. Polymerase chain reactions (PCR) were performed in three replicates, each containing a 1x Phusion HF buffer, 0.4 μM of forward and reverse primers, 200 μM of dNTPs, 0.5 U of Phusion enzyme (Thermo Scientific, Finland) and 20 ng of genomic community DNA as the template and molecular-grade water in a total reaction volume of 20 μl. After an initial denaturation at 98°C for 3 min, the following cycling conditions were used: 38 cycles of 98°C, 10 s; 56°C, 10 s; and 72°C, 20 s. After PCR amplification, the pooled triplicate reactions were purified using an AMPure XP PCR cleanup kit (Agencourt Bioscience, CA, USA) and assessed for DNA size, molarity and quality using an Agilent Bioanalyzer 2100 (Agilent Technologies, CA, USA). Finally, the samples were diluted to equimolar concentrations and sequenced using an Ion 316 Chip Kit v2 with the Ion Torrent PGM platform (Thermo Fisher, Life Technologies, USA).

Mycobiome Analysis

Denoising and amplicon sequence variants (ASV) picking were performed using the default settings of the DADA2 algorithm in QIIME2 (Callahan et al., 2016). Chimeric sequences were removed from the data with the q2-vsearch plugin implementing VSEARCH in QIIME2. Sequences in the data that were not classified into the kingdoms of fungi were removed using a custom Python script. ASVs found only in one sample and had lower than 100 reads across the table were excluded using a feature-table plugin in QIIME2. Two samples had a low number of reads and were excluded from further analysis. Taxonomic classification of fungal sequences was performed using Naive Bayes classifiers with the q2-feature-classifier plugin, trained on an ITS silva database for fungi. Phylogenetic trees were created using the q2-phylogeny plugin, which utilizes MAFFT and FastTree (Katoh and Standley, 2013). Alpha and beta diversity analyses were performed using a rarefying depth of 650 for fungi with the q2-diversity plugin and visualized with custom scripts using the Matplotlib package for Python (Hunter, 2007). Statistical comparisons between groups were performed with custom scripts using the SciPy package for Python (Oliphant, 2007). The observed ASVs, Pielou's evenness, Shannon's diversity index and Faith's phylogenetic diversity were chosen as alpha diversity metrics. Principal coordinate analyses were performed for Bray-Curtis, Jaccard and weighted and unweighted UniFrac distances and visualized in two dimensions. A Venn diagram was drawn using an online tool at Euler Venn Applet using differential abundance data between controls and PFAFA cases (http://bioinformatics.psb.ugent.be/webtools/Venn/). Pie charts of fungal taxonomy were created using Krona (Ondov et al., 2011).

Machine Learning Analysis

Random forest (Breiman, 2001) classifiers were trained on relative abundance tables collapsed to the genus level to differentiate between control and PFAFA case samples based on fungal data. The classifiers' performance was assessed with out-of-bag accuracy and receiver operating characteristic (ROC) curves averaged over stratified tenfold cross-validation on the whole dataset. For the averaged ROC curve, 95% confidence intervals were calculated using a Bayesian method in the Scipy package (Oliphant, 2007). The feature importance of different ASVs was assessed using mean decrease impurity (MDI), where Gini is the impurity metric, averaged for each cross-validation. The classifiers were used to predict the same test set with real and shuffled labels in each fold. The averaged area under the curve (AUC) from each cross-validation for real and shuffled labels was tested with an independent t-test. The entire process was repeated 100 times, the values were averaged, and the resulting 100 p-values were combined using Fisher's method. Samples that had zero relative abundance for every variable were excluded. The machine learning analysis was performed with custom Python scripts using the scikit-learn package (Pedregosa et al., 2011), and figures were plotted with Matplotlib 3.2.1 (Hunter, 2007).

Statistical Analyses

For sample size calculation we anticipated to find out a 40% difference in the presence of *C. albicans* in the tonsils between the groups. With an alfa error of 5% and power of 80%, we calculated a sample size of 24 patients per group to be needed.

We used Student's t-test for independent samples to compare the mean number of observed ASVs and the means of the indices describing the diversity of the mycobiome between PFAFA patients and controls. We calculated the means (or medians) with their standard deviations (SDs) (or ranges) of the relative abundances of fungal phyla, the most abundant genera and selected genera in each group. The statistical significances of the differences were tested using the nonparametric method with the Mann-Whitney U test. The analyses were performed using IBM SPSS Statistics 25 (IBM, Armonk, NY, USA).

RESULTS

General Description of the Mycobiome in Tonsil Samples

Across all samples, we identified 103 amplicon sequence variants (ASVs) from two major fungal phyla, Ascomycota and Basidiomycota, and some unclassified reads. Sequence reads belonging to the phyla Basidiomycota and Ascomycota were present in all samples. The number of ASVs per sample ranged from 2 to 18. The most abundant genera belonging to the phylum Ascomycota were *Candida*, *Gyoerffyella*, *Meliniomyces* and *Rhexocercosporidium*, while those belonging to the phylum Basidiomycota were *Malassezia*, *Telephora*, *Suillus* and *Rhodotorula* (**Figure 1** and **Supplementary Figure 1**). The relative abundance of representative genera, class and family of control and cases are presented in **Table 2**.

COMPARISONS BETWEEN THE TONSIL MYCOBIOMES OF PERIODIC FEVER, APHTHOUS STOMATITIS, PHARYNGITIS, AND ADENITIS SYNDROME CASES AND CONTROLS

There were no significant differences in the number of ASVs or the alpha diversity indices between tonsil samples from PFAPA cases and controls (**Figure 2**). Both groups shared 68 unique taxa, while 25 ASVs were present only in the cases and 10 only in the controls (**Figure 3**). The most abundant phyla in both groups were Basidiomycota and Ascomycota. Basidiomycota were more prevalent in cases and Ascomycota in controls; however, the differences were not statistically significant (**Figure 1**). The Bray-Curtis dissimilarity, Jaccard distance and weighted and unweighted UniFrac distances showed slight clustering in the top two respective principal coordinates. However, they did not differentiate between PFAPA and controls (**Figure 4**).

Microbial Diversity and Comparison of *Candida albicans* Abundance

At the species level, *C. albicans* was present in 43% of the cases and 27% of the controls (p = 0.235). The mean relative abundance of *Candida albicans* was 11% (95% CI: 19 to 27%) in cases and 3.4 % (95% CI: -0.8% to 8%) in controls, p =0.104. There were no statistically significant differences in the mean

FIGURE 1 | The relative abundance of fungal mycobiome community in 52 tonsil tissue samples [30 pharyngitis and adenitis syndrome (PFAPA) and 22 controls]. Less abundant (<0.5 %) genera are not shown or are combined.

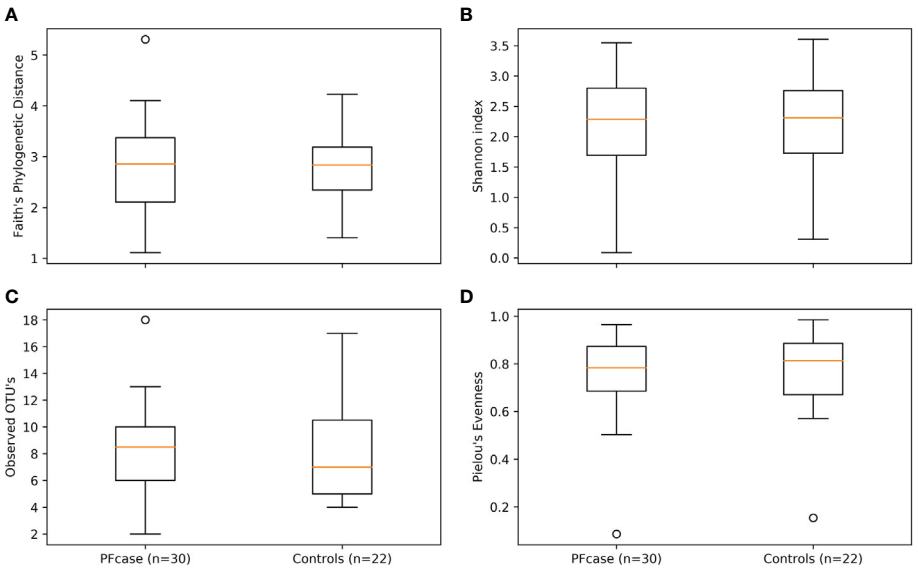


FIGURE 2 | Alpha diversity boxplots of PFcase and control samples. Metrics used are **(A)** Faith's phylogenetic distance, **(B)** Shannon index, **(C)** Observed amplicon sequence variant's (ASV's) and **(D)** Pielou's evenness. Outlier samples are shown as dots.

abundance of other fungal species between the cases and controls. The indices describing alpha diversity and the relative abundance of the phyla did not differ significantly in subjects with or without antibiotic courses during the year before tonsillectomy.

Machine Learning Analysis of the Mycobiome

A classifier was trained to differentiate cases from controls according to the mycobiome data collapsed to the genus level using the silva database. The random classifier Random forest differentiation between PFAPA cases and controls showed no statistical difference in performance compared to a random chance classifier (area under the curve (AUC) = 0.47, SD = 0.05; p = 0.809; **Figure 5**).

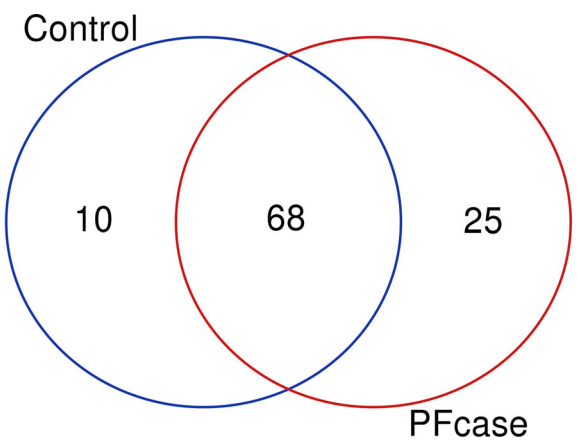


FIGURE 3 | Venn diagram illustrating overlap and number of unique taxa between controls and PFcases. The red and green circles represent PFcases and controls respectively.

DISCUSSION

We have earlier shown that tonsil bacterial microbiome composition is associated with PFAPA syndrome (Tejesvi et al., 2016). In the previous study, only bacterial microbiome was investigated. We hypothesized that mycobiome may associate with the pathogenesis of PFAPA as well. In this controlled study, we showed tonsil mycobiome and its composition in PFAPA and controls. However, we could not confirm the association of mycobiome with the pathogenesis of PFAPA syndrome. Machine learning analysis performed on mycobiome data did not classify PFAPA cases and controls.

PFAPA has been suggested to be an autoinflammatory disease, a condition characterized by abnormally high or uncontrolled inflammation (Wekell et al., 2016). *Candida albicans* is one of the strongest and best verified triggers of inflammasome activity (Joly et al., 2009; Rehaume et al., 2010). It can stimulate inflammasomes, especially in hyphenal forms (Joly et al., 2009). It has also been suggested to play a role in the pathogenesis of inflammatory gut diseases (Sokol et al., 2017). Using culture-based methods and electron microscopy, we previously reported more culture-positive *C. albicans* findings and biofilm formation in tonsillar tissue of PFAPA patients than in tissue of controls (Lantto et al., 2015). In another case-control study, we found that PFAPA patients report clinical oral thrush, an oral fungal infection, in their medical histories more frequently than healthy controls (Lantto et al., 2018). In this study, we hypothesized that the excessive presence of *C. albicans* observed in our previous culture-based study would be even

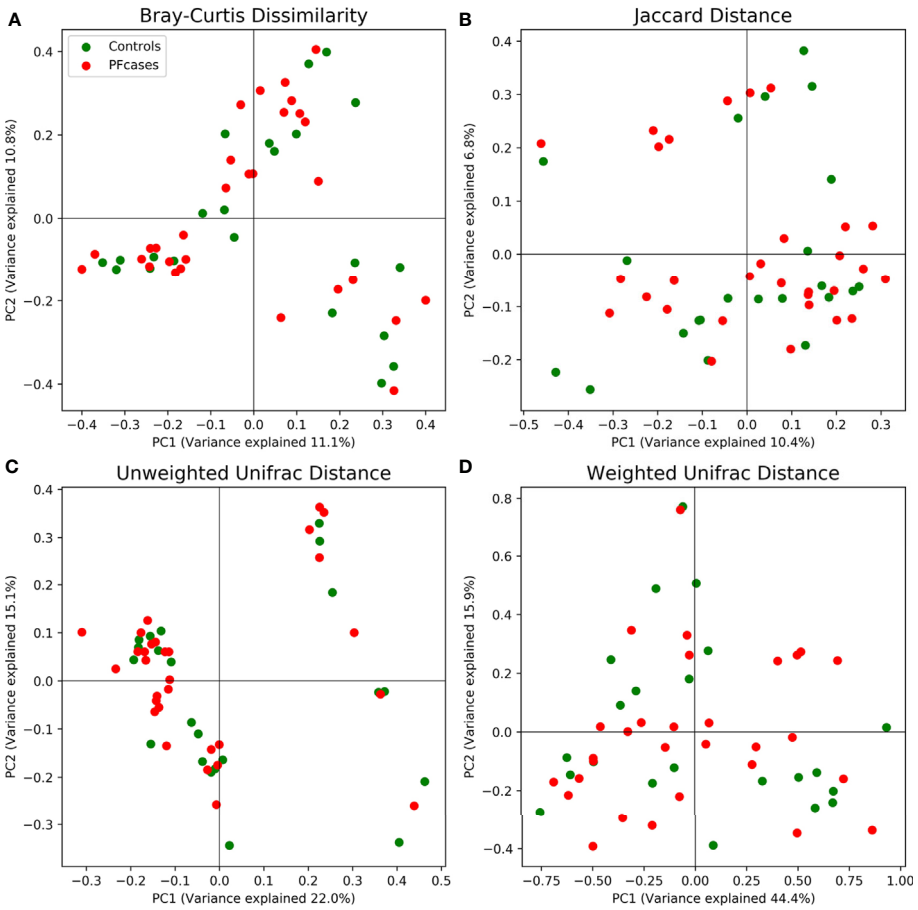


FIGURE 4 | Principal coordinate analysis results plotted for two of the most variance explaining components. Red and green dots represent PFcases and controls, respectively. In **(A)** Bray-Curtis dissimilarity, **(B)** Jaccard distance, **(C)** unweighted Unifrac distance and **(D)** weighted Unifrac distance are shown.

more evident with modern sequencing and machine learning techniques (Rehaume et al., 2010; Wekell et al., 2016; Kocheturov et al., 2019). *Candida albicans* was indeed more abundant in the tonsil mycobiome of PFAPA patients than of controls, but the difference was not statistically significant. Thus, we could not confirm the hypothesis with our sample size; however, the observed difference warrants further studies. Notably, a fluctuation in the inflammatory response might result in a fluctuation in the abundance of microbial triggers as well.

Using traditional statistical models, it is difficult to identify the fungal populations associated with a disease. There is moreover significant variation in microbiome structures between individuals (Walters et al., 2014). To overcome these challenges, machine learning analysis is now being widely used (Johnson et al., 2016; Ai et al., 2017). In our earlier studies on the pathogenesis of PFAPA, we used conventional microbiology of tonsils (Lantto et al., 2015) and then microbiome analysis with conventional statistics (Tejesvi et al., 2016). In this study, we used a machine learning approach and our results does not show a significant difference in discerning PFAPA cases between machine learning analysis and random chance.

Before the next-generation sequencing era, mycobiome was traditionally investigated by *in vitro* culturing. Significant discrepancies in culture-dependent and independent methods are likely, as most fungi are not cultivable in laboratory media (Chen et al., 2011; Browne et al., 2016). Recent studies using next-generation sequencing technology have overcome the bias of culture-dependent methods but have mostly focused on the bacteriome. Recent studies have investigated the role of the human mycobiome in the pathogenesis of gastrointestinal diseases in immunocompromised hosts, diabetes and obesity, revealing the functional diversity of fungi associated with different human body sites (Mukherjee et al., 2014; Kowalewska et al., 2016).

The mycobiome may vary in different parts of the tonsillar tissue, namely, in the superficial layers and the crypts. However, we were unable to determine whether samples for DNA extraction had been taken from superficial layers or crypts.

Fungal dysbiosis and homeostasis are dynamic processes that are probably more common than actual fungal infections and therefore continually shape the immune response (Iliev and Leonardi, 2017). Diversity in the oral mycobiota is lower than

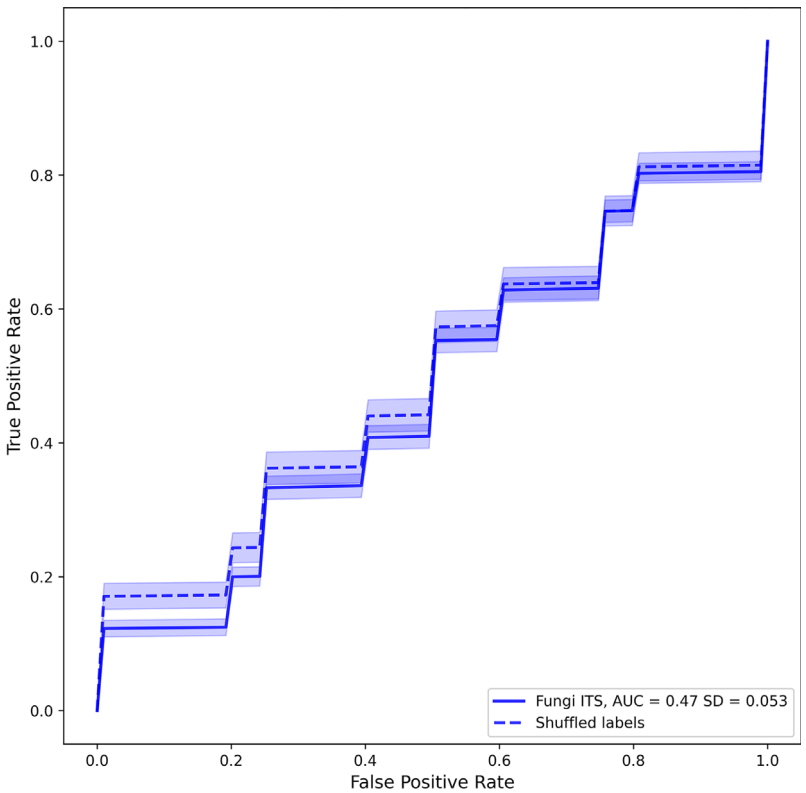


FIGURE 5 | Averaged receiver operating characteristic curves of fungal classifiers trained to differentiate positives (PFcase samples) from negatives (control samples). Solid and dashed lines represent real and shuffled labels. Transparent areas around each line represent the 95% confidence intervals of the curve.

in the bacteriome. In healthy adults, the oral mycobiota is dominated by members of the Ascomycota phylum, mainly *Candida* species (Ghannoum et al., 2010; Bandara et al., 2019). These phyla are also dominant in other parts of the human body (Nash et al., 2017). In a culture-based study, about 12% of infants were found to be oral carriers of *Candida* species (Stecksen-Blicks et al., 2015). However, the oral mycobiome of small children has not been studied with modern techniques. *Candida albicans* is more prevalent in oral samples of patients with rheumatoid arthritis (Bishu et al., 2014) and periodontal disease than in control samples (Peters et al., 2017).

Th17 cells are known to participate in host defense against fungi and extracellular bacteria, and their role in maintaining homeostasis between commensal microorganisms and the host has also been studied (Zielinski, 2014). In a study on the differentiation of Th17 cells *in vivo*, *C. albicans* induced pro-inflammatory Th17 cells that produced IL-1b. However, these cells were incapable of self-regulatory IL-10 production. In our previous culture-based study, tonsil samples from PFAPA patients yielded *C. albicans* more often than those from controls. These results may indicate an important role of the tonsil microbiota in the pathogenesis of PFAPA syndrome.

The role of the mycobiome in various diseases is still largely unknown. This was the first study to examine the mycobiome of children's tonsils. The strength of this study was its controlled design. One limitation was that it was not possible to collect control samples from healthy children. Thus, we chose to use children with hypertrophic tonsils as controls. Another limitation was that the mean age of the control group was older than that of the patient group. Moreover, the sample size was not large enough to perform analyses stratified by age. Furthermore, exposure to different antimicrobials before tonsillectomy may have affected our microbiological findings. In our series, the cases had more often received antimicrobial courses 12 months before tonsillectomy; however, the difference was not statistically significant.

In conclusion, the tonsil mycobiome of PFAPA children did not statistically differ from that of controls in this study. Thus, we could not confirm that *Candida*, earlier associated with PFAPA in epidemiological and conventional microbiological studies, is a trigger of excessive, fluctuating inflammatory response to the mycobiome in the tonsils in PFAPA. The tonsil mycobiome is less diverse and candida is the major genera present in the tonsils. Machine learning performed on mycobiome data did not classify PFAPA cases. However, future studies with a larger sample size may classify and accurately predict PFAPA cases.

DATA AVAILABILITY STATEMENT

We have deposited the Ion Torrent fungal raw data in the Sequence Read Archive (SRA) with accession number SRP132771.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Northern Ostrobothnia Hospital District, Oulu, Finland. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

MR, TT, UL, and PK planned the clinical study and tonsil sample collection. MT, TT, PV, and MR wrote the manuscript. MT did the laboratory work, and MS sequenced the mycobiome. MT and PV performed the bioinformatics analysis. MU critically reviewed the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Society for Pediatric Research and the Finnish Medical Foundation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2020.616814/full#supplementary-material

SUPPLEMENTARY FIGURE 1 | The relative abundance of most abundant taxa in the control and PFcase.

REFERENCES

Adrovic, A., Sahin, S., Barut, K., and Kasapcopur, O. (2019). Familial Mediterranean fever and periodic fever, aphthous stomatitis, pharyngitis, and adenitis (PFAPA) syndrome: shared features and main differences. *Rheumatol. Int.* 39, 29–36. doi: 10.1007/s00296-018-4105-2

Ai, L., Tian, H., Chen, Z., Chen, H., Xu, J., and Fang, J. Y. (2017). Systematic evaluation of supervised classifiers for fecal microbiota-based prediction of colorectal cancer. *Oncotarget* 8, 9546–9556. doi: 10.18632/oncotarget.14488

Bandara, H. M. H. N., Panduwawala, C. P., and Samaranayake, L. P. (2019). Biodiversity of the human oral mycobiome in health and disease. *Oral. Dis.* 25, 363–371. doi: 10.1111/odi.12899

Bishu, S., Su, E. W., Wilkerson, E. R., Reckley, K. A., Jones, D. M., McGeachy, M. J., et al. (2014). Rheumatoid arthritis patients exhibit impaired Candida albicans-specific Th17 responses. *Arthritis Res. Ther.* 16, R50. doi: 10.1186/ar4480

Breiman, L. (2001). Random Forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Brown, K. L., Wekell, P., Karlsson, A., and Berg, S. (2011). On the road to discovery in periodic fever, aphthous stomatitis, pharyngitis and adenitis (PFAPA) syndrome. *Proc. Natl. Acad. Sci. U. S. A.* 108, E525. doi: 10.1073/pnas.1107233108

Browne, H. P., Forster, S. C., Anonye, B. O., Kumar, N., Neville, B. A., Stares, M. D., et al. (2016). Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature* 533, 543–546. doi: 10.1038/nature17645

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., and Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869

Cantarini, L., Vitale, A., Bartolomei, B., Galeazzi, M., and Rigante, D. (2012). Diagnosis of PFAPA syndrome applied to a cohort of 17 adults with unexplained recurrent fevers. *Clin. Exp. Rheumatol.* 30 (2), 269–271.

Chen, Y., Chen, Z., Guo, R., Chen, N., Lu, H., Huang, S., et al. (2011). Correlation between gastrointestinal fungi and varying degrees of chronic hepatitis B virus infection. *Diagn. Microbiol. Infect. Dis.* 70, 492–498. doi: 10.1016/j.diagmicrobio.2010.04.005

El Mouzan, M. I., Korolev, K. S., Al Mofarreh, M. A., Menon, R., Winter, H. S., Al Sarkhy, A. A., et al. (2018). Fungal dysbiosis predicts the diagnosis of pediatric Crohn's disease. *World J. Gastroenterol.* 24, 4510–4516. doi: 10.3748/wjg.v24.i39.4510

Garavello, W., Pignataro, L., Gaini, L., Torretta, S., Somigliana, E., and Gaini, R. (2011). Tonsillectomy in children with periodic fever with aphthous stomatitis, pharyngitis, and adenitis syndrome. *J. Pediatr.* 159, 138–142. doi: 10.1016/j.jpeds.2010.12.014

Ghannoum, M. A., Jurevic, R. J., Mukherjee, P. K., Cui, F., Sikaroodi, M., Naqvi, A., et al. (2010). Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. *PLoS Pathog.* 6, e1000713. doi: 10.1371/journal.ppat.1000713

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* 9, 90–95. doi: 10.1109/MCSE.2007.55

Huseyin, C. E., O'Toole, P. W., Cotter, P. D., and Scanlan, P. D. (2017). Forgotten fungi-the gut mycobiome in human health and disease. *FEMS Microbiol. Rev.* 41, 479–511. doi: 10.1093/femsre/fuw047

Iliev, I. D., and Leonardi, I. (2017). Fungal dysbiosis: immunity and interactions at mucosal barriers. *Nat. Rev. Immunol.* 17, 635–646. doi: 10.1038/nri.2017.55

Iliev, I. D., Funari, V. A., Taylor, K. D., Nguyen, Q., Reyes, C. N., Strom, S. P., et al. (2012). Interactions between commensal fungi and the C-type lectin receptor Dectin-1 influence colitis. *Science* 336, 1314–1317. doi: 10.1126/science.1221789

Johnson, H. R., Trinidad, D. D., Guzman, S., Khan, Z., Parziale, J. V., DeBruyn, J. M., et al. (2016). A Machine Learning Approach for Using the Postmortem Skin Microbiome to Estimate the Postmortem Interval. *PLoS One* 11, e0167370. doi: 10.1371/journal.pone.0167370

Joly, S., Ma, N., Sadler, J. J., Soll, D. R., Cassel, S. L., and Sutterwala, F. S. (2009). Cutting edge: Candida albicans hyphae formation triggers activation of the Nlrp3 inflammasome. *J. Immunol.* 183, 3578–3581. doi: 10.4049/jimmunol.0901323

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kocheturov, A., Pardalos, P. M., and Karakitsiou, A. (2019). Massive datasets and machine learning for computational biomedicine: trends and challenges. *Ann. Operations Res.* 276, 5–34. doi: 10.1007/s10479-018-2891-2

Kolly, L., Busso, N., von Scheven-Gete, A., Bagnoud, N., Moix, I., Holzinger, D., et al. (2013). Periodic fever, aphthous stomatitis, pharyngitis, cervical adenitis syndrome is linked to dysregulated monocyte IL-1beta production. *J. Allergy Clin. Immunol.* 131, 1635–1643. doi: 10.1016/j.jaci.2012.07.043

Kowalewska, B., Zorena, K., Szmigiero-Kawko, M., Waz, P., and Mysliwiec, M. (2016). Higher diversity in fungal species discriminates children with type 1 diabetes mellitus from healthy control. *Patient Prefer. Adherence* 10, 591–599. doi: 10.2147/PPA.S97852

Kumamoto, C. A. (2016). The Fungal Mycobiota: Small Numbers, Large Impacts. *Cell Host Microbe* 19, 750–751. doi: 10.1016/j.chom.2016.05.018

Lantto, U., Koivunen, P., Tapiainen, T., Glumoff, V., Hirvikoski, P., Uhari, M., et al. (2015). Microbes of the tonsils in PFAPA (Periodic Fever, Aphthous stomatitis, Pharyngitis and Adenitis) syndrome - a possible trigger of febrile episodes. *APMIS* 123, 523–529. doi: 10.1111/apm.12383

Lantto, U., Kettunen, S., Tapiainen, T., Koivunen, P., Uhari, M., and Renko, M. (2018). Comorbidity of PFAPA (periodic fever, aphthous stomatitis, pharyngitis and adenitis) patients: a case control study. *Clin. Exp. Rheumatol.* 36 (6 Suppl 115), 129–134.

Levy, M., Thaiss, C. A., and Elinav, E. (2015). Metagenomic cross-talk: the regulatory interplay between immunogenomics and the microbiome. *Genome Med.* 7, 120–129. doi: 10.1186/s13073-015-0249-9

Manthiram, K., Preite, S., Dedeoglu, F., Demir, S., Ozen, S., Edwards, K. M., et al. (2020). Common genetic susceptibility loci link PFAPA syndrome, Behcet's disease, and recurrent aphthous stomatitis. *Proc. Natl. Acad. Sci. U. S. A.* 117, 14405–14411. doi: 10.1073/pnas.2002051117

Marshall, G. S., Edwards, K. M., and Lawton, A. R. (1989). PFAPA syndrome. *Pediatr. Infect. Dis. J.* 8, 658–659. doi: 10.1097/00006454-198909000-00026

Mukherjee, P. K., Chandra, J., Retuerto, M., Sikaroodi, M., Brown, R. E., Jurevic, R., et al. (2014). Oral mycobiome analysis of HIV-infected patients: identification of Pichia as an antagonist of opportunistic fungi. *PLoS Pathog.* 10, e1003996. doi: 10.1371/journal.ppat.1003996

Nash, A. K., Auchtung, T. A., Wong, M. C., Smith, D. P., Gesell, J. R., Ross, M. C., et al. (2017). The gut mycobiome of the Human Microbiome Project healthy cohort. *Microbiome* 5, 153–154. doi: 10.1186/s40168-017-0373-4

Oliphant, T. E. (2007). Python for Scientific Computing. *Comput. Sci. Eng.* 9, 10–20. doi: 10.1109/MCSE.2007.58

Ondov, B. D., Bergman, N. H., and Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinf.* 12, 385–. doi: 10.1186/1471-2105-12-385

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.

Peters, B. A., Wu, J., Hayes, R. B., and Ahn, J. (2017). The oral fungal mycobiome: characteristics and relation to periodontitis in a pilot study. *BMC Microbiol.* 17, 157–159. doi: 10.1186/s12866-017-1064-9

Polvi, E. J., Li, X., O'Meara, T. R., Leach, M. D., and Cowen, L. E. (2015). Opportunistic yeast pathogens: reservoirs, virulence mechanisms, and therapeutic strategies. *Cell Mol. Life Sci.* 72, 2261–2287. doi: 10.1007/s00018-015-1860-z

Rehaume, L. M., Jouault, T., and Chamaillard, M. (2010). Lessons from the inflammasome: a molecular sentry linking Candida and Crohn's disease. *Trends Immunol.* 31, 171–175. doi: 10.1016/j.it.2010.01.007

Renko, M., Salo, E., Putto-Laurila, A., Saxen, H., Mattila, P. S., Luotonen, J., et al. (2007). A randomized, controlled trial of tonsillectomy in periodic fever, aphthous stomatitis, pharyngitis, and adenitis syndrome. *J. Pediatr.* 151, 289–292. doi: 10.1016/j.jpeds.2007.03.015

Sokol, H., Leducq, V., Aschard, H., Pham, H. P., Jegou, S., Landman, C., et al. (2017). Fungal microbiota dysbiosis in IBD. *Gut* 66, 1039–1048. doi: 10.1136/gutjnl-2015-310746

Stecksen-Blicks, C., Granstrom, E., Silfverdal, S. A., and West, C. E. (2015). Prevalence of oral Candida in the first year of life. *Mycoses* 58, 550–556. doi: 10.1111/myc.12355

Stojanov, S., Lapidus, S., Chitkara, P., Feder, H., Salazar, J. C., Fleisher, T. A., et al. (2011). Periodic fever, aphthous stomatitis, pharyngitis, and adenitis (PFAPA) is a disorder of innate immunity and Th1 activation responsive to IL-1 blockade. *Proc. Natl. Acad. Sci. U. S. A.* 108, 7148–7153. doi: 10.1073/pnas.1103681108

Tejesvi, M. V., Uhari, M., Tapiainen, T., Pirttila, A. M., Suokas, M., Lantto, U., et al. (2016). Tonsillar microbiota in children with PFAPA (periodic fever, aphthous stomatitis, pharyngitis, and adenitis) syndrome. *Eur. J. Clin. Microbiol. Infect. Dis.* 35, 963–970. doi: 10.1007/s10096-016-2623-y

Thomas, K. T., Feder, H. M., Lawton, A. R., and Edwards, K. M. (1999). Periodic fever syndrome in children. *J. Pediatr.* 135, 15–21. doi: 10.1016/s0022-3476(99)70321-5

Walters, W. A., Xu, Z., and Knight, R. (2014). Meta-analyses of human gut microbes associated with obesity and IBD. *FEBS Lett.* 588, 4223–4233. doi: 10.1016/j.febslet.2014.09.039

Ward, T. L., Knights, D., and Gale, C. A. (2017). Infant fungal communities: current knowledge and research opportunities. *BMC Med.* 15, 30–3z. doi: 10.1186/s12916-017-0802-z

Wekell, P., Karlsson, A., Berg, S., and Fasth, A. (2016). Review of autoinflammatory diseases, with a special focus on periodic fever, aphthous stomatitis, pharyngitis and cervical adenitis syndrome. *Acta Paediatr.* 105, 1140–1151. doi: 10.1111/apa.13531

Zielinski, C. E. (2014). Autoimmunity beyond Th17: GM-CSF producing T cells. *Cell Cycle* 13, 2489–2490. doi: 10.4161/15384101.2014.946377

Conflict of Interest: Authors MT and PV are co-founders of the company Genobiomics LLC, Finland.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cellular and Infection Microbiology

ORIGINAL RESEARCH
published: 09 March 2021
doi: 10.3389/fcimb.2021.643092

Microbiota in Gut, Oral Cavity, and Mitral Valves Are Associated With Rheumatic Heart Disease

Xue-Rui Shi[1†], Bo-Yan Chen[2,3†], Wen-Zhen Lin[2,3], Yu-Lin Li[2,3], Yong-Li Wang[2,3], Yan Liu[2,3], Jing-Juan Huang[1], Wei-Wei Zhang[1], Xiao-Xin Ma[2,3], Shuai Shao[2,3], Ruo-Gu Li[1] and Sheng-Zhong Duan[2,3*]*

[1] Department of Cardiology, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China, [2] Laboratory of Oral Microbiota and Systemic Diseases, Shanghai Ninth People's Hospital, College of Stomatology, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [3] National Clinical Research Center for Oral Diseases, Shanghai Key Laboratory of Stomatology & Shanghai Research Institute of Stomatology, Shanghai, China

Rheumatic heart disease refers to the long-term damage of heart valves and results from an autoimmune response to group A *Streptococcus* infection. This study aimed to analyze the microbiota composition of patients with rheumatic heart disease and explore potential function of microbiota in this disease. First, we revealed significant alterations of microbiota in feces, subgingival plaques, and saliva of the patients compared to control subjects using 16S rRNA gene sequencing. Significantly different microbial diversity was observed in all three types of samples between the patients and control subjects. In the gut, the patients possessed higher levels of genera including *Bifidobacterium* and *Eubacterium*, and lower levels of genera including *Lachnospira*, *Bacteroides*, and *Faecalibacterium*. *Coprococcus* was identified as a super-generalist in fecal samples of the patients. Significant alterations were also observed in microbiota of subgingival plaques and saliva of the patients compared to control subjects. Second, we analyzed microbiota in mitral valves of the patients and identified microbes that could potentially transmit from the gut or oral cavity to heart valves, including *Streptococcus*. Third, we further analyzed the data using random forest model and demonstrated that microbiota in the gut, subgingival plaque or saliva could distinguish the patients from control subjects. Finally, we identified gut/oral microbes that significantly correlated with clinical indices of rheumatic heart disease. In conclusion, patients with rheumatic heart disease manifested important alterations in microbiota that might distinguish the patients from control subjects and correlated with severity of this disease.

Keywords: rheumatic heart disease, microbiota, gut, subgingival plaque, saliva, mitral valves

OPEN ACCESS

Edited by:
Naile Dame-Teixeira, University of Brasilia, Brazil

Reviewed by:
Roney Orismar Sampaio, University of São Paulo, Brazil
Xin Xu, Sichuan University, China

***Correspondence:**
Ruo-Gu Li
13564565961@163.com
Sheng-Zhong Duan
duansz@shsmu.edu.cn

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Microbiome in Health and Disease, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: *17 December 2020*
Accepted: *04 February 2021*
Published: *09 March 2021*

Citation:
Shi X-R, Chen B-Y, Lin W-Z, Li Y-L, Wang Y-L, Liu Y, Huang J-J, Zhang W-W, Ma X-X, Shao S, Li R-G and Duan S-Z (2021) Microbiota in Gut, Oral Cavity, and Mitral Valves Are Associated With Rheumatic Heart Disease. Front. Cell. Infect. Microbiol. 11:643092. doi: 10.3389/fcimb.2021.643092

INTRODUCTION

Rheumatic heart disease (RHD) is a cardiovascular disease characterized by damages to heart valves, triggered by an autoimmune response to group A *Streptococcus* (GAS) infection (Woldu and Bloomfield, 2016). RHD causes an estimated 300,000 death annually with 30–40 million current cases globally and imposes significant disease burden on low-income countries and some indigenous populations (Watkins et al., 2017). The backbone of RHD therapy is penicillin-based treatment for acute rheumatic fever and replacement of severely damaged heart valves, without an effective vaccine (Tompkins et al., 1972; Nishimura et al., 2014; Woldu and Bloomfield, 2016). There has not been any significant advance in recent history of this field (Watkins et al., 2018). Thus, it is imperative to consider novel strategies for better control of RHD.

Human microbiota has been proven to play essential roles in a wide range of diseases including cardiovascular diseases (Tang et al., 2019). Colon harbors the vast majority of commensal bacteria (Belkaid and Hand, 2014), which can influence immune homeostasis, trigger inflammation, and invade extra-intestinal tissues (Hooper et al., 2012; Anhe et al., 2020). Many studies have focused on revealing the role of gut microbial dysbiosis in the development of cardiovascular diseases such as pulmonary arterial hypertension (Kim et al., 2020), unruptured intracranial aneurysms (Li et al., 2020), chronic heart failure (Kummen et al., 2018), and atherosclerosis (Koren et al., 2010). However, the role of the gut microbiota in RHD has not been studied.

The second most complex population of microbes of human body resides in the oral cavity and influences both oral and systemic health (Zhang et al., 2018; Bui et al., 2019). Periodontal pathogens can alter subgingival microbial composition and host–microorganism interactions, leading to local inflammation and subsequent destruction of periodontal tissues (Hajishengallis et al., 2012). The presence of periodontal bacteria DNA in cardiac tissues and atherosclerotic plaques has suggested connections between oral infections and cardiovascular diseases (Koren et al., 2010; Ziebolz et al., 2018). Periodontitis, a common inflammatory oral disease, increases the risk of adverse pregnancy outcomes, atherosclerosis, stroke, rheumatoid arthritis, diabetes, and other systemic diseases (Pihlstrom et al., 2005). However, there is a paucity of information related the potential role of oral microbiota in RHD.

Here, we aimed to investigate the microbiota composition and structure of RHD patients and explore potential function of microbiota in RHD. We first analyzed microbiota of fecal samples, subgingival plaques and saliva by 16S rRNA gene sequencing to detect alterations between RHD patients and control subjects. We then analyzed microbiota of mitral valves and compared the results with gut and oral microbiota in RHD patients to determine microbial connections between mitral valves and gut/oral cavity. Finally, we explored the possibility using microbiota to discriminate RHD patients from control subjects and analyzed correlations between gut/oral microbiota and severity of RHD.

MATERIALS AND METHODS

Study Cohort

A total of 20 RHD patients and 20 age- and sex-matched control subjects were enrolled in this study. RHD patients with symptomatic severe mitral valvular disease and with a history of acute rheumatic fever were enrolled (Reményi et al., 2012; Nishimura et al., 2017). The exclusion criteria were: inflammatory bowel diseases, irritable bowel syndrome, autoimmune diseases, diarrhea, liver diseases, renal diseases, acute infection, smoking, and use of antibiotics or probiotics 3 months before sample collection.

The study protocol was reviewed and approved by the Human Ethics Committee, Shanghai Chest Hospital, Shanghai Jiaotong University and conducted in accordance with the Principles of Good Clinical Practice and the Declaration of Helsinki. Written informed consent was obtained from all the subjects who participated in the study.

Sample Collection and DNA Extraction

Fecal samples, subgingival plaques, and saliva were collected from RHD patients and control subjects. Fecal samples were collected in falcon tubes within 24 h of patients' admission to the hospital and immediately frozen in −80°C. All subjects were instructed to avoid eating, drinking, and use of a toothbrush or mouth rinse 1 hour before sampling of subgingival plaques and saliva. Subgingival plaques were collected using a dental explorer. Saliva was collected and mixed with 2× lysis buffer at a ratio of 1:1. A total of 16 mitral valves were collected during valve replacement surgeries and were immediately frozen in liquid nitrogen under sterile conditions. Samples were kept in sterile containers and stored at −80°C. Bacterial DNA was extracted using Tiangen kits according to the manufacturer's recommendations and stored at −80°C until further analyses.

High-Throughput Sequencing and Processing

The V3/V4 regions of 16S rRNA genes were amplified with specific primers of 338F-806R. The PCR products were quantified with PicoGreen dsDNA Assay Kit (Invitrogen, Carlsbad, USA) and sequenced on Illumina Novaseq PE250 platform to generate paired-end reads (2 × 480 bp).

Procession of the sequencing data was performed on QIIME2 platform. Analysis of sequencing data was based on amplicon sequence variants (ASVs) (Bokulich et al., 2018). After chimera detection, high-quality sequences with 97% similarity were clustered into the same ASV. Classification of ASVs was performed based on the Greengenes Database.

Data Analysis

Richness and α-diversity were measured by Chao1 and Shannon index based on the genus profiles. β-diversity was visualized using principal coordinate analysis (PCoA) based on the Bray–Curtis distances. ZP-plot was used to identify of key module members (Deng et al., 2012). Within-module connectivity (Zi) and among-module connectivity (Pi) were calculated as

previously described (Deng et al., 2012). Microbes were classified into four categories: peripherals (Zi ≤ 2.5, Pi ≤ 0.62), connectors (Zi ≤ 2.5, Pi > 0.62), module hubs (Zi > 2.5, Pi ≤ 0.62), and network hubs (Zi > 2.5, Pi > 0.62) according to the criteria previously set (Deng et al., 2012). Network hubs were considered as super-generalists, which were highly connected within their own modules and among modules (Deng et al., 2012). Raw data of relative abundances of genera were transformed to "$\log_{10}$" and $\text{Log}_{10}0$ was shown as "−8" in figures.

Linear discriminant analysis effect size (LEfSe) was used to identify features that differed between groups. The threshold of the logarithmic LDA score for discriminative features was set to 3.0. Identified taxa were plotted in cladograms. A random forest classifier was trained to distinguish the two groups based on the genera abundance profile of RHD patients and control subjects. The important genera of the classifier were ranked by Gini index. The performance of the classifier model was evaluated by 10-fold cross-validations and further applied to construct receiver operating characteristic (ROC) curve. The cross-validation accuracy was measured as the area under the ROC curve (AUC). Heatmaps were hierarchically clustered to represent the microbe-clinical indices associations based on the Spearman correlation coefficients.

Statistics

α-diversity was compared using non-parametric Kruskal–Wallis tests and *post hoc* Dunn's test with FDR correction. β-diversity was tested by the ANOSIM method. Raw data of relative abundances of genera were compared using non-parametric Mann–Whitney tests. Correlation analysis was performed using Spearman's correlation. *Post hoc* power analysis was used to calculate power with significance level set to 0.05 and effect size based on Chao1. The power was 0.88, 0.18 and 0.93 for gut, subgingival plaques and saliva, respectively. Statistical analyses were performed using SPSS 23.0, QIIME2, R package (V3.5.1), or GPower (V3.1.9.4).

RESULTS

Alterations of Gut Microbiota in Rheumatic Heart Disease Patients

The demographic data for the enrolled RHD patients and control subjects were summarized in **Supplemental Table 1**. To detect gut microbial alterations, we analyzed the microbiota in fecal samples of control subjects and RHD patients using 16S rRNA gene sequencing. Gut microbiota of RHD patients showed significantly lower richness illustrated by Chao1 and α-diversity (within-sample diversity) illustrated by Shannon index compared to those of control subjects (**Figures 1A, B**). Principal coordinate analysis (PCoA) based on Bray–Curtis distance was performed to determine β-diversity (between-sample diversity) of gut microbiota, which demonstrated significant difference between RHD patients and control subjects (**Figure 1C**). To understand interactions among different microbes in gut microbiota of RHD patients, ZP-plot was used to analyze topological roles, and *Coprococcus* was identified as a super-generalist (shown as network hubs) (**Figure 1D**). Among the 15 most abundant genera, relative abundance of *Faecalibacterium* and *Bacteroides* significantly decreased, while that of *Shigella*, *Gemmiger*, *Bifidobacterium*, *Ruminococcus*, *Streptococcus* and *Dorea* significantly increased in gut microbiota of RHD patients compared to control subjects (**Figure 1E**). In addition, relative abundance of *Blauita* showed a trend of increase without statistical significance in gut microbiota of RHD patients (**Figure 1E**). We next employed linear discriminant analysis effect size (LEfSe) to identify taxa that discriminate microbial composition between disease states. LEfSe identified higher levels of taxa including Arthrobacter, *Bifidobacterium*, *Porphyromonas*, *Melissococccus*, *Eubacterium*, *Ruminococcus*, *Dorea*, *Gemmiger*, *etc*., as well as lower levels of taxa such as *Staphylococcus*, *Lachnospira*, *Faecalibacterium*, and *Oxalobacter* in fecal samples of RHD patients (**Figure 1F**). These results demonstrated considerable changes of the gut microbiota in RHD patients compared to control subjects.

Alterations of Subgingival Plaque Microbiota in Rheumatic Heart Disease Patients

Next we analyzed the microbiota in subgingival plaque samples of control subjects and RHD patients using 16S rRNA gene sequencing. Subgingival plaque microbiota showed no differences of richssness and α-diversity between the two groups illustrated by Chao1 and Shannon index (**Figures 2A, B**). Results of PCoA based on Bray–Curtis distance demonstrated moderate difference in β-diversity between RHD patients and control subjects (**Figure 2C**). There was no super-generalist based on modular topological roles in subgingival plaque microbiota of RHD patients (**Figure 2D**). Among the 15 most abundant genera, relative abundance of *Corynebacterium* and *Selenomonas* significantly decreased, while that of *Streptococcus* and *Blautia* significantly increased in subgingival plaque microbiota of RHD patients compared to control subjects (**Figure 2E**). LEfSe identified higher levels of taxa including *Streptococcus*, *Ruminococcus*, *Blautia*, *Dorea*, *Lachnoanaerobaculum*, *Roseburia*, *Gemmiger*, as well as lower levels of taxa such as *Corynebacterium*, *Staphylococcus*, *Lactobacillus* and *Selenomonas* in subgingival plaque samples of RHD patients (**Figure 2F**).

Alterations of Salivary Microbiota in Rheumatic Heart Disease Patients

We further analyzed microbiota in saliva of control subjects and RHD patients using 16S rRNA gene sequencing. Salivary microbiota of RHD patients showed significantly higher richness illustrated by Chao1 and similar α-diversity illustrated by Shannon index compared to those of control subjects (**Figures 3A, B**). Results of PCoA based on Bray–Curtis distance demonstrated significantly different β-diversity between RHD patients and control subjects (**Figure 3C**). There was no super-generalist based on modular topological roles in salivary microbiota of RHD patients (**Figure 3D**). Among the 15 most abundant genera, relative abundance of *Prevotella*, *Haemophilus*, *Veillonella*, *Campylobacter*, and *Actinomyces* significantly decreased, while that of *Streptococcus* and *Rothia* significantly increased in salivary microbiota of RHD

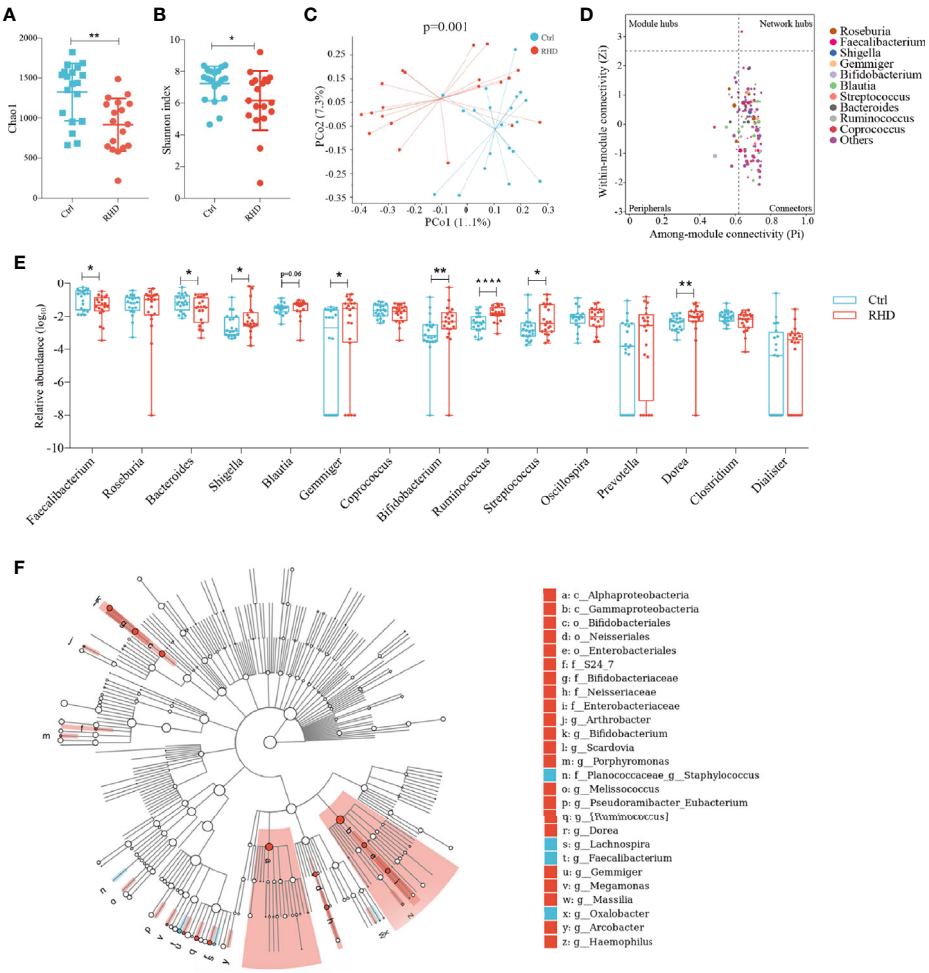


FIGURE 1 | Alterations of gut microbiota in patients with rheumatic heart disease. Microbiota of fecal samples from patients with rheumatic heart disease (RHD) and control subjects (Ctrl) were analyzed using 16S rRNA gene sequencing. **(A)** Richness of fecal microbiota assessed by Chao1. **(B)** α-diversity of fecal microbiota assessed by Shannon index. **(C)** β-diversity analyzed by principal coordinate analysis (PCoA) based on Bray–Curtis distance of fecal microbiota at genus level. **(D)** Determination of module-based topological roles (peripherals, connectors, module hubs, or network hubs) of fecal microbiota in RHD patients using ZP-plot at genus level. The size of dots represents abundance of each genus. **(E)** Relative abundances of the fifteen most abundant genera in fecal microbiota. **(F)** Taxonomic cladogram of fecal microbiota based on LEfSe. Red color indicates increase and blue indicates decrease of taxa in RHD compared to Ctrl. LDA = 3. Mann–Whitney test was used. n = 20: 20. *P < 0.05; **P < 0.01; ****P < 0.0001.

patients compared to control subjects (**Figure 3E**). LEfSe identified higher levels of taxa including *Piscicoccus*, *Rothia*, *Abiotrophia*, *Melissococcus*, *Streptococcus*, *Parvimonas*, *Clostridium*, *Lachnoanaerobaculum*, *etc.*, as well as lower levels of taxa such as *Actinomyces*, *Rhodococcus*, *Prevotella*, *Staphylococcus*, *Veillonella*, *Afipia*, *Sphingomonas*, *etc.* (**Figure 3F**).

Comparisons of Microbiota Composition Between Mitral Valves and Other Body Sites in Rheumatic Heart Disease Patients

Subsequently, we analyzed the microbiota in mitral valves of RHD patients (n = 16) using 16S rRNA gene sequencing. The 15 most abundant genera in mitral valves of RHD patients were *Ralstonia*, *Pelomonas*, *Acinetobacter*, *Neisseria*, *Sphingomonas*, *Streptococcus*, *Agrobacterium*, *Thermus*, *Shigella*, *Rothia*, *Fusobacterium*, *Prevotella*, *Afipia*, *Caulobacter*, and *Burkholderia* (**Supplemental Figure 1A**). There was no super-generalist based on modular topological roles in microbiota of mitral valves of RHD patients (**Supplemental Figure 2**). We next compared microbiota in mitral valves with gut and oral microbiota in these 16 RHD patients. PCoA based on Bray–Curtis distance demonstrated distinct β-diversity among microbiota of mitral valves, gut, and oral cavity, although microbiota of mitral valves in some RHD patients had similar β-diversity with oral microbiota (**Figure 4A**). As expected, β-diversity of subgingival plaque microbiota and salivary microbiota are similar (**Figure 4A**). Taxonomic composition plots at phylum level showed that mitral valves contained significantly more *Proteobacteria* and less *Firmicutes* compared to gut,

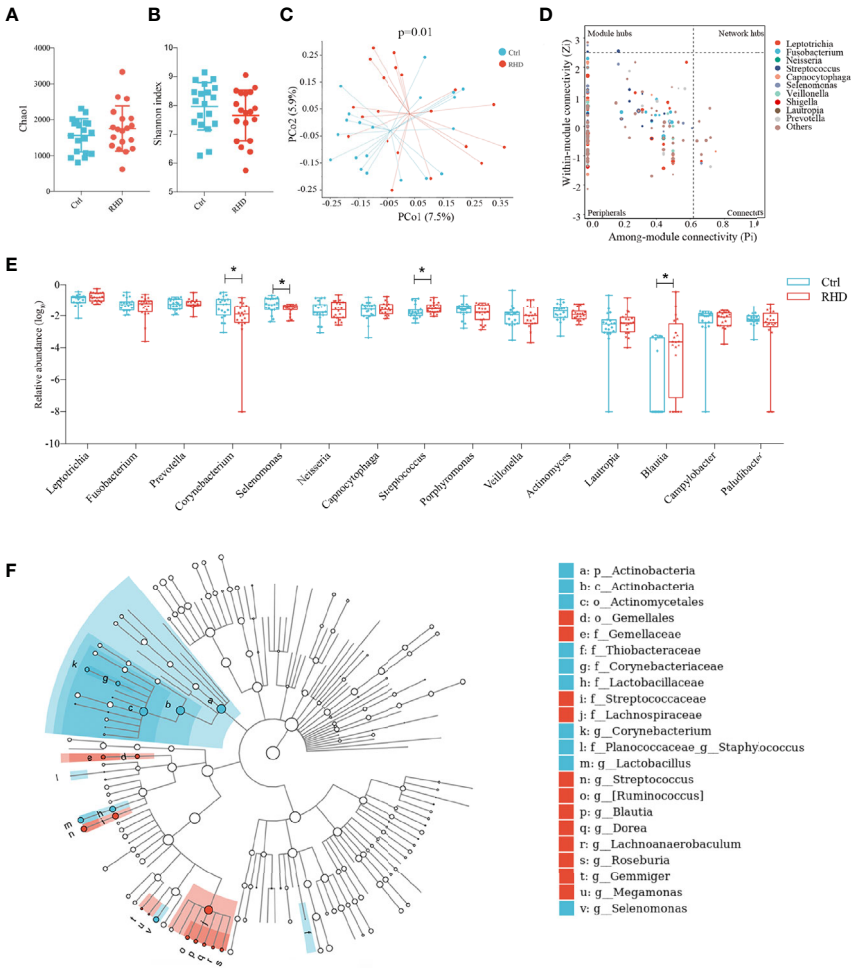


FIGURE 2 | Alterations of subgingival plaque microbiota in patients with rheumatic heart disease. Microbiota of subgingival plaques (SP) from patients with rheumatic heart disease (RHD) and control subjects (Ctrl) were analyzed using 16S rRNA gene sequencing. **(A)** Richness of SP microbiota assessed by Chao1. **(B)** α-diversity of SP microbiota assessed by Shannon index. **(C)** β-diversity analyzed by principal coordinate analysis (PCoA) based on Bray–Curtis distance of SP microbiota at genus level. **(D)** Determination of module-based topological roles (peripherals, connectors, module hubs, or network hubs) of SP microbiota in RHD patients using ZP-plot at genus level. The size of dots represents abundance of each genus. **(E)** Relative abundances of the fifteen most abundant genera in SP microbiota. **(F)** Taxonomic cladogram of SP microbiota based on LEfSe. Red color indicates increase and blue indicates decrease of taxa in RHD compared to Ctrl. LDA = 3. Mann–Whitney test was used. n = 20: 20. *P < 0.05.

subgingival plaques, and saliva (**Figure 4B**), similar to the results of a previous study in atherosclerotic plaques (Koren et al., 2010). We further searched for microbes shared by mitral valves and other body sites using three criteria: (1) detected in mitral valve and at least another body site of an individual RHD patient; (2) detected in mitral valves and another body site of more than eight RHD patients (50%) simultaneously; (3) relative abundance ranked among top 35 (**Supplemental Figure 1**). Mitral valves and fecal samples shared *Streptococcus*, *Shigella*, *Lactobacillus*, and *Bacteroides* in all 16 patients, *Oscillospira* in 15 patients, *Blautia* in 11 patients, and *Prevotella* in nine patients (**Figure 4C**). Mitral valves and subgingival plaques shared *Streptococcus* and *Fusobacterium* in all 16 patients, *Neisseria* and *Shigella* in 15 patients, Prevotella in 13 patients, *Bacteroides* in 12 patients, *Campylobacter* in 10 patients, and *Veillonella* in nine patients (**Figure 4D**). Mitral valves and saliva shared *Streptococcus* and *Fusobacterium* in all 16 patients, *Neisseria* in 15 patients, *Prevotella* in 13 patients, *Campylobacter* in 10 patients, as well as *Rothia* and *Veillonella* in nine patients (**Figure 4E**).

Microbiota Discriminates Rheumatic Heart Disease Patients From Control Subjects

To explore the diagnostic value of microbiota in discriminating RHD patients from healthy controls, we constructed random forest classifiers. According to random forest classifier based on gut microbiota, the top 15 important gut genera were *Ruminococcus*, *Dorea*, *Arthrobacter*, *Lachnospira*, *Shigella*, *Gemmiger*, *Bifidobacterium*, *Ralstonia*, *Faecalibacterium*, *Subdoligranulum*, *Burkholderia*, *Blautia*, *Turicibacter*, *Streptococcus*, and *Selenomonas* (**Supplemental Figure 3A**).

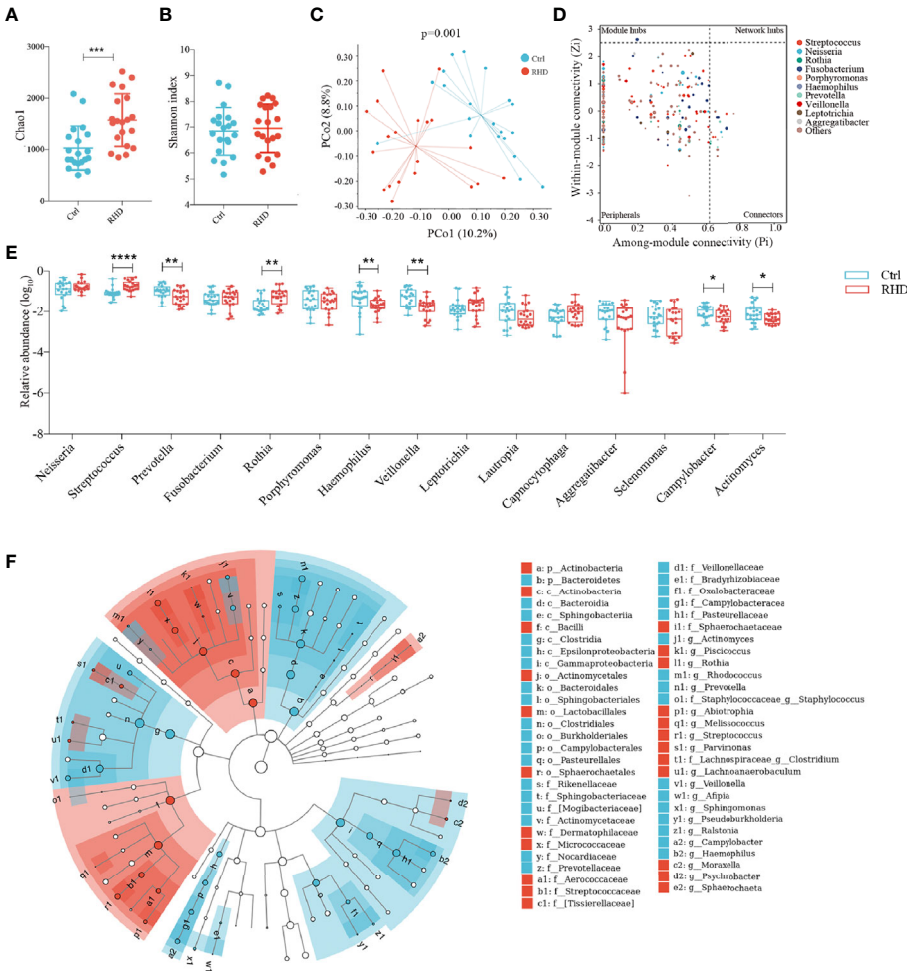


FIGURE 3 | Alterations of salivary microbiota in patients with rheumatic heart disease. Microbiota of saliva from patients with rheumatic heart disease (RHD) and control subjects (Ctrl) were analyzed using 16S rRNA gene sequencing. **(A)** Richness of salivary microbiota assessed by Chao1. **(B)** α-diversity of salivary microbiota assessed by Shannon index. **(C)** β-diversity analyzed by principal coordinate analysis (PCoA) based on Bray–Curtis distance of salivary microbiota at genus level. **(D)** Determination of module-based topological roles (peripherals, connectors, module hubs, or network hubs) of salivary microbiota in RHD patients using ZP-plot at genus level. The size of dots represents abundance of each genus. **(E)** Relative abundances of the fifteen most abundant genera in salivary microbiota. **(F)** Taxonomic cladogram of salivary microbiota based on LEfSe. Red color indicates increase and blue indicates decrease of taxa in RHD compared to Ctrl. LDA = 3. Mann–Whitney test was used. n = 20: 20. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

Area under the curve (AUC) of the gut microbiota was 0.74 (**Figure 5A**). As for subgingival plaque microbiota, the top 15 important subgingival plaque genera were *Roseburia*, *Acinetobacter*, *Turicibacter*, *Bifidobacterium*, *Lactobacillus*, *Leptotrichia*, *Faecalibacterium*, *Lachnoanaerobaculum*, *Clostridium*, *Streptococcus*, *Atopobium*, *Blautia*, *Gemmiger*, *Filifactor*, and *Granulicatella* (**Supplemental Figure 3B**). AUC of subgingival plaque microbiota was 0.89 (**Figure 5B**). In saliva, the top 15 important salivary genera were *Streptococcus*, *Shigella*, *Rothia*, *Pelomonas*, *Lachnoanaerobaculum*, *Parvimonas*, *Arthrobacter*, *Clostridium*, *Abiotrophia*, *Afipia*, *Streptobacillus*, *Ralstonia*, *Turicibacter*, *Cardiobacterium*, and *Haemophilus* (**Supplemental Figure 3C**). AUC of salivary microbiota was 1.00 (**Figure 5C**).

Microbial Taxa Correlate With Clinical Indices in Rheumatic Heart Disease Patients

Mitral valves are the most common heart valves affected by RHD (Russell et al., 2017). Mitral stenosis and regurgitation cause depressed left atrial compliance and elevated left atrial pressure, leading to atrial remodelling and increased left atrial diameter (LAD). Furthermore, pulmonary hypertension is one of the most frequent medical complications in RHD (Watkins et al., 2018). We explored the relationship between these clinical indices and microbial taxa differentially enriched or important to discriminate RHD patients from healthy controls. The results showed that *Bacteroides* and *Eubacterium* in the gut negatively correlated with LAD; *Roseburia* and *Lachnoanaerobaculum* in subgingival

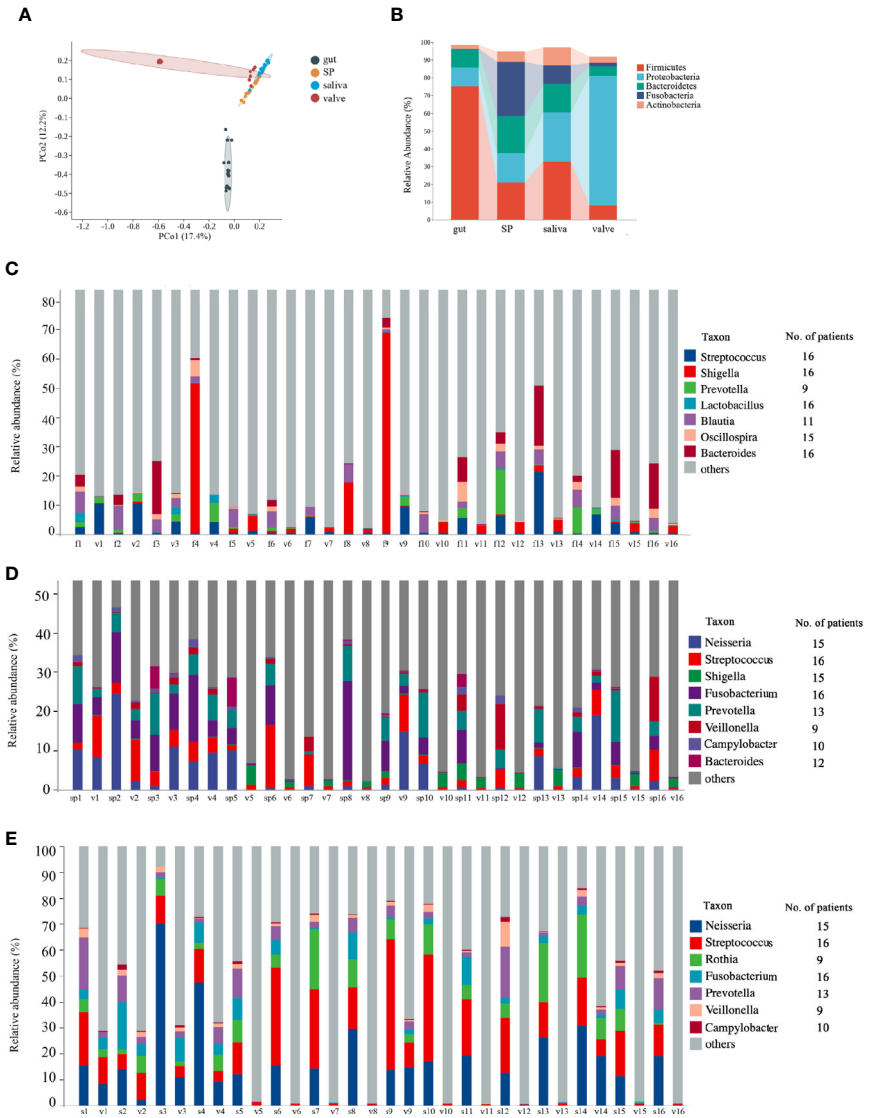


FIGURE 4 | Comparisons of microbiota composition between mitral valves and other body sites of patients with rheumatic heart disease. Microbiota of mitral valves from patients with rheumatic heart disease (RHD) were analyzed using 16S rRNA gene sequencing. **(A)** Comparison of microbial β-diversity (PCoA) among gut, subgingival plaques (SP), saliva, and mitral valves (valve) of RHD patients. **(B)**Taxonomic composition of microbiota at different body sites of RHD patients at phylum level. **(C)** Illustration of genera detected in both fecal sample (f) and mitral valve (v) of each RHD patient. The numbers (1–16) denote RHD patients. The column on right indicates the number of patients who possess the corresponding genus in both sites. **(D)** Illustration of genera detected in both subgingival plaque (sp) and mitral valve (v) of each RHD patient. **(E)** Illustration of genera detected in both saliva (s) and mitral valve (v) of each RHD patient. n = 16.

plaques, on the other hand, positively correlated with LAD (**Figure 6A**). *Blautia* in the gut negatively correlated with pulmonary artery systolic pressure (PASP); *Corynebacterium* and *Roseburia* in subgingival plaques, on the other hand, positively correlated with PASP (**Figure 6B**).

DISCUSSION

To our knowledge, this work is the first one to present comprehensive characterization of the microbiota in RHD patients. Our study demonstrated that RHD patients had altered gut and oral microbiota, which was likely to translocate to mitral valves and correlate with severity of the disease. These results provided new insights on etiology, diagnosis, prevention, and treatment of RHD.

We identified significant alterations in the microbial profile of gut in RHD patients. We observed that relative abundances of *Bifidobacterium* and *Eubacterium* increased and those of *Faecalibacterium* and *Bacteroides* decreased in RHD patients. Moreover, *Coprococcus* served as a super-generalist in gut microbiota of RHD patients. *Bifidobacterium* has been

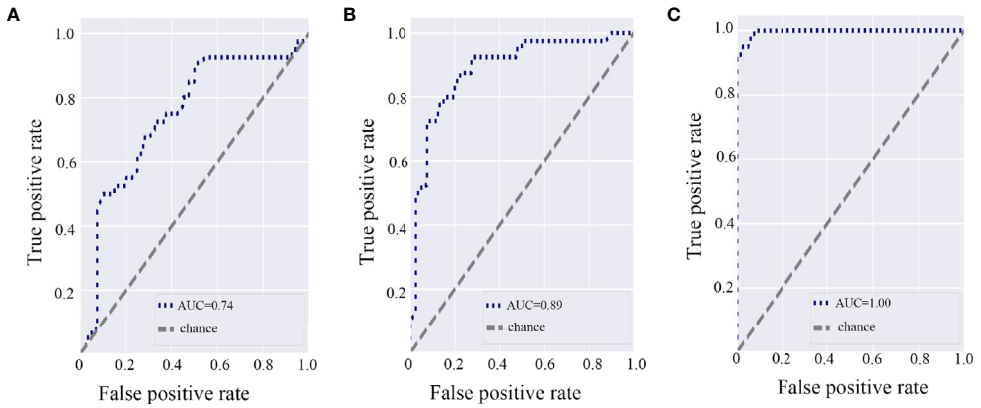


FIGURE 5 | Gut and oral microbiota differentiate rheumatic heart disease patients from control subjects. **(A)** Receiver operating characteristic (ROC) curve according to random forest model for fecal microbiota. A greater area under the ROC curve (AUC) indicates better performance. **(B)** ROC curve according to random forest model for subgingival plaque microbiota. **(C)** ROC curve according to random forest model for salivary microbiota.

considered a probiotic that can alter host microbiota in favor of a "healthier" composition (Toscano et al., 2017). *Eubacterium*, *Faecalibacterum*, *Bacteroides*, *Coprococcus*, and *Blautia* are important propionate- or butyrate-producing bacteria (Hoyles and Swann, 2019; Blaak et al., 2020). Propionate can enhance generation of macrophage and dendritic cell precursors in the bone marrow (Trompette et al., 2014) and butyrate can facilitate extrathymic generation of T-reg cells (Arpaia et al., 2013). Butyrate also helps to maintain the integrity of the intestinal epithelium and reduce injury in distant organs such as lungs (Haak et al., 2018; Hoyles and Swann, 2019). Therefore, it seemed counterintuitive that *Bifidobacterium* and *Eubacterium* increased in the gut microbiota of RHD patients. We postulated that during the progression of RHD the decrease of some beneficial genera such as *Faecalibacterum* and *Bacteroides* contributed to disruption of immune homeostasis, resulting in passive increase of some other beneficial genera such as *Bifidobacterium* and *Eubacterium* in response to lasting immune dysbiosis. Similarly, previous studies have demonstrated an increase of *Bifidobacterium* or *Eubacterium* in patients with inflammatory bowel disease (Wang et al., 2013) or systemic lupus erythematosus (He et al., 2016). We also observed that *Bacteroides*, *Eubacterium*, and *Blautia* in the gut negatively correlated with severity of RHD, suggesting beneficial roles of these three genera in RHD. However, neither *Bifidobacterium* nor *Faecalibacterum* significantly correlated with any clinical indices of RHD (Data not shown). The detailed functions and mechanisms of these individual microbes in RHD remain to be further delineated.

The gut microbiota of RHD patients manifested some similarities but more differences comparing to that of patients with other diseases. We observed increases in most of the abundant genera in the gut microbiota of RHD patients. The reason may be that RHD patients were subjected to high levels of stress and gut inflammation that reduced gastrointestinal motility and clearance capacity, leading to bacterial overgrowth in turn. This phenomenon has been previously observed in patients with type 1 diabetes, chronic heart failure, or other critical illness (Btaiche et al., 2010; Pasini et al., 2016; Malik et al., 2018). The gut microbial profile of RHD patients differed significantly from other cardiovascular diseases such as chronic heart failure and pulmonary arterial hypertension (Kummen et al., 2018; Kim et al., 2020). For example, gut microbiota of patients with chronic heart failure had relatively low abundance in *Bifidobacterium* and high in *Prevotella* (Kummen et al., 2018), whereas our results demonstrated that the gut microbiota of RHD patients had relatively high abundance in *Bifidobacterium* and no difference in *Prevotella*. Some genera in RHD patients had similar alterations as in patients with pulmonary arterial hypertension, including decreased *Bacteroides* and increased *Bifidobacterium* (Kim et al., 2020). However, alterations of some other genera were completely different between RHD patients and patients with pulmonary arterial hypertension. For example, the abundance of *Eubacterium* increased in RHD patients, while it decreased in patients with pulmonary arterial hypertension (Kim et al., 2020). Thus, we propose that the same bacteria may respond to different stimuli specifically and play different roles in various disease conditions. Further animal experiments should be implemented to confirm their specific functions. It is the unique microbial profile that makes it possible to discriminate RHD patients from healthy controls or even other diseases.

We also identified significant microbial alterations in oral cavity of RHD patients. A large variety of microbiota resides in different locations of the oral cavity (Avila et al., 2009). We analyzed the microbiota of saliva and subgingival plaques in this study. Salivary microbiota of RHD patients showed higher richness than that of control subjects, likely because of poor oral hygiene (Maharaj and Vayej, 2012; Belstrom et al., 2018). The difference in β-diversity of salivary microbiota between RHD and control subjects was more dramatic in comparison with that of gut and subgingival plaque microbiota. In line with this, the diagnostic value of salivary microbiota was the greatest according to random forest analysis. The results underpinned the theory that saliva represented a significant source of

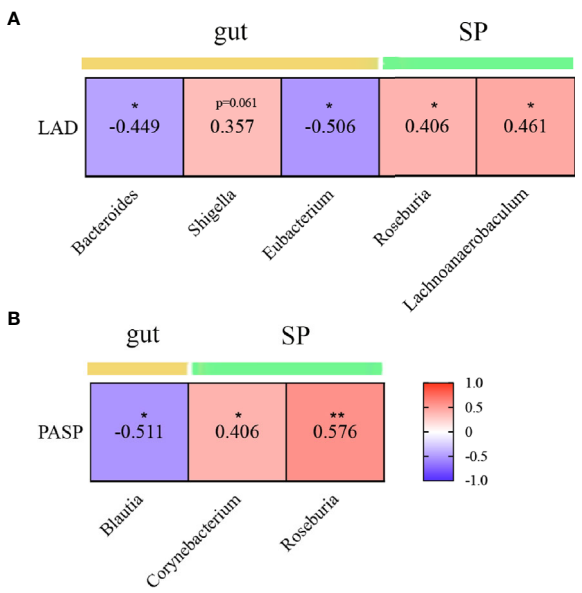


FIGURE 6 | Correlations between microbial genera and clinical indices in patients with rheumatic heart disease. **(A)** Heatmap showing correlations between different genera and LAD. **(B)** Heatmap showing correlations between different genera and PASP. LAD indicates left atrial diameter; and PASP, pulmonary artery systolic pressure. SP indicates subgingival plaque. Spearman's rank correlation coefficients are illustrated in squares. *P < 0.05; **P < 0.01.

discriminatory biomarkers for oral and systemic diseases (Yoshizawa et al., 2013). For example, salivary microbes were able to correlate with clinical indices and stratify active and moderately active patients of rheumatoid arthritis (Zhang et al., 2015). Similar results were obtained in Crohn's disease, another autoimmune disease (Zhang et al., 2020). Although microbes in saliva had large potential to differentiate RHD from control subjects, there was no correlation between any single salivary microbe and RHD severity. The reason may be that salivary microbes closely interacted with each other and functioned as a whole (Jenkinson and Lamont, 2005). Our data showed that genus such as *Roseburia*, *Lachnoanaerobaculum*, and *Corynebacterium* in subgingival plaques correlated with RHD severity, although the functions of these genera in periodontal disease or RHD remain to be further explored.

Our results also showed that *Streptococcus* significantly increased in saliva of RHD patients. There are 74 species under the genus of *Streptococcus* (Wong and Yuen, 2012). GAS is responsible for pharyngitis and the post-infection sequela, including RHD (Soderholm et al., 2018). GAS can survive the oral immune defense system and remain viable for long periods (Walker et al., 2014). Not all *Streptococcus* spp. are harmful. For instance, *Streptococcus salivarius* K12 is a probiotic intended for use in the oral cavity and can antagonize the growth of GAS (Di Pierro et al., 2016). Therefore, it needs to be further determined which species of *Streptococcus* contributed to the changes we observed in the oral microbiota of RHD patients. *Streptococcus* also increased in subgingival plaques of RHD patients, likely a consequence of the increase of this genus in saliva (Avila et al., 2009).

Our data suggested potential transmission of microbes from the gut or oral cavity to heart valves of RHD patients. Some genera were shared by mitral valves and gut or oral cavity in RHD patients, indicating that microbes in the gut and/or oral cavity might translocate to the mitral valves. In the context of immunological cascade caused by GAS infection, gut microbial dysbiosis may increase endotoxin production and weaken gut barrier function, leading to increase of intestinal permeability and subsequent bacterial translocation (Camilleri, 2019). Oral cavity is another potential origin of the microbiota in mitral valves. Transmission of oral microbes to other body sites can be caused by invasive dental procedure or even normal daily activities such as tooth brushing and food intake (Smith and Nehring, 2020). Our results demonstrated that the microbiota in mitral valves partially overlapped with that in oral cavity. Intriguingly, *Campylobacter* was distributed in most mitral valves and oral cavity but only presented in one fecal sample of RHD patients, pointing to the possibility of oral-to-valve but not gut-to-valve translocation of this microbe. *Streptococcus* was abundantly distributed in mitral valves of every RHD patient. This is consistent with the theory that GAS or its components can enter the circulation and gain access to the subendothelial collagen matrix (Tandon et al., 2013). *Neisseria, Lactobacillus, Campylobacter and Prevotella* were detected in the mitral valves of most RHD patients. It would be reasonable to speculate that these microbes, together with *Streptococcus*, could translocate from the gut and/or oral cavity to the mitral valves, creating antigens that provoke autoimmune response against host cardiac tissues over the progression of RHD.

Our data provided new insights on the etiology, diagnosis, prevention, and treatment of RHD. First, unique microbial profiles of RHD broadened the concept of genetic susceptibility to RHD. Only a minority (1–2%) of populations living in GAS-endemic areas develop RHD (Carapetis et al., 2000) and specific genetic markers have been linked to RHD (Gray et al., 2017). Microbiota, considered as the second genome of the human body, may play a role in genetic predisposition to RHD (Grice and Segre, 2012). Second, unique microbial profiles may serve as a supplemental diagnostic tool for RHD patients. Our results illustrated that it was feasible to differentiate RHD patients from control subjects using microbiota. Third, our data suggested that the microbiota played important roles in RHD treatment. Broad-spectrum antibiotics may eliminate beneficial microbes and sometimes cause secondary infections (Kelly and LaMont, 2008). Antibiotics designed with the narrowest spectra targeting GAS or regular use of probiotics may reduce the ecologically undesirable side effects of non-discriminative antibiotic chemoprophylaxis (Lemon et al., 2012). At last, the microbiota may serve as a viable therapeutic target for halting progression to severe mitral stenosis and for post-surgery management. The microbiota may influence immune homeostasis, trigger inflammation, invade tissues, and create antigens (Hooper et al., 2012; Tandon et al., 2013;

Anhe et al., 2020). It is feasible to treat diseases through manipulation of the microbiota. For example, fecal microbiota transplantation has been safely performed in the treatment of diarrhea and sepsis in critical care (Li et al., 2015). Moreover, microbiota affects drug pharmacokinetics. The rate of absorption and bioavailability of many oral drugs depends on their exposure to bacterial enzymes before entering the circulation (Li and Jia, 2013). For RHD patients with prosthetic valves, the gut microbiota has been implicated to affect anticoagulant therapy that is mandatory after surgery (Wang et al., 2020). Our study is a start point for potential microbiota-targeting therapies, although more studies are required to establish causative links between the microbiota and RHD.

In conclusion, we described important alterations in the microbiota of RHD patients, provided evidence that microbiota in gut and oral cavity might translocate to mitral valves, demonstrated the possibility of distinguishing patients from control subjects using microbiota, and identified gut/oral microbes that correlated with severity of RHD. Our study paved a promising path for using microbiota as a potential diagnostic, prophylactic, and therapeutic tool for RHD.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov; PRJNA682705.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Human Ethics Committee, Shanghai Chest Hospital, Shanghai Jiaotong University, and conducted in accordance with the Principles of Good Clinical Practice and the Declaration of Helsinki. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

R-GL and S-ZD designed and supervised the project. X-RS and B-YC performed the statistical analyses. W-ZL, Y-LL, Y-LW, J-JH, and W-WZ collected the clinical samples and extracted the DNA. YL, X-XM, and SS interpreted the data. X-RS wrote the manuscript. R-GL and S-ZD read and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (81991500, 81991503, 81921002, 82070262), Shanghai's Top Priority Clinical Medicine Center (2017ZZ01011), Science and Technology Commission of Shanghai Municipality (19140905302), Shanghai Shen Kang Hospital Development Center Clinical Research Plan of SHDC (2020CR3026B), and the Innovative Research Team of High-Level Local Universities in Shanghai (SSMU-ZDCX20180900).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2021.643092/full#supplementary-material

REFERENCES

Anhe, F. F., Jensen, B. A. H., Varin, T. V., Servant, F., Van Blerk, S., Richard, D., et al. (2020). Type 2 diabetes influences bacterial tissue compartmentalisation in human obesity. *Nat. Metab.* 2 (3), 233–242. doi: 10.1038/s42255-020-0178-9

Arpaia, N., Campbell, C., Fan, X. Y., Dikiy, S., van der Veeken, J., deRoos, P., et al. (2013). Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* 504 (7480), 451–45+. doi: 10.1038/nature12726

Avila, M., Ojcius, D. M., and Yilmaz, O. (2009). The Oral Microbiota: Living with a Permanent Guest. *DNA Cell Biol.* 28 (8), 405–411. doi: 10.1089/dna.2009.0874

Belkaid, Y., and Hand, T. W. (2014). Role of the Microbiota in Immunity and Inflammation. *Cell* 157 (1), 121–141. doi: 10.1016/j.cell.2014.03.011

Belstrom, D., Sembler-Moller, M. L., Grande, M. A., Kirkby, N., Cotton, S. L., Paster, B. J., et al. (2018). Impact of Oral Hygiene Discontinuation on Supragingival and Salivary Microbiomes. *JDR Clin. Trans. Res.* 3 (1), 57–64. doi: 10.1177/2380084417723625

Blaak, E. E., Canfora, E. E., Theis, S., Frost, G., Groen, A. K., Mithieux, G., et al. (2020). Short chain fatty acids in human gut and metabolic health. *Benef. Microbes* 11 (5), 411–455. doi: 10.3920/BM2020.0057

Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., et al. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2 ' s q2-feature-classifier plugin. *Microbiome* 6 (1), 90. doi: 10.1186/s40168-018-0470-z

Btaiche, I. F., Chan, L. N., Pleva, M., and Kraft, M. D. (2010). Critical Illness, Gastrointestinal Complications, and Medication Therapy during Enteral Feeding in Critically Ill Adult Patients. *Nutr. Clin. Pract.* 25 (1), 32–49. doi: 10.1177/0884533609357565

Bui, F. Q., Coutinho Almeida-da-Silva, C. L., Huynh, B., Trinh, A., Liu, J., Woodward, J., et al. (2019). Association between periodontal pathogens and systemic disease. *Biomed. J.* 42 (1), 27–35. doi: 10.1016/j.bj.2018.12.001

Camilleri, M. (2019). Leaky gut: mechanisms, measurement and clinical implications in humans. *Gut* 68 (8), 1516–1526. doi: 10.1136/gutjnl-2019-318427

Carapetis, J. R., Currie, B. J., and Mathews, J. D. (2000). Cumulative incidence of rheumatic fever in an endemic region: a guide to the susceptibility of the population? *Epidemiol. Infect.* 124 (2), 239–244. doi: 10.1017/s0950268800003514

Deng, Y., Jiang, Y. H., Yang, Y., He, Z., Luo, F., and Zhou, J. (2012). Molecular ecological network analyses. *BMC Bioinf.* 13, 113. doi: 10.1186/1471-2105-13-113

Di Pierro, F., Colombo, M., Zanvit, A., and Rottoli, A. S. (2016). Positive clinical outcomes derived from using Streptococcus salivarius K12 to prevent streptococcal pharyngotonsillitis in children: a pilot investigation. *Drug Healthcare Patient Saf.* 8, 77–81. doi: 10.2147/dhps.S117214

Gray, L.-A., D'Antoine, H. A., Tong, S. Y. C., McKinnon, M., Bessarab, D., Brown, N., et al. (2017). Genome-Wide Analysis of Genetic Risk Factors for Rheumatic Heart Disease in Aboriginal Australians Provides Support for Pathogenic Molecular Mimicry. *J. Infect. Dis.* 216 (11), 1460–1470. doi: 10.1093/infdis/jix497

Grice, E. A., and Segre, J. A. (2012). The human microbiome: our second genome. *Annu Rev Genomics Hum Genet* (2012) 13, 151–170. doi: 10.1146/annurev-genom-090711-163814

Haak, B. W., Littmann, E. R., Chaubard, J.-L., Pickard, A. J., Fontana, E., Adhi, F., et al. (2018). Impact of gut colonization with butyrate-producing microbiota on respiratory viral infection following allo-HCT. *Blood* 131 (26), 2978–2986. doi: 10.1182/blood-2018-01-828996

Hajishengallis, G., Darveau, R. P., and Curtis, M. A. (2012). The keystone-pathogen hypothesis. *Nat. Rev. Microbiol.* 10 (10), 717–725. doi: 10.1038/nrmicro2873

He, Z., Shao, T., Li, H., Xie, Z., and Wen, C. (2016). Alterations of the gut microbiome in Chinese patients with systemic lupus erythematosus. *Gut Pathog.* 8 (1), 64. doi: 10.1186/s13099-016-0146-9

Hooper, L. V., Littman, D. R., and Macpherson, A. J. (2012). Interactions Between the Microbiota and the Immune System. *Science* 336 (6086), 1268–1273. doi: 10.1126/science.1223490

Hoyles, L., and Swann, J. (2019). "Influence of the Human Gut Microbiome on the Metabolic Phenotype," in *The Handbook of Metabolic Phenotyping* (Elsevier Press), 535–560.

Jenkinson, H. F., and Lamont, R. J. (2005). Oral microbial communities in sickness and in health. *Trends Microbiol.* 13 (12), 589–595. doi: 10.1016/j.tim.2005.09.006

Kelly, C. P., and LaMont, J. T. (2008). Clostridium difficile - More difficult than ever. *N. Engl. J. Med.* 359 (18), 1932–1940. doi: 10.1056/NEJMra0707500

Kim, S., Rigatto, K., Gazzana, M. B., Knorst, M. M., Richards, E. M., Pepine, C. J., et al. (2020). Altered Gut Microbiome Profile in Patients With Pulmonary Arterial Hypertension. *Hypertension* 75 (4), 1063–1071. doi: 10.1161/HYPERTENSIONAHA.119.14294

Koren, O., Spor, A., Felin, J., Fak, F., Stombaugh, J., Tremaroli, V., et al. (2010). Human oral, gut, and plaque microbiota in patients with atherosclerosis. *Proc. Natl. Acad. Sci.* 108 (Supplement_1), 4592–4598. doi: 10.1073/pnas.1011383107

Kummen, M., Mayerhofer, C. C. K., Vestad, B., Broch, K., Awoyemi, A., Storm-Larsen, C., et al. (2018). Gut Microbiota Signature in Heart Failure Defined From Profiling of 2 Independent Cohorts. *J. Am. Coll. Cardiol.* 71 (10), 1184–1186. doi: 10.1016/j.jacc.2017.12.057

Lemon, K. P., Armitage, G. C., Relman, D. A., and Fischbach, M. A. (2012). Microbiota-Targeted Therapies: An Ecological Perspective. *Sci. Trans. Med.* 4 (137), 137rv135. doi: 10.1126/scitranslmed.3004183

Li, H., and Jia, W. (2013). Cometabolism of Microbes and Host: Implications for Drug Metabolism and Drug-Induced Toxicity. *Clin. Pharmacol. Ther.* 94 (5), 574–581. doi: 10.1038/clpt.2013.157

Li, Q., Wang, C., Tang, C., He, Q., Zhao, X., Li, N., et al. (2015). Successful treatment of severe sepsis and diarrhea after vagotomy utilizing fecal microbiota transplantation: a case report. *Crit. Care* 19 (1), 37 doi: 10.1186/s13054-015-0738-7

Li, H., Xu, H., Li, Y., Jiang, Y., Hu, Y., Liu, T., et al. (2020). Alterations of gut microbiota contribute to the progression of unruptured intracranial aneurysms. *Nat. Commun.* 11 (1), 328. doi: 10.1038/s41467-020-16990-3

Maharaj, B., and Vayej, A. C. (2012). Oral health of patients with severe rheumatic heart disease. *Cardiovasc. J. Afr.* 23 (6), 336–339. doi: 10.5830/cvja-2012-009

Malik, A., Morya, R. K., Bhadada, S. K., and Rana, S. (2018). Type 1 diabetes mellitus: Complex interplay of oxidative stress, cytokines, gastrointestinal motility and small intestinal bacterial overgrowth. *Eur. J. Clin. Invest.* 48 (11), e13021. doi: 10.1111/eci.13021

Nishimura, R. A., Otto, C. M., Bonow, R. O., Carabello, B. A., Erwin, J. P., Guyton, R. A., et al. (2014). 2014 AHA/ACC Guideline for the Management of Patients With Valvular Heart Disease: Executive Summary. *J. Am. Coll. Cardiol.* 63 (22), 2438–2488. doi: 10.1016/j.jacc.2014.02.537

Nishimura, R. A., Otto, C. M., Bonow, R. O., Carabello, B. A., Erwin, J. P., Fleisher, L. A., et al. (2017). 2017 AHA/ACC Focused Update of the 2014 AHA/ACC Guideline for the Management of Patients With Valvular Heart Disease. *J. Am. Coll. Cardiol.* 70 (2), 252–289. doi: 10.1016/j.jacc.2017.03.011

Pasini, E., Aquilani, R., Testa, C., Baiardi, P., Angioletti, S., Boschi, F., et al. (2016). Pathogenic Gut Flora in Patients With Chronic Heart Failure. *Jacc-Heart Failure* 4 (3), 220–227. doi: 10.1016/j.jchf.2015.10.009

Pihlstrom, B. L., Michalowicz, B. S., and Johnson, N. W. (2005). Periodontal diseases. *Lancet* 366 (9499), 1809–1820. doi: 10.1016/s0140-6736(05)67728-8

Reményi, B., Wilson, N., Steer, A., Ferreira, B., Kado, J., Kumar, K., et al. (2012). World Heart Federation criteria for echocardiographic diagnosis of rheumatic heart disease—an evidence-based guideline. *Nat. Rev. Cardiol.* 9 (5), 297–309. doi: 10.1038/nrcardio.2012.7

Russell, E. A., Walsh, W. F., Reid, C. M., Tran, L., Brown, A., Bennetts, J. S., et al. (2017). Outcomes after mitral valve surgery for rheumatic heart disease. *Heart Asia* 9 (2), e010916. doi: 10.1136/heartasia-2017-010916

Smith, D. A., and Nehring, S. M. (2020). "Bacteremia," in *StatPearls* (Treasure Island, FL: StatPearls Publishing LLC).

Soderholm, A. T., Barnett, T. C., Sweet, M. J., and Walker, M. J. (2018). Group A streptococcal pharyngitis: Immune responses involved in bacterial clearance and GAS-associated immunopathologies. *J. Leukocyte Biol.* 103 (2), 193–213. doi: 10.1189/jlb.4MR0617-227RR

Tandon, R., Sharma, M., Chandrashekhar, Y., Kotb, M., Yacoub, M. H., and Narula, J. (2013). Revisiting the pathogenesis of rheumatic fever and carditis. *Nat. Rev. Cardiol.* 10 (3), 171–177. doi: 10.1038/nrcardio.2012.197

Tang, W. H. W., Backhed, F., Landmesser, U., and Hazen, S. L. (2019). Intestinal Microbiota in Cardiovascular Health and Disease. *J. Am. Coll. Cardiol.* 73 (16), 2089–2105. doi: 10.1016/j.jacc.2019.03.024

Tompkins, D. G., Boxerbaum, B., and Liebman, J. (1972). Long-term prognosis of rheumatic fever patients receiving regular intramuscular benzathine penicillin. *Circulation* 45 (3), 543–551. doi: 10.1161/01.Cir.45.3.543

Toscano, M., De Grandi, R., Stronati, L., De Vecchi, E., and Drago, L. (2017). Effect of Lactobacillus rhamnosus HN001 and Bifidobacterium longum BB536 on the healthy gut microbiota composition at phyla and species level: A preliminary study. *World J. Gastroenterol.* 23 (15), 2696–2704. doi: 10.3748/wjg.v23.i15.2696

Trompette, A., Gollwitzer, E. S., Yadava, K., Sichelstiel, A. K., Sprenger, N., Ngom-Bru, C., et al. (2014). Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. *Nat. Med.* 20 (2), 159–166. doi: 10.1038/nm.3444

Walker, M. J., Barnett, T. C., McArthur, J. D., Cole, J. N., Gillen, C. M., Henningham, A., et al. (2014). Disease Manifestations and Pathogenic Mechanisms of Group A Streptococcus. *Clin. Microbiol. Rev.* 27 (2), 264–301. doi: 10.1128/cmr.00101-13

Wang, W., Chen, L., Zhou, R., Wang, X., Song, L., Huang, S., et al. (2013). Increased Proportions of Bifidobacterium and the Lactobacillus Group and Loss of Butyrate-Producing Bacteria in Inflammatory Bowel Disease. *J. Clin. Microbiol.* 52 (2), 398–406. doi: 10.1128/jcm.01500-13

Wang, L., Liu, L., Liu, X., Xiang, M., Zhou, L., Huang, C., et al. (2020). The gut microbes, Enterococcus and Escherichia-Shigella, affect the responses of heart valve replacement patients to the anticoagulant warfarin. *Pharmacol. Res.* 159, 104979. doi: 10.1016/j.phrs.2020.104979

Watkins, D. A., Johnson, C. O., Colquhoun, S. M., Karthikeyan, G., Beaton, A., Bukhman, G., et al. (2017). Global, Regional, and National Burden of Rheumatic Heart Diseas-2015. *N. Engl. J. Med.* 377 (8), 713–722. doi: 10.1056/NEJMoa1603693

Watkins, D. A., Beaton, A. Z., Carapetis, J. R., Karthikeyan, G., Mayosi, B. M., Wyber, R., et al. (2018). Rheumatic Heart Disease Worldwide JACC Scientific Expert Panel. *J. Am. Coll. Cardiol.* 72 (12), 1397–1416. doi: 10.1016/j.jacc.2018.06.063

Woldu, B., and Bloomfield, G. S. (2016). Rheumatic Heart Disease in the Twenty-First Century. *Curr. Cardiol. Rep.* 18 (10), 11. doi: 10.1007/s11886-016-0773-2

Wong, S. S. Y., and Yuen, K. Y. (2012). Streptococcus pyogenes and re-emergence of scarlet fever as a public health problem. *Emerg. Microbes Infect.* 1:10. doi: 10.1038/emi.2012.9

Yoshizawa, J. M., Schafer, C. A., Schafer, J. J., Farrell, J. J., Paster, B. J., and Wong, D. T. (2013). Salivary biomarkers: toward future clinical and diagnostic utilities. *Clin. Microbiol. Rev.* 26 (4), 781–791. doi: 10.1128/CMR.00021-13

Zhang, X., Zhang, D., Jia, H., Feng, Q., Wang, D., Liang, D., et al. (2015). The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nat. Med.* 21 (8), 895–905. doi: 10.1038/nm.3914

Zhang, Y. H., Wang, X., Li, H. X., Ni, C., Du, Z. B., and Yan, F. H. (2018). Human oral microbiota and its modulation for oral health. *Biomed. Pharmacother.* 99, 883–893. doi: 10.1016/j.biopha.2018.01.146

Zhang, T., Kayani, M. U. R., Hong, L., Zhang, C., Zhong, J., Wang, Z., et al. (2020). Dynamics of the Salivary Microbiome During Different Phases of Crohn's Disease. *Front. Cell Infect. Microbiol.* 10, 544704. doi: 10.3389/fcimb.2020.544704

Ziebolz, D., Jahn, C., Pegel, J., Semper-Pinnecke, E., Mausberg, R. F., Waldmann-Beushausen, R., et al. (2018). Periodontal bacteria DNA findings in human cardiac tissue - Is there a link of periodontitis to heart valve disease? *Int. J. Cardiol.* 251, 74–79. doi: 10.1016/j.ijcard.2017.09.001

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cellular and Infection Microbiology

ORIGINAL RESEARCH
published: 19 March 2021
doi: 10.3389/fcimb.2021.625229

Dysbiosis of the Human Oral Microbiome During the Menstrual Cycle and Vulnerability to the External Exposures of Smoking and Dietary Sugar

Nagihan Bostanci[1†], Maria Christine Krog[2,3†], Luisa W. Hugerth[4,5†], Zahra Bashir[2,6], Emma Fransson[4,5], Fredrik Boulund[4,5], Georgios N. Belibasakis[1], Kristin Wannerberger[7], Lars Engstrand[4,5], Henriette Svarre Nielsen[2,8,9‡] and Ina Schuppe-Koistinen[4,5‡]*

[1] Division of Oral Diseases, Department of Dental Medicine, Karolinska Institutet, Stockholm, Sweden, [2] The Recurrent Pregnancy Loss Units, Copenhagen University Hospitals, Rigshospitalet and Hvidovre Hospital, Copenhagen, Denmark, [3] Department of Clinical Immunology, Copenhagen University Hospital, Rigshospitalet, Denmark, [4] Centre for Translational Microbiome Research, Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden, [5] Science for Life Laboratory, Stockholm, Sweden, [6] Department of Obstetrics and Gynaecology, Holbæk Hospital, Holbæk, Denmark, [7] Ferring International Center SA, Saint-Prex, Switzerland, [8] Department of Obstetrics and Gynaecology, Hvidovre Hospital, Copenhagen, Denmark, [9] Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

Physiological hormonal fluctuations exert endogenous pressures on the structure and function of the human microbiome. As such, the menstrual cycle may selectively disrupt the homeostasis of the resident oral microbiome, thus compromising oral health. Hence, the aim of the present study was to structurally and functionally profile the salivary microbiome of 103 women in reproductive age with regular menstrual cycle, while evaluating the modifying influences of hormonal contraceptives, sex hormones, diet, and smoking. Whole saliva was sampled during the menstrual, follicular, and luteal phases (n = 309) of the cycle, and the participants reported questionnaire-based data concerning their life habits and oral or systemic health. No significant differences in alpha-diversity or phase-specific clustering of the overall microbiome were observed. Nevertheless, the salivary abundances of genera *Campylobacter*, *Haemophilus*, *Prevotella*, and *Oribacterium* varied throughout the cycle, and a higher species-richness was observed during the luteal phase. While the overall community structure maintained relatively intact, its functional properties were drastically affected. In particular, 11 functional modules were differentially abundant throughout the menstrual cycle, including pentose phosphate metabolism, and biosynthesis of cobalamin and neurotransmitter gamma-aminobutyric acid. The menstrual cycle phase, but not oral contraceptive usage, was accountable for greater variations in the metabolic pathways of the salivary microbiome. Further co-risk factor analysis demonstrated that *Prevotella* and *Veillonella* were increased in current smokers, whereas high dietary sugar consumption modified the richness and diversity of the microbiome during the cycle. This is the first large study to systematically address dysbiotic variations of the oral microbiome during the course of menstrual cycle, and

OPEN ACCESS

Edited by:
Thuy Do,
University of Leeds, United Kingdom

Reviewed by:
Janina P. Lewis,
Virginia Commonwealth University, United States
Gena D. Tribble,
University of Texas Health Science Center at Houston, United States
André L. Teixeira,
University of Guelph, Canada

***Correspondence:**
Henriette Svarre Nielsen
henriette.svarre.nielsen@regionh.dk

[†]These authors have contributed equally to this work

[‡]These authors have contributed equally to this work

Specialty section:
This article was submitted to Microbiome in Health and Disease, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: *02 November 2020*
Accepted: *11 February 2021*
Published: *19 March 2021*

Citation:
Bostanci N, Krog MC, Hugerth LW, Bashir Z, Fransson E, Boulund F, Belibasakis GN, Wannerberger K, Engstrand L, Nielsen HS and Schuppe-Koistinen I (2021) Dysbiosis of the Human Oral Microbiome During the Menstrual Cycle and Vulnerability to the External Exposures of Smoking and Dietary Sugar. Front. Cell. Infect. Microbiol. 11:625229. doi: 10.3389/fcimb.2021.625229

document the additive effect of smoking and sugar consumption as environmental risk factors. It reveals the structural resilience and functional adaptability of the oral microbiome to the endogenous hormonal pressures of the menstrual cycle, while revealing its vulnerability to the exogenous exposures of diet and smoking.

Keywords: menstrual cycle, oral microbiome, saliva, hormonal contraceptives, sugar, diet, women's health, shotgun sequencing

INTRODUCTION

The oral cavity is a special ecological habitat composed of soft and non-shedding hard tissues, colonized with a plethora of microorganisms (Wade, 2013; Belibasakis et al., 2019). The oral microbiota constitutes the second most diverse microbial community of the human body, which under normal circumstances remains mostly stable (Human Microbiome Project, 2012; David et al., 2014; Rosier et al., 2018; Lif Holgerson et al., 2020). The continuous presence of saliva, a unique biological medium, has a crucial role in maintaining the stability of the oral microbiota. Higher microbial diversity in saliva, with altered pH and proteolytic enzyme activity was proposed to indicate early dysbiosis towards inflammatory oral diseases (Zaura et al., 2009). Salivary composition and flow rate vary with age, sex, daily rhythm, dietary habits as well as female hormones (Yeh et al., 1998; Lukacs and Largaespada, 2006; Gumus et al., 2015). Female hormones, specifically estrogens, may suppress the physiological salivary flow rate (Streckfus et al., 1998; Lu et al., 1999). This results in reduction of the natural antimicrobial capacity of saliva, thus disrupting the local microbial homeostasis and increasing susceptibility to gingivitis and dental caries in the affected women (Lukacs and Largaespada, 2006; Gursoy et al., 2008; Silva de Araujo Figueiredo et al., 2017).

The reproductive age of a woman's lifetime, spanning from puberty until menopause is characterized by major changes in circulating female hormone levels and frequently accompanied by emotional and physiological changes, including heightened inflammatory status (Oertelt-Prigione, 2012; Clancy et al., 2013). The salivary and plasma kinetics of sex steroids mirror one-another across the menstrual cycle (Gandara et al., 2007). Of note, salivary estradiol levels peak at the time of ovulation, aligning with compositional changes of uterine endocervical gland secretions (Saibaba et al., 2017). Among the most common oral signs observed during menstruation are gingival inflammation, reduced salivary flow, altered pH and oral ulcers (Holm-Pedersen and Loe, 1967). Furthermore, the use of hormonal contraceptives has been associated with poorer oral health in young women, yet the mechanistic links between such hormone supplementation and oral microbial dysbiosis is not elucidated (Brusca et al., 2010). Long-term use of oral contraceptives can lead to accelerated progression of periodontal disease, yet this may also be concentration-dependent (Preshaw et al., 2001; Preshaw et al., 2013). Early studies performed shortly after the introduction of oral contraceptives, when hormone doses were very high, found evidence of increased gingival inflammation and possibly also increased probing depths, often despite better oral hygiene in the users of the oral contraceptives (Preshaw et al., 2001). Low doses of estrogen and progesterone are now being widely used in contraceptive pills, and these have little impact on the extent of periodontal inflammation in response to plaque accumulation (Preshaw et al., 2001). Most studies in the field are based on clinical parameters of disease severity and do not take into account direct measures of oral infection, such as qualitative and quantitative microbiome changes in the oral milieu (Holm-Pedersen and Loe, 1967). The few studies available on the dynamic interplay between menstrual cycle hormones or hormonal contraceptives and selected components of the oral microbiota did not reach a clear consensus (Jensen et al., 1981; Fischer et al., 2008; Kumar, 2013). Interestingly, ovulation also has been linked to the increased levels of anaerobic bacterial counts in saliva independent of flow rate (Prout and Hopps, 1970). Jensen et al. reported that women who are taking oral contraceptives had up to sixteen times higher level of *Bacteroides* species in their dental plaque than the control group (Jensen et al., 1981).

Data regarding the dynamics of the salivary microbiome during the regular menstrual cycle using high-throughput sequencing technologies in young women is entirely missing. Therefore, the aim of the present study was to examine the structural and functional dynamics of the salivary microbiome during one full menstrual cycle in women of reproductive age under different contraceptive regimens.

MATERIALS AND METHODS

Study Design and Ethics Statement

Women were recruited by advertisements in student magazines, university notice boards, and social media and included between September 2017 and January 2018 at Rigshospitalet, Copenhagen, Denmark. All data were collected and managed using REDCap electronic data capture tools, hosted at the Capital Region of Denmark. The study is approved by The Regional Committee on Health Research Ethics (H-17017580) and the Data Protection Agency in the Capital Region of Denmark (2012-58-0004). All participants gave oral and written consent to participate. All participants were asked to phone the clinic when spotting/beginning of bleeding. They were then scheduled for a hospital visit on cycle day 1–3. An ultrasound scan was performed to confirm that the participant was not pregnant and to confirm the menstruation with shedding of the endometrium.

Additionally, hormone levels (plasma estradiol and progesterone) were measured. The two following visits were booked (day 8–12) and (day 18–22) every time with an ultrasound scan to confirm cycle phase and absence of pregnancy. One of the inclusion criteria was to have a regular cycle (median 26 days). The subjects were divided into 3 groups: group 1 (n = 43) that did not use any hormonal contraceptives, group 2 (n = 41) that used combined oral contraception (COC: estrogen + progestin) for at least half a year prior to the start of the study and group 3 (n = 19) using levonegestrel intra-uterine system (LNG-IUS). Women were excluded from the study if they were pregnant or had an intention to become pregnant during the course of the study or had oligomenorrhea or irregular menstrual cycles or spotting. Subjects were also excluded if they had any systemic disease or medical condition or been treated by antibiotics in the past three months prior to the beginning of the study or during its duration.

Variables Reported by the Subjects Through Questionnaires

Participants reported data concerning their life habits and health history. The food frequency questionnaire was based on four-week recall, with frequencies given on a 9-point scale from "0 times in the past four weeks" to ">3 times/day for the past four weeks". Frequency of free-sugar consumption was derived from the sum of the frequencies for the following food items: chocolate milk, juice, soda with sugar, ice-cream, biscuits and cookies, sweet bread and rolls, dry cake, cake with filling and candy (including chocolate, licorice, jelly and other candy). On quantitative analyses, sugar consumption was divided into low (up to and including the first quartile), high (above the third quartile), or intermediate. In relation to oral health, participants were asked whether they had been to a dentist or dental hygienist during the 3 months prior to answering the questionnaire. Smoking was coded as "daily smoker", "occasional smoker", "former smoker", and "never smoker".

Blood and Saliva Collection and Processing

Women were followed during a full menstrual cycle including three hospital visits. The first hospital visit was at cycle day (CD) 1–3. The second visit was CD 8–12 and the third CD 18–22. The patients were fasting 30 min before saliva collection, including drinking, chewing gum or chewing tobacco, and smoking. Whole saliva samples (2 ml) were collected using a SalivaGene Collector (STRATEC Molecular GmbH, Germany) containing lyophilized DNA stabilization buffer, according to the instructions of the manufacturer, and were frozen at -80°C. Blood samples were drawn at every hospital visit. Blood was collected in 9 ml EDTA tubes, left until separated and spinned for 15 min at 3,000 rpm and plasma was aliquoted and frozen at -80°C. Plasma estradiol and progesterone were measured using the standard automated system (Cobas® 8000 by Roche Diagnostics).

Extraction of Salivary DNA and Next-Generation Sequencing

Saliva aliquots of 600 µl were shipped to CoreBiome (OraSure, Bethlehem, PA, USA) where they were extracted with MO Bio PowerFecal (Qiagen, Hilden, Germany) automated for high throughput on QiaCube (Qiagen), with bead-beating in 0.1 mm glass bead plates. Three spaced negative controls and one positive control were included in each extraction. All negative extraction controls had undetectable amounts of DNA, and all positive controls were also approved. DNA concentration (for samples and controls) was quantified using Quant-iT Picogreen dsDNA Assay (Invitrogen, ThermoFisher Scientific, Carlsbad, CA, USA). Libraries were prepared using an adapted Nextera (Illumina Inc, San Diego, CA, USA) procedure and sequenced on an Illumina NextSeq using single-end 150 bp reads with a NextSeq 500/550 High Output v2 kit. Reads were processed with CoreBiome's BoosterShot shallow shotgun sequencing technology. The raw sequencing reads are available from the European Nucleotide Archive under project PRJEB37731, samples SAMEA6662389-SAMEA6662857.

Bioinformatics Analysis

Human reads were removed by mapping to the hg19 release of the human genome using BBTools (available at https://sourceforge.net/projects/bbmap/). Because BoosterShot technology is optimized for fecal samples, taxonomy was reannotated the reads using Kraken2 (26) with confidence set to 0.5 and Bracken, based on the Human Oral Microbiome Database v9.0.3 (27). Functional annotation based on KEGG modules was used as provided by CoreBiome.

Statistical Analyses

All statistical analyses were performed in R v. 3.5.2. Alpha-diversity was calculated as the observed number of species as well as Simpson's inverted index. Comparisons for the same individual across time were calculated as paired t-tests, while comparisons between individuals were calculated using Welch's t-test. Differences between the three contraception groups were calculated with Pearson's chi-square test or Fishers Exact test for count data and Kruskal-Wallis test for continuous data. Differences in variation were quantified using Levene's test of equality of variance, using the Brown-Forsythe variant with R package lawsat (v3.2). Correlations were calculated with Pearson's product moment. All tests were performed with a 95% confidence interval and a significance cutoff of 0.05. Multiple testing correction was conducted with the Benjamini-Hochberg procedure where applicable. Beta-diversity was calculated on Bray-Curtis distances and clustered on complete linkage. The relative impact of metadata factors on beta-diversity dispersion assessed through Permanova. Alpha- and Beta-diversity analyses were calculated with package Vegan (v2.5-3) and graphs were generated with packages RColorBrewer (v1.1-2) and Vioplot (v0.2). Associations between specific taxa and the metadata were calculated in Maaslin2, treating the contraceptive and phase of the cycle as fixed effects and individual's identities, smoking, and sugar intake as random effects. A minimum abundance of 0.1% in at least six samples was required to keep a taxon in the analysis. Furthermore, clustering of species was done according to the "color complex" classification of Socransky et al., (1998). (i.e., six-color cluster-lists of bacteria according to their frequency of detection and levels in periodontitis and health) and according to the core microbiome

classification data from Abusleme et al. (health-associated core species and periodontitis-associated core species in the subgingival microbiome) (Abusleme et al., 2013). Correlations between color groups and metadata were calculated initially as multivariate ANOVA, and where differences were found, investigated as beta-regressions treating contraceptive, phase of the cycle, smoking and sugar intake as fixed effects and individual's identities as random effects, using R package glmmTMB (v0.2.2.0).

RESULTS

Demographic and Clinical Characteristics of the Study Participants

Demographic and clinical characteristics of the 103 study participants are categorized based on contraceptive use and summarized in **Table 1**. The participants reported normal menstrual cycles of approximately 26 days (range, 23 to 34 days). There were no significant differences ($p > 0.05$) in the age, BMI or frequency of tobacco or cannabis usage between the women in each contraceptive group (**Table 1**). For participants using COC, 25 of 41 reported a daily dosage of 150 μg levonorgestrel and 30 μg ethinylestradiol. The other participants had several different progestins, combined with a daily dosage of ethinylestradiol of 20–35 μg. A single participant was on a multiphasic pill, while three could not give the name of their pill. For participants with an IUS, three brand names were reported (Jaydess, Kyleena and Mirena, all produced by Bayer AB, Leverkusen, Germany), with four participants unable to name the brand of their device.

A few of the participants had been to a dental health professional during the three months prior to the beginning of the study, with no difference between groups (**Supplementary Table 1**). Self-rated health was overall high and not significantly different between groups (**Supplementary Table 1**). Sugar consumption did not vary throughout the menstrual cycle or between the contraceptive groups (**Supplementary Table 1**).

Microbial Community Structure and Composition

The total number of annotated sequence reads for the overall cohort of 103 women was 102,605,212 (median 242,834 reads per sample, IQR 106,839–482,885). There were 209 microbial taxa identified in the saliva samples. The taxonomic classification is presented in **Supplementary Table 2**. In brief, the OTUs were collectively represented by eight bacterial phyla, namely Actinobacteria (0–70%, median 4%), Bacteroidetes (0–66%, median 37%), Firmicutes (0–77%, median 31%), Proteobacteria (0–65%, median 20%), Fusobacteria (0–8.6%), Spirochaetes (0–1.4%), and candidate divisions *SR1* (0–0.4%) and TM7 (0–14%). Fifty genera were identified, of which the most abundant genera across all samples were *Haemophilus* (0–48%, median 7%, specially *H. parainfluenzae*, 0–23%), *Neisseria* (0–53%, median 8%, specially *N. flavescens*, 0–21% and *N. subflava*, 0–23%), *Prevotella* (0–65%, median 34%, specially *P. histicola*, 0–24%, *P. melaninogenica*, 0–26%, *P. pallens*, 0–20% and oral taxon 313, 0–25%), *Streptococcus* (0–60%, median 13%, specially *S. mitis*, 0–60% and *S. parasanguinis*, 0–27%) and *Veillonella* (0–42%, median 11%, specially *V. atypica*, 0–29%) (**Figure 1**). In addition to these genera, a few samples presented with high levels of *Rothia mucilaginosa* (0–70%, median 1%) and *R. dentocariosa* (0–26%, median 0.2%).

To assess the relative effects of several metadata parameters on the structure of the microbiome, we ran a permutational analysis of variance (Permanova) including contraceptive usage, phase of the menstrual cycle, smoking status, free-sugar consumption, and individual identity. All factors are partially explanatory of the distance between samples (**Table 2**). While unspecific individual factors are dominating, contraception and cycle phase play a comparable role to well-known determinants of oral health, such as free-sugar consumption and smoking (**Table 2**). Therefore, the remaining analyses were adjusted for these factors when possible.

To assess whether any specific taxa differed in abundance according to female hormonal cycles and contraception, we ran Maaslin2 using contraceptive and cycle phase as fixed effects and individual identity, smoking and sugar intake as random effects (see the methods section for details). We found that *Prevotella* and *Veillonella* were increased in current smokers (daily and occasional). Additionally, four genera were found to differ in abundance throughout the menstrual cycle, namely *Campylobacter*, *Haemophilus*, *Prevotella*, and *Oribacterium*, the latter of which was over-represented by the species *O. sinus* (**Figure 2**). In addition, the genus *Atopobium* was found to be more abundant in IUS users ($r = 0.0017$, $p = 0.047$).

Oral Microbiome Richness and Diversity

No significant differences in alpha-diversity (within-sample diversity) were observed as a result of smoking, although daily

TABLE 1 | Demographic and clinical characteristics of the study participants.

	Non-hormonal contraception n = 43	Combined oral contraception n = 41	Levonorgestrel intra-uterine system n = 19	p-value
Age, years (median, IQR)	23.0 (22.0–28.0)	23.0 (22.0–24.0)	24.0 (22.0–25.0)	0.452[1]
BMI, kg/m^2 (median, IQR)	21.8 (20.8–24.6)	22.5 (20.7–24.3)	21.6 (21.6–23.1)	0.725[1]
Tobacco smoking (n)	16	8	7	0.163[2]
Tobacco, Snus* (n)	4	5	0	0.316[3]
Cannabis smoking (n)	5	0	3	0.025[3]

IQR, interquartile range; BMI, Body Mass Index.
[1]Kruskall-Wallis test [2]Chi-square test [3]Fisher's Exact test.
**Snus is a form of moist powder smokeless tobacco product.*

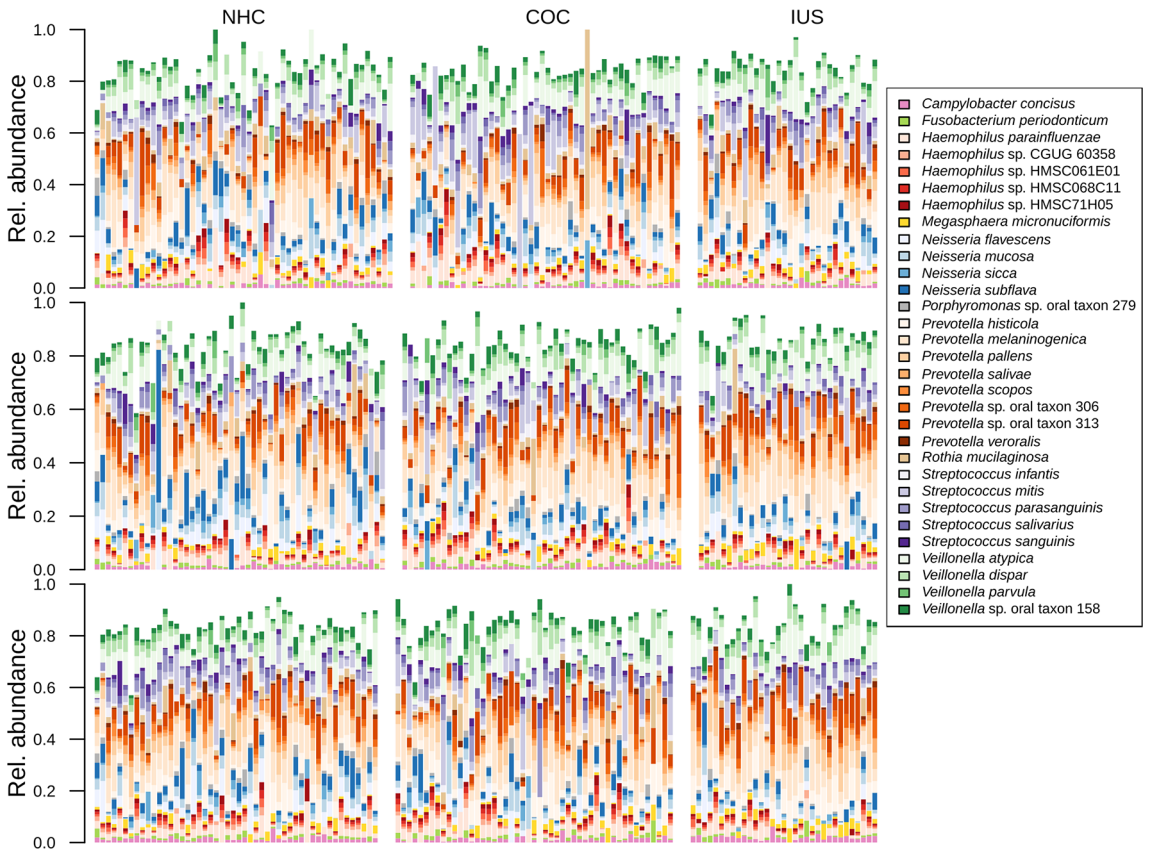


FIGURE 1 | Bar plot displaying the taxonomic composition of samples, sorted by phase of the menstrual cycle and contraceptive method.

TABLE 2 | Result of permutational analysis of variance shows the relative effect of important factors on differentiating the microbiome between samples.

Factor	Degrees of freedom	R^2	p-value
Smoking	3	0.083	0.001
Sugar consumption	2	0.024	0.001
Contraception	2	0.019	0.001
Cycle phase	2	0.012	0.001
Residuals	290	0.862	NA

smokers had slightly higher richness (number of species) than never smokers (means 36 vs. 34, p = 0.06). Participants in the middle and high ranges of sugar consumption had higher richness and diversity than those with low sugar consumption, but this different was only significant for the middle range (mean richness: low 32.8, mid 39.5, high 38.0; mean diversity: low = 14.7, mid = 16.6, high = 15.7. p-values: richness 0.003, diversity 0.005). We next evaluated alpha-diversity (within-sample diversity) according to menstrual phases or contraceptive categories. No significant differences in alpha-diversity were observed across menstrual phases or between contraceptive methods. Higher richness (number of species) was observed in the luteal phase than in the menstrual phase (**Figure 3A**), but this did not prove to be significant by a narrow margin (p = 0.06), neither for the entire cohort nor for any of the contraceptive groups separately. There were no significant differences in diversity (Simpson's inverted index; **Figure 3B**). Total DNA amounts in the samples did not correlate with either richness or diversity (**Supplementary Table 3**). While both contraceptive method and phase of the menstrual cycle were statistically correlated to beta-diversity (between samples; **Table 2**), there was no distinct pattern of clustering by either of these factors (**Figure 4**; **Supplementary Figure 1**).

The separation between samples was mainly driven by gradients in the abundances of a few species. A gradient was evident with high levels of *N. subflava, N. mucosa, N. flavescens, R. mucilagniosa* and *H. parainfluenzae* on one side, and high levels *P. histicola, V. atypica* and *Prevotella* taxon 313 marked the other side. A second gradient displayed *P. pallens* and *P. melaninogenica* on one extreme, and chiefly *S. mitis* on the other. Interestingly, *P. pallens* and *P. melaninogenica* are not related to smoking status. No linear combination of taxa covered a large amount of the variation, with the first two principal components covering only 16% of total variance. Only when including the first 12 principal components was 50% of the total variance covered (**Supplementary Figure 1**).

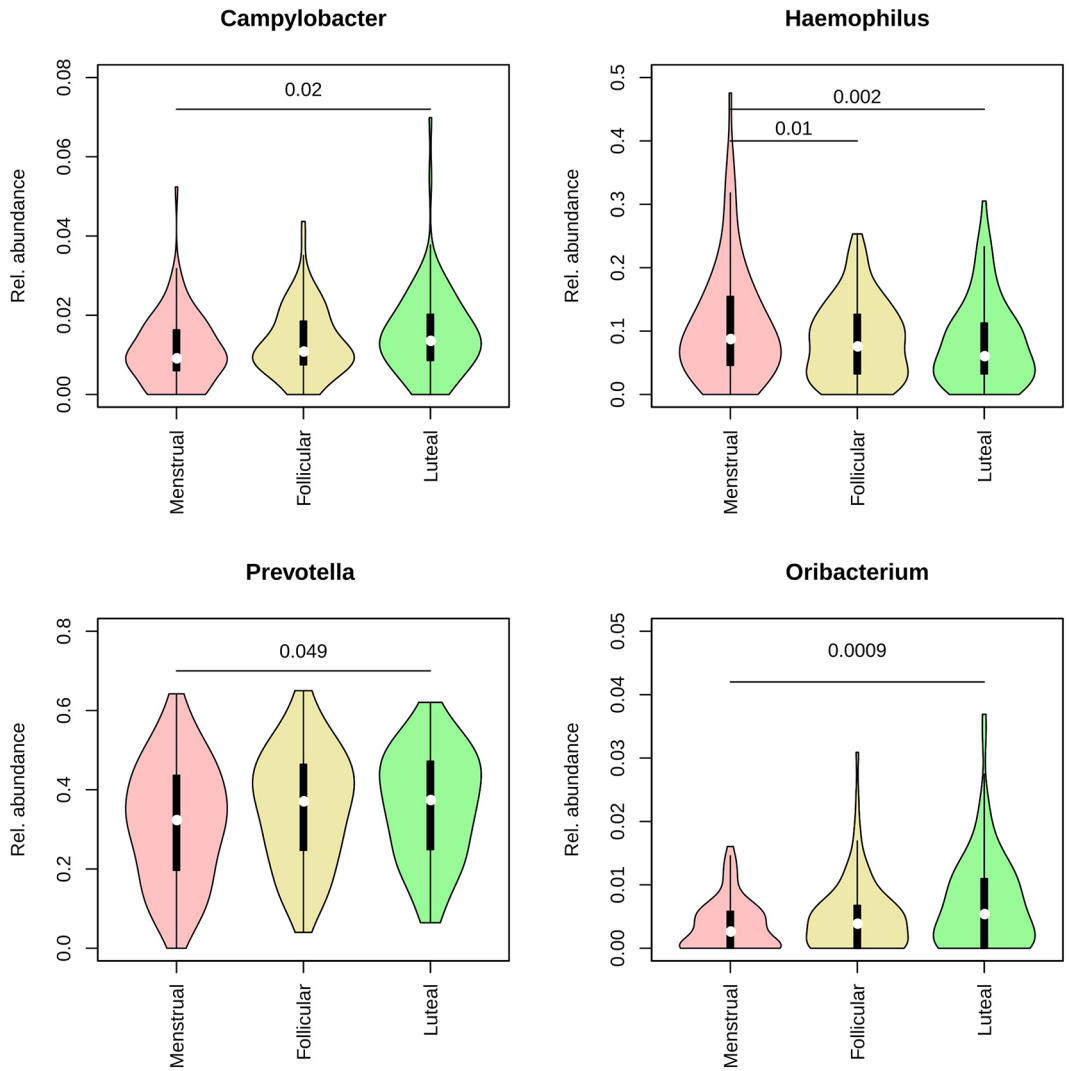


FIGURE 2 | Violin plots representing the distribution in relative abundance for each phase of the menstrual cycle for the four genera found to vary across the cycle. Pink: menstrual. Yellow: follicular. Green: luteal.

Specific Taxa Clustering

Furthermore, clustering of species was done according to the "color complex" classification (6 "color complex"—blue, green, yellow, purple, orange, and red—based on their frequency of detection) and according to the core subgingival microbiome classification (health-associated core species and periodontitis-associated core species in the subgingival microbiome) (28, 29). The blue, yellow, green, and purple complexes are associated with periodontal health, whereas the orange and red complexes are correlated with periodontitis (Socransky et al., 1998). The bacteria in the metagenomic dataset were grouped into the color complexes. The percentages of each of these complexes per sample were analyzed. The most prevalent and abundant groups in this cohort were the yellow (present in 298/300 women, median abundance 9.5%), orange (297/300, 10.7%), and purple complexes (287/300, 3.7%) (**Figure 5**).

Multivariate analysis of variance (MANOVA) revealed the yellow group to be associated to contraception and sugar intake, while the purple group was associated to smoking and sugar consumption. These two groups were analyzed further with a beta-regression, adjusting for subject as a random factor. This analysis found the yellow group to be increased in high and intermediate sugar consumption (high: $r = 0.31$, $p = 0.039$; intermediate: $r = 0.25$, $p = 0.05$). The purple group was found to be decreased in the IUS group ($r = -0.43$, $p = 0.016$) and vary across smoking groups, being decreased among daily smokers ($r = -0.57$, $p = 0.0055$), increased among former smokers ($r = 0.43$, $p = 0.0052$) and not changed for occasional smokers ($r = -0.06$, $p = 0.71$). Furthermore, in relation to the menstrual cycle, we observed: a) a higher abundance of the yellow complex during the menstrual phase, b) a tendency for

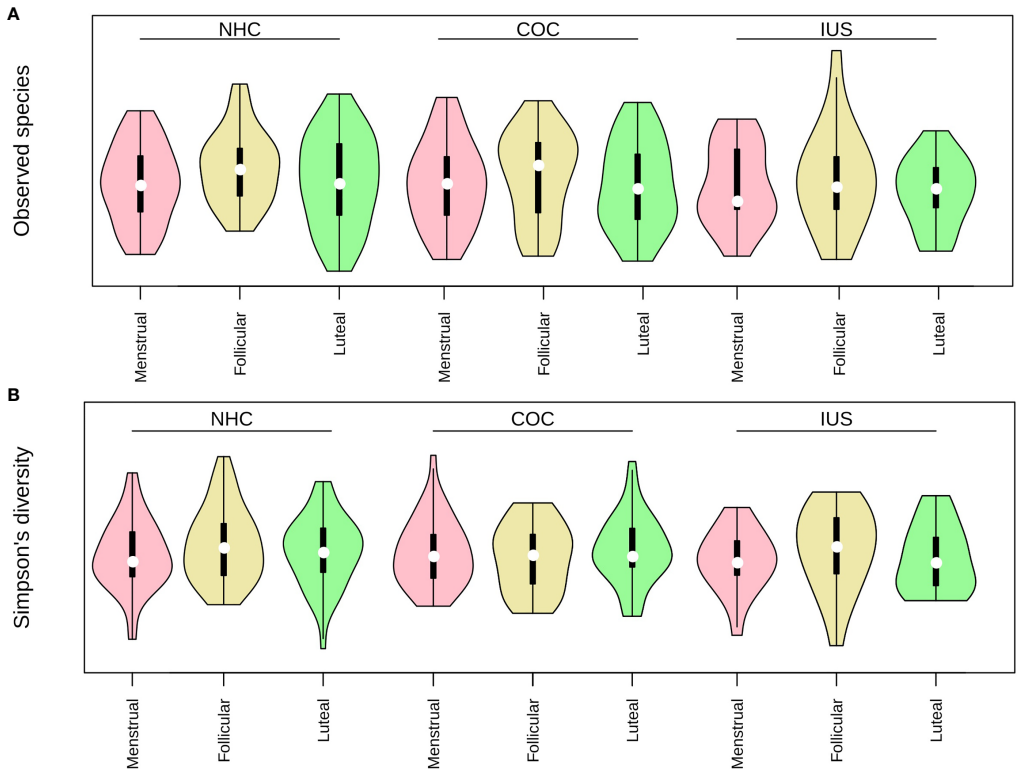


FIGURE 3 | **(A)** Richness (observed species) and **(B)** Diversity (inverted Simpson's index) in each contraceptive group and phase of the menstrual cycle. Pink: menstrual. Yellow: follicular. Green: luteal.

higher abundance of the red and green complexes during the follicular phase, and c) a lower abundance of the blue complex during the follicular phase (**Figure 5**). The interplay between each metadata factor and each of the colored groups is depicted as a heatmap in **Supplementary Figure 2**.

When classified according to the two core microbiome groupings (health- or disease-associated), there were also no notable differences in abundances between phases of the menstrual cycle (**Figure 6**), or between groups of contraceptives (**Supplementary Figure 3**). The variance in the abundance of health-associated bacteria within the COC group was more than twice greater than that of the other two groups (**Figure 6**; **Supplementary Figure 3**; COC, 0.012; NHC, 0.005; IUS 0.004; Levene's test, $p = 10^{-4}$). This difference was mostly driven by the dominance of the yellow complex, typically associated with periodontal health (**Figure 5**). Bacteria associated with caries were only found in low abundances (median 0, IQR 0–0.13%) and did not vary across the cycle (**Supplementary Table 4**). **Supplementary Table 4** lists all observed bacterial species and their association with health and disease.

Shifts of Functional Gene Composition

To examine the differential representation of particular microbial metabolic and biosynthetic pathways during the menstrual cycle or with different contraceptive use, the functional potential of each sample was assessed based on KEGG modules. These were submitted to Maaslin2 with sugar intake, smoking, cycle phase, and contraceptive as fixed effects and the individual as a random effect, as was done for the taxonomic annotation. A single module, pyrimidine deoxyribonucleotide biosynthesis, was increased for daily smokers, and a single module, Phosphatidylethanolamine biosynthesis, decreased with sugar consumption. 11 modules were found to be differentially abundant over the menstrual cycle (**Figure 7**; **Supplementary Table 5**). Amongst these are pathways involving co-factors, such as cobalamin biosynthesis and ascorbate degradation, as well as the biosynthesis of the neurotransmitter *gamma*-aminobutyric acid (GABA). Another 3 modules varied with contraceptive usage, namely the glyoxylate cycle, guanine biosynthesis and the NADH:quinone oxidoreductase step of oxidative phosphorylation (**Figure 7**). Overall, functional analysis indicated that more variation is attributed to the phase of menstrual cycle rather than contraceptive usage.

Hormonal Effects

Because standard hormonal measurements only measure endogenous estradiol and progesterone, not the ethinylestradiol, progestin and levonorgestrel used in COC and IUS, respectively, direct associations between female reproductive hormones and the

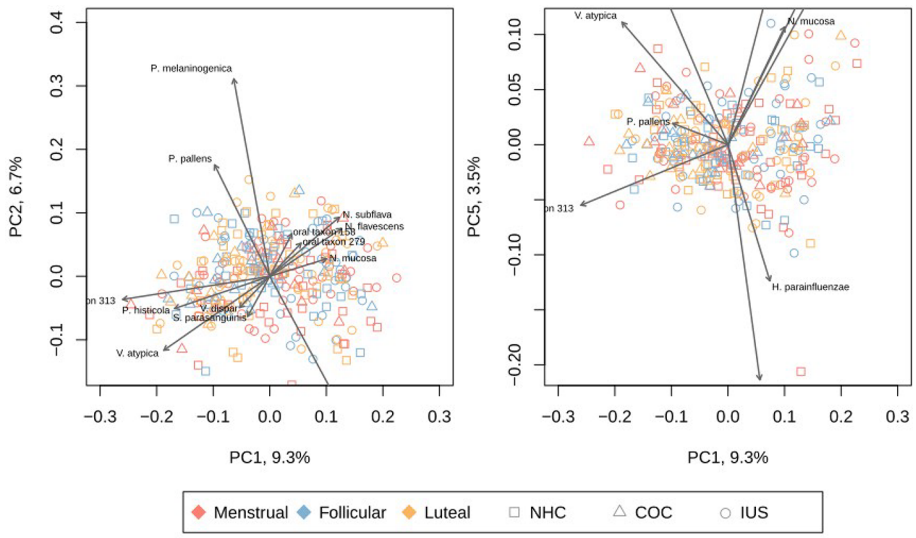


FIGURE 4 | Principal Coordinates Analysis (PCoA) of each sample, based on Bray-Curtis distance. The left panel depicts PC1 and PC2, while the right panel depicts PC1 and PC5. Other principal components are depicted in **Supplementary Figure 3**. The species with highest impact on the depicted principal components are overlaid as grey arrows. Red: menstrual phase. Yellow: follicular phase. Blue: luteal phase. Circle: non-hormonal contraceptives. Square: combined oral contraceptives. Triangle: intra-uterine levonegestrel system. H. parainfluenzae, Haemophilus parainfluenzae; N. flavescens, Neisseria flavescens; N. mucosa, Neisseria mucosa; N. subflava, Neisseria subflava; oral taxon 306, Prevotella sp. oral taxon 306; oral taxon 313, Prevotella sp. oral taxon 313; P. histicola, Prevotella histicola; P. melaninogenica, Prevotella melaninogenica; P. pallens, Prevotella pallens; R. dentocariosa, Rothia dentocariosa; R. mucilaginosa, Rothia mucilaginosa; S. mitis, Streptococcus mitis; S. parasanguinis, Streptococcus parasanguinis; S. salivarius, Streptococcus salivarius; V. atypica, Veillonella atypica; V. dispar, Veillonella dispar.

oral microbiome were only conducted for women not using hormonal contraception (Stanczyk and Clarke, 2010). The estradiol levels were higher in the follicular and luteal phase compared to the menstrual phase (Mean ± STDEV: 0.51 ± 0.24, 0.41 ± 0.31, 0.14 ± 0.05 nmol/L, respectively) (**Supplementary Table 3**). The progesterone concentrations were increased by 19-fold in the luteal phase compared to follicular and menstrual phases (Mean ± SDEV: 32.10 ± 20.45, 1.53 ± 1.21, 1.67 ± 2.78 nmol/L).

The MANOVA indicates that the red group is dependent on estradiol levels. The beta-regression adjusted for sugar intake and smoking did not converge, but visual inspection of the scatter plot between the red group and estradiol levels reveals that, for subjects where red group bacteria are present, they increase with increasing estradiol (**Supplementary Figure 4**). Adjusting for smoking and sugar intake, 24 species were correlated to estradiol and another 30 to progesterone, but only a single bacterial species was found to be significantly correlated to estradiol after correction for multiple testing, namely, *Porphyromonas endodontalis* (r = 0.0018, adj. p = 0.039)

DISCUSSION

In the present study, we longitudinally characterized the salivary microbiome of 103 regularly menstruating women in reproductive age during the course of one full menstrual cycle, using whole genome (shotgun) sequencing, and evaluated its potential deviations caused by the use of oral contraceptives and changes during the menstrual cycle. Such fluctuations may render the oral microbiome more dysbiotic, able to drive more aberrant inflammatory responses by the gingival tissues, a well recognizable clinical feature at phases of the menstrual cycle. To the best of our knowledge, this is the first large study to systematically address dysbiotic variations of the oral microbiome during the course of menstrual cycle, and document the additive effect of smoking and sugar consumption as exogenous risk factors. The analysis of this cohort cumulatively identified 50 bacterial genera belonging to eight phyla, all well represented in the human oral microbiome database and previously identified as members of the human salivary microbiome (Dewhirst et al., 2010; Segata et al., 2012; Hasan et al., 2014).This diversity may be considered much lower compared to most salivary microbiome studies that have been using 16S rRNA gene sequencing instead, a method which may nevertheless overestimate diversity. Conversely, high species-specificity was notable here, with 209 species being detectable in the studied cohort. An earlier metagenomic survey of saliva reports that the salivary microbiome typically contains 175 bacterial species (Hasan et al., 2014). It estimated that the number of species-level phylotypes may vary from 500 to 10,000 and each oral niche may harbor on average 266 species level phylotypes (Zaura et al., 2009; Bik et al., 2010; Hasan et al., 2014). The lack of significant differences in taxa diversities across phases of the menstrual cycle, or use of contraceptives, denote

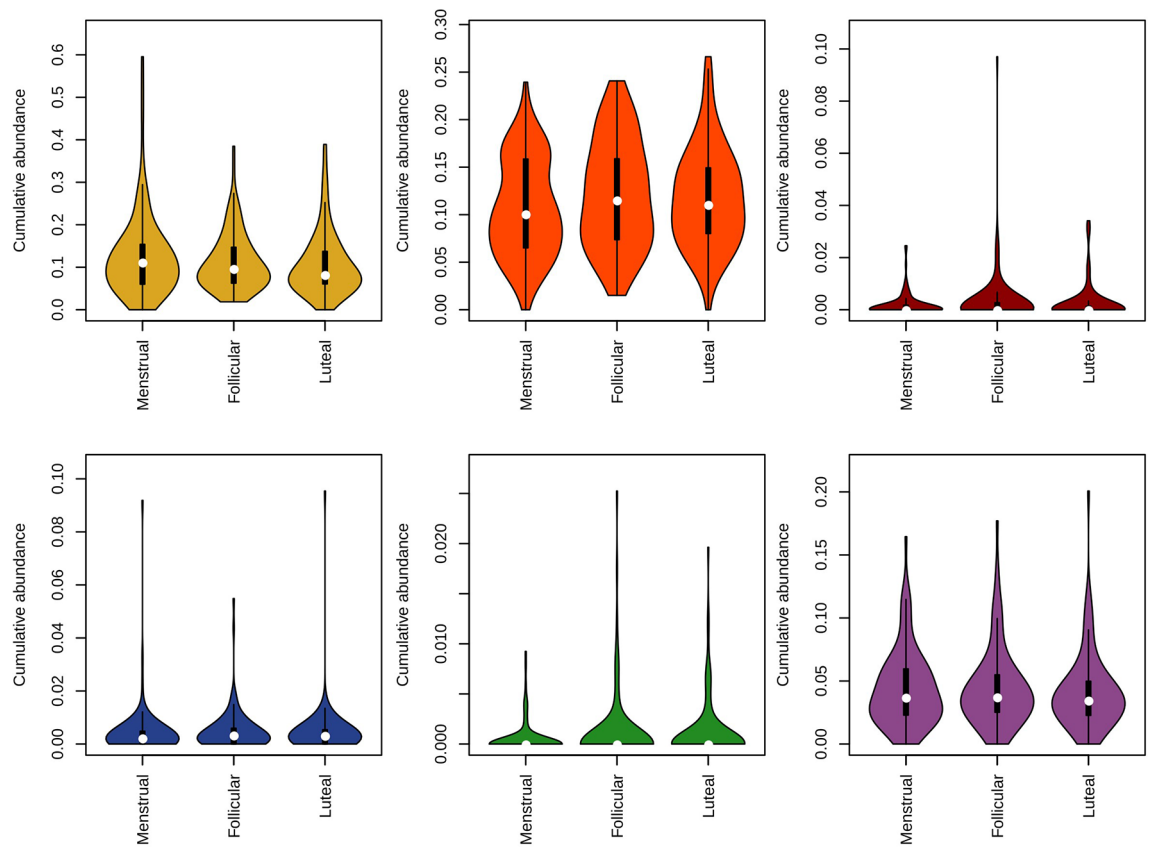


FIGURE 5 | Violin plot representing the distribution in relative abundance for each phase of the menstrual cycle method for species of the Socransky complexes. The blue, yellow, green, and purple complexes are associated with periodontal health, whereas the orange and red complexes are correlated with periodontal disease.

the less susceptibility of the oral cavity to inherent biological or external pharmacological pressures, respectively (Zaura et al., 2014; Zaura et al., 2017). Dietary habits, including sugar intake, appeared to influence the composition of the salivary microbiome during the menstrual cycle, as evaluated by weekly dietary records. The finding that diet did influence the community composition of the salivary microbiome may not come as a surprise, as it has been shown that the influence of diet occurs also at the metabolome level (De Filippis et al., 2014; Tanner et al., 2018). Specific dietary habits i.e., high frequency of carbohydrate exposure can select for more acid-tolerant and acidogenic bacteria such as *Streptococci, Lactobacilli or Bifidobacteria*, which may in turn disturb the enamel mineral equilibrium, leading to irreversible demineralization and dental caries. Still, pairing metagenomic data with conventional nutrient profiles may not be sufficient to infer microbiome variations to diet (Johnson et al., 2019).

Despite the overall relative stabilities in diversity and overall abundance of the microbiome profiles in saliva, some taxon-function specific and significant changes were observed. In particular, four genera were found to differ in abundance throughout the menstrual cycle, namely *Campylobacter*, *Haemophilus*, *Prevotella*, and *Oribacterium*. Higher abundance of *Prevotella* species in saliva may increase pH and stimulate the flow of gingival crevicular fluid (Raber-Durlacher et al., 1994). These changes may favor acid-intolerant, proteolytic species associated with gingival inflammation. *P. intermedia* growth and biosynthetic activity appear to be regulated by progesterone and estradiol, two cycle hormones, as demonstrated in experimental models (Kornman and Loesche, 1982).

The salivary and plasma dynamics of sex steroids seems to mirror one-another across the menstrual cycle, with progesterone levels peaking during the luteal phase and estradiol levels during the follicular phase (Gandara et al., 2007). Further, the estradiol levels in plasma positively correlated with the total number of the bacteria from the red complex bacteria (*P. gingivalis*, *T. denticola*, *T. forsythia)* in women not taking exogenous sex steroid hormones (those found in hormonal contraceptives) (Clark and Soory, 2006). The presence of *P gingivalis* has been associated with gingival inflammation during menstruation and pregnancy and positively correlated with the increase in sex hormones in saliva (Muramatsu and Takaesu, 1994; Carrillo-de-Albornoz et al., 2010). In addition to the red complex bacteria, the closely

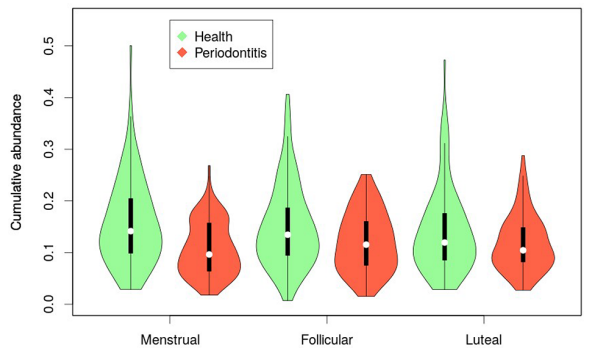


FIGURE 6 | Violin plot representing the distribution in relative abundance for each phase of the menstrual cycle according to the subgingival core microbiome classification (green: health-associated; red: periodontitis-associated).

related *P. endodontalis* also proved to be significantly correlated with increasing estradiol (Lombardo Bedran et al., 2012). Although estradiol has been reported to exert both pro- or anti- inflammatory responses in oral mucosa or modify growth of specific oral species at a dose dependent manner, yet the mechanisms by which this occurs have not been explored in depth. We also observed that the abundancies of *Prevotella* and *Veillonella* were increased in current smokers, as also demonstrated earlier (Kumar, 2013; Paropkari et al., 2016). These changes may not be surprising as smoking leads to a diverse, pathogen-rich anaerobic oral microbiome and depletion of commensals, hence creating an at-risk-for-harm environment for the development of oral diseases (Shchipkova et al., 2010). This additive risk could may well apply to the female population, where hormonal regulations are concurrent with the transient establishment of a dysbiotic microbiota (Paropkari et al., 2016).

While there is a circumstantial body of evidence linking the menstrual cycle to changes in the microbiome of the dental plaque, there is no clear consensus and the sample size of the cohorts were rather limited. An association between ovulation and the increased levels of anaerobic bacterial counts in saliva has been supported (Prout and Hopps, 1970) whereas others did not conclude on a cyclical pattern of subgingival bacterial colonization of any of the 74 species studied (Fischer et al., 2008). Yet, the study has described that *Aggregatibacter actinomycetemcomitans*, a species closely related to the *Haemophilus* genus, was commonly detected at the beginning of the menstruation and peaked during the following 2 weeks. While *A. actinomycetemcomitans* itself was not frequently identified in the present study, *Haemophilus* was among the four differentially abundant genera through the cycle, represented by species other than *A. actinomycetemcomitans*. *Campylobacter rectus* (*C. rectus*) is a gram-negative motile rod associated with periodontal diseases, whose growth can be enhanced by estradiol (Yokoyama et al., 2005; Bostanci et al., 2007; Yokoyama et al., 2008). Young pregnant or postpartum women carry high levels of *C. rectus* in subgingival plaques or saliva (Mitchell-Lewis et al., 2001; Yokoyama et al., 2005). A sharp rise in salivary estradiol levels is shown to occur immediately at the ovulation phase (Lu et al., 1999), and such hormonal peaks may indeed explain the significant fluctuation in the abundances of *Campylobacter* and *Prevotella* genera during the progression of the menstrual cycle observed in the present study. While there are at present no reports linking the presence or growth of *Oribacterium* to progesterone and estradiol, it is interesting to note that the occurrence of this genus has been associated with oral malodor, or halitosis (Sizova et al., 2014; Seerangaiyan et al., 2017). The menstrual cycle has also been pointed out as a factor influencing halitosis. Distinct cyclic variations in volatile sulphur compound concentrations occur during the menstrual cycle, which seem to coincide with the mid-cycle surge of the luteinizing hormone and the mid-luteal phase, corresponding to a peak of progesterone and estrogens, respectively (Tonzetich et al., 1978; Kawamoto et al., 2010).

The present study did not identify any major influences of the intake of oral contraceptives in the composition of the salivary microbiome composition. This is further supported by earlier studies demonstrating that the magnitude of gingival inflammation is not affected by various oral contraceptive formulations (Preshaw et al., 2001). Yet, improved oral hygiene may compensate for the potential hormonal influence on the oral microbiome due to oral contraceptive intake (Preshaw et al., 2001; Preshaw and Bissett, 2013). *Rothia, Haemophilus*, and *Neisseria* were highly abundant among the samples. These genera are known to be associated with good oral health, a finding that is aligned with the data obtained from the questionnaires (Palmer et al., 2017). A few early studies attempted to shed light on potential relationships between oral contraceptives and the oral microbiota, demonstrating a higher percentage of *Bacteroides* spp., such as *Prevotella intermedia* (previously *B. intermedius*), in women under a hormonal contraception regimen (Jensen et al., 1981).

It is becoming increasingly apparent that determining the structural composition of the oral microbiome delivers a finite amount of information, and that evaluating its function is crucial to understanding its totality (Espinoza et al., 2018). The present study identified selective functional variations of the salivary microbiome during menstrual phases, revealing its adaptability to hormonal fluctuations, while the overall community remains structurally intact. The most prevalent microbial functional modules were highly consistent within and across subjects. Yet, a subset of functions significantly differed between cycle phases, but also between contraceptives users. Interestingly, we found that the capacity for the biosynthesis of GABA was decreased during the menstrual phase. GABA is a neurotransmitter of major importance to inhibit functionality in the brain. GABAergic deficits are suggested to contribute to poor mental health and reduced GABA levels have been found in depressed patients, both centrally and in the periphery. Premenstrual dysphoric disorder has been linked to alterations in systems related to GABA (AC et al., 2006; Hofmeister and Bodden, 2016), however most significantly shown in the luteal phase. Notably, there are also previous reports on elevated inflammation as well as subjective health symptoms to be present in a higher level

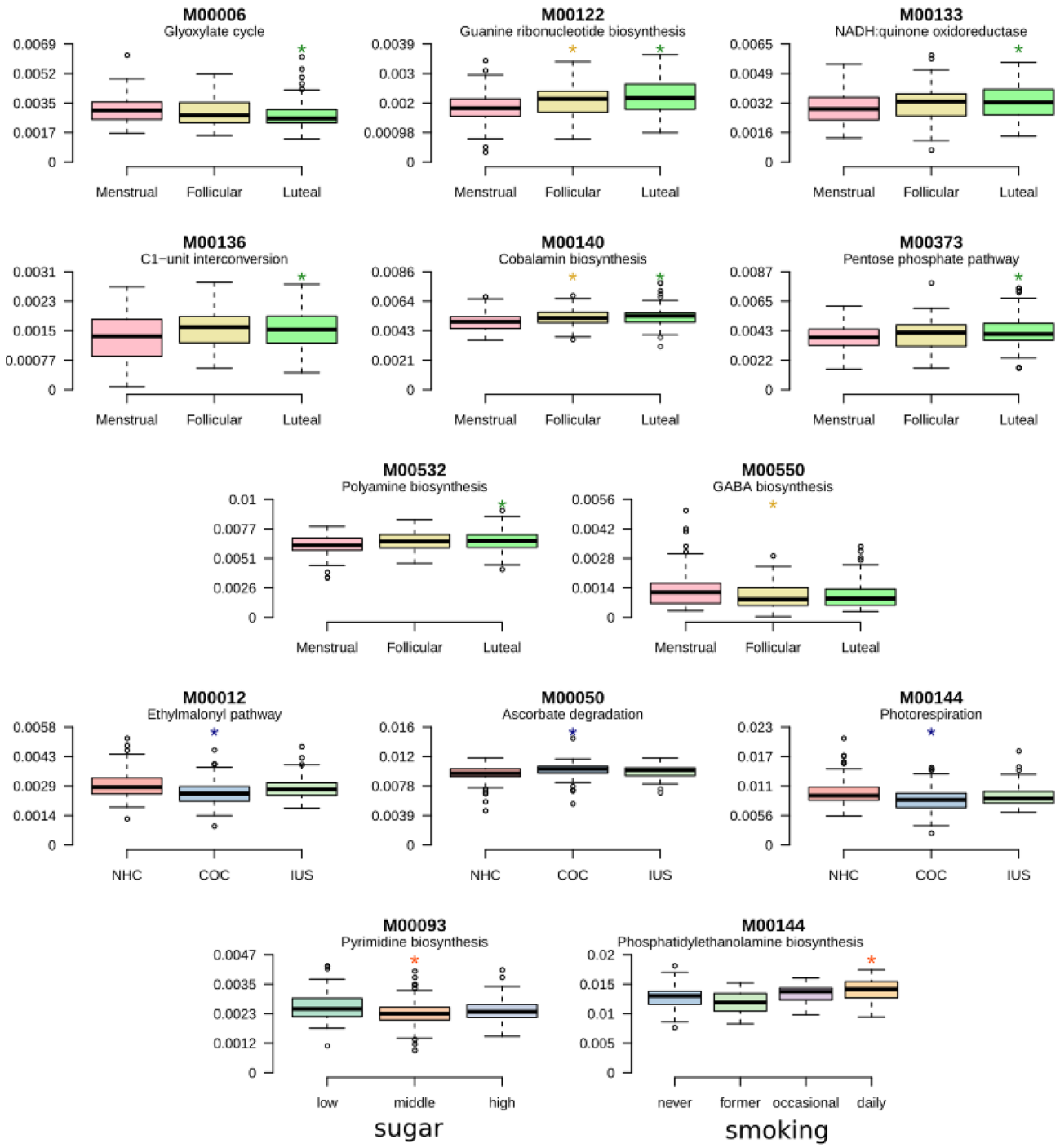


FIGURE 7 | Shifts of functional gene composition. Boxplots presenting pathways that are differentially abundant during the menstrual cycle or according to contraceptive usage. Detailed results are presented in **Supplementary Table 5**.

during the menstrual phase, also in women with no premenstrual syndrome (Puder et al., 2006). Smoking and sugar consumption seem to associated with differences in metabolic profile of the salivary microbiome. Most significantly associated with smoking was the pyrimidine deoxyribonucleotide biosynthesis, which was evidently higher in smokers, and high sugar consumption, which was associated with reduced phosphatidylethanolamine biosynthesis. It is interesting to note that pentose phosphate

metabolism is frequently increased in Gram-positive pathogens such as *Streptococcus* and *Actinomyces* in response to environmental stresses. In the present data set, gram positive bacteria such as *S. mitis* and *S. oralis* were reduced in abundance during the luteal phase that had low gene abundance in the pentose phosphate pathway. On the contrary, genes related to membrane transport including the phosphotransferase system, RTX-toxin transport, RaxAB-RaxC type I secretion were over-represented during the luteal phase, which may be related to bacterial chemotaxis and increased toxin biosynthesis. Although the differences in gene richness did not seem to correspond to differences in bacterial species richness, this is in line with an earlier study indicating that 50% of all genes in a metagenomic sample are individual-specific and the functional differences between individuals are larger than the taxonomic differences (Tierney et al., 2019). A number of species within the oral microbiome community can metabolically exploit hormones as carbon and energy sources, processing them by degradation or chemical modification (Takahashi et al., 2010; Garcia-Gomez et al., 2013; Kumar, 2013).

In conclusion, we longitudinally characterized the ecological shifts associated with hormonal fluctuations in the salivary microbiome of regularly menstruating women, during the course of one full menstrual cycle. This is the first large study to systematically address dysbiotic variations of the oral microbiome during the course of menstrual cycle, and document the additive effect of smoking and sugar consumption as environmental risk factors. It reveals the structural resilience and functional adaptability of the oral microbiome to the endogenous hormonal pressures of the menstrual cycle, while revealing its vulnerability to the exogenous exposures of diet and smoking.

DATA AVAILABILITY STATEMENT

All sequencing data analysed in this study are available from the European Nucleotide Archive under project PRJEB37731, samples SAMEA6662389-SAMEA6662857.

ETHICS STATEMENT

The participants gave both oral and written consent to participate in this study.

AUTHOR CONTRIBUTIONS

MK, ZB, and HN obtained ethics and data protection approval, wrote the procotol, planned and organized the study cohort, included participants and secured informed consent and collected samples. LH, NB, FB, and GB analyzed the sequencing data and wrote the manuscript. IS-K planned experiments and wrote the manuscript. EF and LE wrote the manuscript and supervised students. KW contributed to conception and design of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The Centre for Translational Microbiome Research is partly funded by Ferring Pharmaceuticals (LH, FB, LE, IS-K, and EF). A research grant from Ferring Pharmaceuticals enabled the clinical recruitment and sampling (HN). The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication. The Rigshospitalet Research Fund (HN and MK), Karolinska Institute Strategic Funds (NB and GB), the Swedish Research Council (NB), and KI/SLL Strategic Dental Research Fund (NB and GB).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2021.625229/full#supplementary-material

REFERENCES

Abusleme, L., Dupuy, A. K., Dutzan, N., Silva, N., Burleson, J. A., Strausbaugh, L. D., et al. (2013). The subgingival microbiome in health and periodontitis and its relationship with community biomass and inflammation. *ISME J.* 7, 1016–1025. doi: 10.1038/ismej.2012.174

AC, N. W., Sundstrom-Poromaa, I., and Backstrom, T. (2006). Action by and sensitivity to neuroactive steroids in menstrual cycle related CNS disorders. *Psychopharmacol. (Berl)* 186, 388–401. doi: 10.1007/s00213-005-0185-2

Belibasakis, G. N., Bostanci, N., Marsh, P. D., and Zaura, E. (2019). Applications of the oral microbiome in personalized dentistry. *Arch. Oral. Biol.* 104, 7–12. doi: 10.1016/j.archoralbio.2019.05.023

Bik, E. M., Long, C. D., Armitage, G. C., Loomer, P., Emerson, J., Mongodin, E. F., et al. (2010). Bacterial diversity in the oral cavity of 10 healthy individuals. *ISME J.* 4, 962–974. doi: 10.1038/ismej.2010.30

Bostanci, N., Allaker, R. P., Belibasakis, G. N., Rangarajan, M., Curtis, M. A., Hughes, F. J., et al. (2007). Porphyromonas gingivalis antagonises Campylobacter rectus induced cytokine production by human monocytes. *Cytokine* 39, 147–156. doi: 10.1016/j.cyto.2007.07.002

Brusca, M. I., Rosa, A., Albaina, O., Moragues, M. D., Verdugo, F., and Ponton, J. (2010). The impact of oral contraceptives on women's periodontal health and the subgingival occurrence of aggressive periodontopathogens and Candida species. *J. Periodontol.* 81, 1010–1018. doi: 10.1902/jop.2010.090575

Carrillo-de-Albornoz, A., Figuero, E., Herrera, D., and Bascones-Martinez, A. (2010). Gingival changes during pregnancy: II. Influence of hormonal variations on the subgingival biofilm. *J. Clin. Periodontol.* 37, 230–240. doi: 10.1111/j.1600-051X.2009.01514.x

Clancy, K. B., Baerwald, A. R., and Pierson, R. A. (2013). Systemic inflammation is associated with ovarian follicular dynamics during the human menstrual cycle. *PloS One* 8, e64807. doi: 10.1371/journal.pone.0064807

Clark, D. T., and Soory, M. (2006). The metabolism of cholesterol and certain hormonal steroids by Treponema denticola. *Steroids* 71, 352–363. doi: 10.1016/j.steroids.2005.11.006

David, L. A., Materna, A. C., Friedman, J., Campos-Baptista, M. I., Blackburn, M. C., Perrotta, A., et al. (2014). Host lifestyle affects human microbiota on daily timescales. *Genome Biol.* 15, R89. doi: 10.1186/gb-2014-15-7-r89

De Filippis, F., Vannini, L., La Storia, A., Laghi, L., Piombino, P., Stellato, G., et al. (2014). The same microbiota and a potentially discriminant metabolome in the saliva of omnivore, ovo-lacto-vegetarian and Vegan individuals. *PloS One* 9, e112373. doi: 10.1371/journal.pone.0112373

Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C., Yu, W. H., et al. (2010). The human oral microbiome. *J. Bacteriol.* 192, 5002–5017. doi: 10.1128/JB.00542-10

Espinoza, J. L., Harkins, D. M., Torralba, M., Gomez, A., Highlander, S. K., Jones, M. B., et al. (2018). Supragingival Plaque Microbiome Ecology and Functional Potential in the Context of Health and Disease. *mBio* 9 (6), e01631–18. doi: 10.1128/mBio.01631-18

Fischer, C. C., Persson, R. E., and Persson, G. R. (2008). Influence of the menstrual cycle on the oral microbial flora in women: a case-control study including men as control subjects. *J. Periodontol.* 79, 1966–1973. doi: 10.1902/jop.2008.080057

Gandara, B. K., Leresche, L., and Mancl, L. (2007). Patterns of salivary estradiol and progesterone across the menstrual cycle. *Ann. N. Y. Acad. Sci.* 1098, 446–450. doi: 10.1196/annals.1384.022

Garcia-Gomez, E., Gonzalez-Pedrajo, B., and Camacho-Arroyo, I. (2013). Role of sex steroid hormones in bacterial-host interactions. *BioMed. Res. Int.* 2013, 928290. doi: 10.1155/2013/928290

Gumus, P., Emingil, G., Ozturk, V. O., Belibasakis, G. N., and Bostanci, N. (2015). Oxidative stress markers in saliva and periodontal disease status: modulation during pregnancy and postpartum. *BMC Infect. Dis.* 15, 261. doi: 10.1186/s12879-015-1003-z

Gursoy, M., Pajukanta, R., Sorsa, T., and Kononen, E. (2008). Clinical changes in periodontium during pregnancy and post-partum. *J. Clin. Periodontol.* 35, 576–583. doi: 10.1111/j.1600-051X.2008.01236.x

Hasan, N. A., Young, B. A., Minard-Smith, A. T., Saeed, K., Li, H., Heizer, E. M., et al. (2014). Microbial community profiling of human saliva using shotgun metagenomic sequencing. *PloS One* 9, e97699. doi: 10.1371/journal.pone.0097699

Hofmeister, S., and Bodden, S. (2016). PremenstrualSyndrome and Premenstrual Dysphoric Disorder. *Am. Fam. Physician* 94, 236–240.

Holm-Pedersen, P., and Loe, H. (1967). Flow of gingival exudate as related to menstruation and pregnancy. *J. Periodontal. Res.* 2, 13–20. doi: 10.1111/j.1600-0765.1967.tb01991.x

Human Microbiome Project, C. (2012). A framework for human microbiome research. *Nature* 486, 215–221. doi: 10.1038/nature11209

Jensen, J., Liljemark, W., and Bloomquist, C. (1981). The effect of female sex hormones on subgingival plaque. *J. Periodontol.* 52, 599–602. doi: 10.1902/jop.1981.52.10.599

Johnson, A. J., Vangay, P., Al-Ghalith, G. A., Hillmann, B. M., Ward, T. L., Shields-Cutler, R. R., et al. (2019). Daily Sampling Reveals Personalized Diet-Microbiome Associations in Humans. *Cell Host Microbe* 25, 789–802. doi: 10.1016/j.chom.2019.05.005

Kawamoto, A., Sugano, N., Motohashi, M., Matsumoto, S., and Ito, K. (2010). Relationship between oral malodor and the menstrual cycle. *J. Periodontal. Res.* 45, 681–687. doi: 10.1111/j.1600-0765.2010.01285.x

Kornman, K. S., and Loesche, W. J. (1982). Effects of estradiol and progesterone on Bacteroides melaninogenicus and Bacteroides gingivalis. *Infect. Immun.* 35, 256–263. doi: 10.1128/IAI.35.1.256-263.1982

Kumar, P. S. (2013). Sex and the subgingival microbiome: do female sex steroids affect periodontal bacteria? *Periodontol* 61, 103–124. doi: 10.1111/j.1600-0757.2011.00398.x

Lif Holgerson, P., Esberg, A., Sjodin, A., West, C. E., and Johansson, I. (2020). A longitudinal study of the development of the saliva microbiome in infants 2 days to 5 years compared to the microbiome in adolescents. *Sci. Rep.* 10, 9629. doi: 10.1038/s41598-020-66658-7

Lombardo Bedran, T. B., Marcantonio, R. A., Spin Neto, R., Alves Mayer, M. P., Grenier, D., Spolidorio, L. C., et al. (2012). Porphyromonas endodontalis in chronic periodontitis: a clinical and microbiological cross-sectional study. *J. Oral. Microbiol.* 4, 10123. doi: 10.3402/jom.v4i0.10123

Lu, Y., Bentley, G. R., Gann, P. H., Hodges, K. R., and Chatterton, R. T. (1999). Salivary estradiol and progesterone levels in conception and nonconception cycles in women: evaluation of a new assay for salivary estradiol. *Fertil. Steril.* 71, 863–868. doi: 10.1016/S0015-0282(99)00093-X

Lukacs, J. R., and Largaespada, L. L. (2006). Explaining sex differences in dental caries prevalence: saliva, hormones, and "life-history" etiologies. *Am. J. Hum. Biol.* 18, 540–555. doi: 10.1002/ajhb.20530

Mitchell-Lewis, D., Engebretson, S. P., Chen, J., Lamster, I. B., and Papapanou, P. N. (2001). Periodontal infections and pre-term birth: early findings from a cohort of young minority women in New York. *Eur. J. Oral. Sci.* 109, 34–39. doi: 10.1034/j.1600-0722.2001.00966.x

Muramatsu, Y., and Takaesu, Y. (1994). Oral health status related to subgingival bacterial flora and sex hormones in saliva during pregnancy. *Bull. Tokyo Dent. Coll.* 35, 139–151.

Oertelt-Prigione, S. (2012). Immunology and the menstrual cycle. *Autoimmun. Rev.* 11, A486–A492. doi: 10.1016/j.autrev.2011.11.023

Palmer, R. J. Jr., Shah, N., Valm, A., Paster, B., Dewhirst, F., Inui, T., et al. (2017). Interbacterial Adhesion Networks within Early Oral Biofilms of Single Human Hosts. *Appl. Environ. Microbiol.* 83 (11), e00407–17. doi: 10.1128/AEM.00407-17

Paropkari, A. D., Leblebicioglu, B., Christian, L. M., and Kumar, P. S. (2016). Smoking, pregnancy and the subgingival microbiome. *Sci. Rep.* 6, 30388. doi: 10.1038/srep30388

Preshaw, P. M., and Bissett, S. M. (2013). Periodontitis: oral complication of diabetes. *Endocrinol. Metab. Clin. North Am.* 42, 849–867. doi: 10.1016/j.ecl.2013.05.012

Preshaw, P. M., Knutsen, M. A., and Mariotti, A. (2001). Experimental gingivitis in women using oral contraceptives. *J. Dent. Res.* 80, 2011–2015. doi: 10.1177/00220345010800111201

Preshaw, P. M., Holliday, R., Law, H., and Heasman, P. A. (2013). Outcomes of non-surgical periodontal treatment by dental hygienists in training: impact of site- and patient-level factors. *Int. J. Dent. Hyg.* 11, 273–279. doi: 10.1111/idh.12032

Prout, R. E., and Hopps, R. M. (1970). A relationship between human oral bacteria and the menstrual cycle. *J. Periodontol.* 41, 98–101. doi: 10.1902/jop.1970.41.2.98

Puder, J. J., Blum, C. A., Mueller, B., De Geyter, C., Dye, L., and Keller, U. (2006). Menstrual cycle symptoms are associated with changes in low-grade inflammation. *Eur. J. Clin. Invest.* 36, 58–64. doi: 10.1111/j.1365-2362.2006.01591.x

Raber-Durlacher, J. E., van Steenbergen, T. J., Van der Velden, U., de Graaff, J., and Abraham-Inpijn, L. (1994). Experimental gingivitis during pregnancy and post-partum: clinical, endocrinological, and microbiological aspects. *J. Clin. Periodontol.* 21, 549–558. doi: 10.1111/j.1600-051X.1994.tb01172.x

Rosier, B. T., Marsh, P. D., and Mira, A. (2018). Resilience of the Oral Microbiota in Health: Mechanisms That Prevent Dysbiosis. *J. Dent. Res.* 97, 371–380. doi: 10.1177/0022034517742139

Saibaba, G., Srinivasan, M., Priya Aarthy, A., Silambarasan, V., and Archunan, G. (2017). Ultrastructural and physico-chemical characterization of saliva during menstrual cycle in perspective of ovulation in human. *Drug Discovery Ther.* 11, 91–97. doi: 10.5582/ddt.2017.01008

Seerangaiyan, K., van Winkelhoff, A. J., Harmsen, H. J. M., Rossen, J. W. A., and Winkel, E. G. (2017). The tongue microbiome in healthy subjects and patients with intra-oral halitosis. *J. Breath Res.* 11, 036010. doi: 10.1088/1752-7163/aa7c24

Segata, N., Haake, S. K., Mannon, P., Lemon, K. P., Waldron, L., Gevers, D., et al. (2012). Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol.* 13, R42. doi: 10.1186/gb-2012-13-6-r42

Shchipkova, A. Y., Nagaraja, H. N., and Kumar, P. S. (2010). Subgingival microbial profiles of smokers with periodontitis. *J. Dent. Res.* 89, 1247–1253. doi: 10.1177/0022034510377203

Silva de Araujo Figueiredo, C., Goncalves Carvalho Rosalem, C., Costa Cantanhede, A. L., Abreu Fonseca Thomaz, E. B., and Fontoura Nogueira da Cruz, M. C. (2017). Systemic alterations and their oral manifestations in pregnant women. *J. Obstet. Gynaecol. Res.* 43, 16–22. doi: 10.1111/jog.13150

Sizova, M. V., Muller, P. A., Stancyk, D., Panikov, N. S., Mandalakis, M., Hazen, A., et al. (2014). Oribacterium parvum sp. nov. and Oribacterium asaccharolyticum sp. nov., obligately anaerobic bacteria from the human oral cavity, and emended description of the genus Oribacterium. *Int. J. Syst. Evol. Microbiol.* 64, 2642–2649. doi: 10.1099/ijs.0.060988-0

Socransky, S. S., Haffajee, A. D., Cugini, M. A., Smith, C., and Kent, R. L.Jr. (1998). Microbial complexes in subgingival plaque. *J. Clin. Periodontol.* 25, 134–144. doi: 10.1111/j.1600-051X.1998.tb02419.x

Stanczyk, F. Z., and Clarke, N. J. (2010). Advantages and challenges of mass spectrometry assays for steroid hormones. *J. Steroid Biochem. Mol. Biol.* 121, 491–495. doi: 10.1016/j.jsbmb.2010.05.001

Streckfus, C. F., Baur, U., Brown, L. J., Bacal, C., Metter, J., and Nick, T. (1998). Effects of estrogen status and aging on salivary flow rates in healthy Caucasian women. *Gerontology* 44, 32–39. doi: 10.1159/000021980

Takahashi, N., Washio, J., and Mayanagi, G. (2010). Metabolomics of supragingival plaque and oral bacteria. *J. Dent. Res.* 89, 1383–1388. doi: 10.1177/0022034510377792

Tanner, A. C. R., Kressirer, C. A., Rothmiller, S., Johansson, I., and Chalmers, N. I. (2018). The Caries Microbiome: Implications for Reversing Dysbiosis. *Adv. Dent. Res.* 29, 78–85. doi: 10.1177/0022034517736496

Tierney, B. T., Yang, Z., Luber, J. M., Beaudin, M., Wibowo, M. C., Baek, C., et al. (2019). The Landscape of Genetic Content in the Gut and Oral Human Microbiome. *Cell Host Microbe* 26, 283–295. doi: 10.1016/j.chom.2019.07.008

Tonzetich, J., Preti, G., and Huggins, G. R. (1978). Changes in concentration of volatile sulphur compounds of mouth air during the menstrual cycle. *J. Int. Med. Res.* 6, 245–254. doi: 10.1177/030006057800600313

Wade, W. G. (2013). The oral microbiome in health and disease. *Pharmacol. Res.* 69, 137–143. doi: 10.1016/j.phrs.2012.11.006

Yeh, C. K., Johnson, D. A., and Dodds, M. W. (1998). Impact of aging on human salivary gland function: a community-based study. *Aging (Milano)* 10, 421–428. doi: 10.1007/BF03339889

Yokoyama, M., Hinode, D., Masuda, K., Yoshioka, M., and Grenier, D. (2005). Effect of female sex hormones on Campylobacter rectus and human gingival fibroblasts. *Oral. Microbiol. Immunol.* 20, 239–243. doi: 10.1111/j.1399-302X.2005.00222.x

Yokoyama, M., Hinode, D., Yoshioka, M., Fukui, M., Tanabe, S., Grenier, D., et al. (2008). Relationship between Campylobacter rectus and periodontal status during pregnancy. *Oral. Microbiol. Immunol.* 23, 55–59. doi: 10.1111/j.1399-302X.2007.00391.x

Zaura, E., Keijser, B. J., Huse, S. M., and Crielaard, W. (2009). Defining the healthy "core microbiome" of oral microbial communities. *BMC Microbiol.* 9, 259. doi: 10.1186/1471-2180-9-259

Zaura, E., Nicu, E. A., Krom, B. P., and Keijser, B. J. (2014). Acquiring and maintaining a normal oral microbiome: current perspective. *Front. Cell Infect. Microbiol.* 4, 85. doi: 10.3389/fcimb.2014.00085

Zaura, E., Brandt, B. W., Prodan, A., Teixeira de Mattos, M. J., Imangaliyev, S., Kool, J., et al. (2017). On the ecosystemic network of saliva in healthy young adults. *ISME J.* 11, 1218–1231. doi: 10.1038/ismej.2016.199

Conflict of Interest: Author KW was employed by Ferring International Center SA.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cellular and Infection Microbiology

ORIGINAL RESEARCH
published: 22 March 2021
doi: 10.3389/fcimb.2021.629723

The Oral Microbiome in Periodontal Health

Magdalena Lenartova[1,2], *Barbora Tesinska*[1,2], *Tatjana Janatova*[2,3], *Ondrej Hrebicek*[1], *Jaroslav Mysak*[3], *Jiri Janata*[1,4] *and Lucie Najmanova*[1,4]*

[1] *Institute of Microbiology, Academy of Sciences of the Czech Republic, Prague, Czechia,* [2] *Department of Genetics and Microbiology, Faculty of Science, Charles University, Prague, Czechia,* [3] *Institute of Dental Medicine, First Faculty of Medicine, Charles University and General University Hospital, Prague, Czechia,* [4] *Institute of Microbiology v. v. i., BIOCEV, Academy of Sciences of the Czech Republic, Vestec, Czechia*

The estimation of oral microbiome (OM) taxonomic composition in periodontally healthy individuals can often be biased because the clinically periodontally healthy subjects for evaluation can already experience dysbiosis. Usually, they are included just based on the absence of clinical signs of periodontitis. Additionally, the age of subjects is used to be higher to correspond well with tested groups of patients with chronic periodontitis, a disorder typically associated with aging. However, the dysbiosis of the OM precedes the clinical signs of the disease by many months or even years. The absence of periodontal pockets thus does not necessarily mean also good periodontal health and the obtained image of "healthy OM" can be distorted.To overcome this bias, we taxonomically characterized the OM in almost a hundred young students of dentistry with precise oral hygiene and no signs of periodontal disease. We compared the results with the OM composition of older periodontally healthy individuals and also a group of patients with severe periodontitis (aggressive periodontitis according to former classification system). The clustering analysis revealed not only two compact clearly separated clusters corresponding to each state of health, but also a group of samples forming an overlap between both well-pronounced states. Additionally, in the cluster of periodontally healthy samples, few outliers with atypical OM and two major stomatotypes could be distinguished, differing in the prevalence and relative abundance of two main bacterial genera: *Streptococcus* and *Veillonella*. We hypothesize that the two stomatotypes could represent the microbial succession from periodontal health to starting dysbiosis. The old and young periodontally healthy subjects do not cluster separately but a trend of the OM in older subjects to periodontitis is visible. Several bacterial genera were identified to be typically more abundant in older periodontally healthy subjects.

Keywords: oral microbiome, periodontal health, periodontitis, core microbiome, stomatotype, taxonomic composition, aging

OPEN ACCESS

Edited by:
Thuy Do,
University of Leeds, United Kingdom

Reviewed by:
Philip Marsh,
Public Health England, United Kingdom
Tobias Boehm,
Western University of Health Sciences, United States

***Correspondence:**
Lucie Najmanova
lucie.najmanova@biomed.cas.cz

Specialty section:
This article was submitted to Microbiome in Health and Disease, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: *15 November 2020*
Accepted: *02 March 2021*
Published: *22 March 2021*

Citation:
Lenartova M, Tesinska B, Janatova T, Hrebicek O, Mysak J, Janata J and Najmanova L (2021) The Oral Microbiome in Periodontal Health. Front. Cell. Infect. Microbiol. 11:629723. doi: 10.3389/fcimb.2021.629723

INTRODUCTION

Periodontitis is the sixth most common disease worldwide (Frencken et al., 2017). The major forms of periodontal disease are gingivitis, chronic periodontitis (which can be the result of untreated gingivitis) and according to former classification also aggressive periodontitis, which differs from the chronic variant by faster and more extensive disease progression, lower age of patients, and obvious familial aggregation (Armitage and Cullinan, 2010; Van der Velden, 2017). Periodontal disease poses a set of inflammatory conditions affecting the tissues surrounding the teeth. It spreads from the gingiva into the deeper, supportive components of the periodontium: the gum, connective tissue, and the alveolar bone surrounding and supporting a tooth (Hernández et al., 2011), and in more severe cases it can lead to a tooth loss (Kirst et al., 2015). It is a complex infectious disease, where specific pathogenic bacteria growing in biofilms play a key role. It is thus the result of the interplay between subgingival biofilm and host immune response and is further affected by other local, environmental, and genetic factors (Griffen et al., 2012).

The oral cavity has, after the gut, the second largest and diverse microbiota harboring over 700 species of bacteria (Deo and Deshmukh, 2019). In health, the oral microbiome (OM) represents a well-balanced dynamic ecosystem that generally tends to keep within its typical values (Najmanova et al., 2021). The dysbiosis then leads to gingivitis and finally periodontitis (Hajishengallis, 2015). The microbial composition shift precedes the clinical signs of the disease (Liu et al., 2012). The relationship between specific groups of taxa and periodontitis has been thoroughly studied: The OM undoubtedly associated with severe periodontitis is characterized by the presence of the so-called "red complex" bacteria: *Porphyromonas gingivalis*, *Tannerella forsythia*, and *Treponema denticola* (Socransky et al., 1998). Other bacteria highly abundant in periodontal disease belong to the phyla Synergistetes, Firmicutes, Bacteroidetes, Chloroflexi (Griffen et al., 2012; Abusleme et al., 2013; Pérez-Chaparro et al., 2014; Kirst et al., 2015). On the other hand, only a few taxa have been unambiguously associated with periodontal health - mainly some members of genera *Actinomyces* and *Streptococcus* (Griffen et al., 2012; Abusleme et al., 2013; Kirst et al., 2015; Meuric et al., 2017). In addition, little is known about the succession of steps leading from periodontal health to disease. For many taxa, the unambiguous assignment to periodontal health or periodontitis has proven difficult because they often exhibited equal prevalence and relative abundance in both states of health. These include the highly abundant *Fusobacterium nucleatum*, *Veillonella parvula* and some members of *Streptococcus* sp., but also the less frequent *Lautropia mirabilis*, *Campylobacter gracilis*, or *Granulicatella adjacens*, (Abusleme et al., 2013; Tsai et al., 2018). The reason could be that in the studies the "periodontal health" is often characterized rather as opposite to periodontitis, i.e. the absence of clinical signs of the disease. Mainly the older cohort of healthy controls, however, could already experience the dysbiosis even though yet without clinical symptoms. Another substantial problem is posed by inconsistent criteria for the diagnosis of patients (variable depth and a number of periodontal pockets required for assignment to periodontitis group, individual experience of examining periodontologist) (Pérez-Chaparro et al., 2014; Deo and Deshmukh, 2019). Previously we tried to solve the problem by assignment of the taxa to the health state using a selection of the most diseased patients from a wider cohort, and, as the opposite a selection of healthy individuals with taxonomically most distant OM (Najmanova et al., 2021). A panel of thirty oral taxa undoubtedly associated with periodontitis corresponded well to previously published data but the list of taxa unambiguously assigned to periodontal health was surprisingly poor containing only seventeen species or so-called "combined taxa" (groups of taxa that could not be distinguished from each other by the used sequencing method). The assignment of dozens of other species frequently identified in the oral cavity thus remains questionable.

In this work, we aimed to describe a typical periodontally healthy oral microbiome and to extend the panel of oral taxa unambiguously associated with periodontal health. It is generally known, that changes in the OM composition naturally occur with aging (Feres et al., 2016; Belibasakis, 2018) which is likely related to the fact, that chronic periodontitis manifests mainly in older people (Eke et al., 2016; Feres et al., 2016; López et al., 2017). To avoid a possible age bias, we employed in our study a cohort of 91 periodontally healthy students of dentistry, i.e. young subjects (average age 23 years) having a very high standard of oral hygiene and thus with lower risk of dysbiosis (HY; healthy young). To verify the impact of age on the taxonomic OM composition, we also analyzed a group of 17 samples from periodontally healthy subjects older than 40 years (HO; healthy old), and for comparison, we also included a group of 45 patients with severe (former aggressive) periodontitis (AP). Two distinct health-associated microbial communities were identified and the dysbiotic changes that could lead to periodontitis onset were described.

MATERIALS AND METHODS

Characteristics of Human Subjects and Sample Collection

Samples from 153 subjects were included and analyzed in this study (**Supplementary Table 1**): 91 periodontally healthy students of dentistry from the 1st faculty of medicine, Charles University in Prague (average age 23 years, marked HY), 17 periodontally healthy people older than 40 years (average age 46 years, marked HO), and 45 patients with severe (former aggressive) periodontitis (average age 33, marked AP). All subjects live in Czech Republic, but besides being students of the same university in HY group, they have no other general relation among each other in terms of living area, employment or any similar parameter. All subjects were examined by a single experienced periodontologist. To be included in the HY or HO group the subjects were required to have no periodontal pocket on probing depth >3 mm, in AP group the subjects had at least two periodontal pockets on probing depth >5mm. The subjects

had not received any antibiotic treatment or periodontal therapy in the three months before the beginning of the study. The HO and AP subjects were obtained within the study approved by the Ethics Committee of the First Faculty of Medicine of Charles University and General University Hospital in Prague as a part of project No. 17-30753A of the Czech Health Research Council and besides periodontitis in the AP group, they were of good general health, the sampling of HY subjects was approved within a project No. 486417 from the Grant Agency of Charles University. All subjects involved in the study signed the informed written consent.

The healthy subjects were sampled from the vestibular side of sulcus gingivalis, the samples from patients with severe periodontitis (AP) were taken from the deepest periodontal pocket. The healthy subjects fulfilled the criteria of periodontal health, as defined by Caton et al., (Caton et al., 2018), i.e. no positive bleeding on probing index (BOP), and no signs of inflammation (erythema and edema), the PPD (periodontal pocket depth) was < 2 mm, as well as the CAL (clinical attachment loss) index. The probands included in the AP group have never been treated for periodontitis prior to inclusion in our study and according to their clinical examination, their severe periodontitis diagnosis was confirmed. The clinical examination contained PPD, BOP, CAL (**Supplementary Table 1**), evaluation of the plaque and tartar amount (low for all probands), number of teeth after preservation treatment (low number of teeth with dental filling for all probands), rtg examination (prevailing vertical character of bone resorption) and family anamnesis with predominated preterm teeth loss in parents of our probands before the age of forty.

Additional information on the tooth sampled (identification of sampled tooth, CAL, BOP, PPD) and other relevant conditions including health state, smoking status, pregnancy, nationality or specific diet are listed in the **Supplementary Table 1**. The descriptive statistics on demographics and clinical data related to tested subjects is given at list "descriptive statistics" of this **Supplementary Table 1**. Each sample was obtained using two sterile paper points (BECHT, Germany). The paper points were left in the gingival sulcus or periodontal pockets for 10s to soak the fluid. Both paper points from one sampling were stored together in -20°C prior to further processing.

DNA Isolation, 16S rDNA Gene Library Preparation, and Sequencing

The DNA was extracted using DNeasy Blood&Tissue kit (Qiagen, Germany) according to the modified manufacturer's instructions (decreased final elution volume of AE buffer from 200 to 120 µl) and stored in -20°C. Isolated DNA was used as a template for PCR 16S rDNA amplification. The universal primers 530f (GTGCCAGCMGCNGCGG) (Dowd et al., 2008) and 907R (CCGTCAATTCMTTTGAGTTT) (Lane et al., 1985) were used to amplify the V4-V5 region of bacterial 16S rDNA in primary PCR. Primers for secondary PCR amplification contained additionally five to seven nucleotide long sample tags, separated from primers by two nucleotide long spacers (**Supplementary Table 2**). PCR reactions were performed in two steps according to Baldrian et al. (Baldrian et al., 2012). First PCR amplification was performed in 3 independent reactions for each sample in 12,5 µl Plain Combi PP Master Mix (Top-Bio, Czech Republic) containing 0,4% Phusion polymerase (New England Biolabs, USA) with 2 µl of template, 2 µl of each primer (0.25 mM) and 6,5 µl H_2O. Cycling conditions were 94°C for 5 min; 35 cycles of 94°C for 1 min, 58°C for 50 s, 72°C for 30 s, followed by 72°C for 10 min. Pooled PCR products were purified after the electrophoretic separation from the agarose gel using the Wizard SV Gel and PCR Clean-Up System (Promega, USA). 3 µl of isolated DNA were used as a template in the second PCR in reaction containing 2 µl of each tagged primer and 25 µl of Plain Combi PP Master Mix enriched by Phusion DNA polymerase and 18 µl of H_2O. Cycling conditions were the same except that cycle number was 10 and number of independent reactions per sample was two. PCR products were separated by electrophoresis and purified using the Wizard SV Gel and PCR Clean-Up System and then concentrated into the volume of 16 µl using the MinElute PCR Purification Kit (QIAGEN, Germany). The concentration of DNA was measured by Qubit 2.0 Fluorometer using dsDNA BR Assay Kit (both Thermo Fisher Scientific, USA). The purified solutions of tagged amplicons from different samples were mixed in equimolar concentrations, the amplicon library was constructed using TruSeq DNA Library Preparation Kit v2 (Illumina, USA) and sequenced by Illumina MiSeq platform (paired-end reads, 2×250 bp).

Analysis of the Sequencing Data

The amplicon raw sequencing data were processed using the pipeline SEED 2.0.3 (Větrovský et al., 2018). The pair-end reads were joined using fastq-join program as a part of ea-utils package (Aronesty, 2013). All sequences were trimmed to 365 nt starting from the first nucleotide after the end of the forward primer sequence. The trimmed sequences were further clustered to 98.5% sequence identity by USEARCH (Edgar, 2010) implemented in SEED and the chimeric sequences were removed (chimera check is a part of the clustering method using Uparse algorithm). The quality-filtration to the mean quality Phred score treshold 30 was applied. Consensus sequences were constructed for each cluster and then compared to HOMD database (Chen et al., 2010) using the Blastn tool. Because not the whole length, but only a 365 nt long portion of the 16S rDNA sequence was analyzed, more human oral taxa (HMTs) often cluster to each consensus at the 98.5% level of identity resulting in the ambiguous taxonomic assignment. To overcome this problem, we defined the "combined taxa" (CTs) in cases when the precise assignment to a single HMT was not possible. The CTs were defined as follows: 889 sequences from HOMD version 14.51 were trimmed to 365 nt equally to our testing sequences and clustered at 98.5% identity resulting in 293 single HMTs and 106 clusters containing from 2 to 15 HMTs. These clusters were named CT1 – CT106 (**Supplementary Table 3**). The identified numbers

of reads per each taxon were further normalized to taxon relative abundance value with respect to a 16S rDNA copy number per genome. This procedure was demonstrated to result in a more realistic estimate of the relative abundances of the bacterial taxa, as it takes into account the variation of 16S rDNA copy numbers among taxa (ranging from 1 to 15). The bacterial genome count estimates were calculated based on the 16S rDNA copy numbers in the closest available sequenced genome as described previously by Vetrovsky and Baldrian (Větrovský and Baldrian, 2013). For the purposes of this publication, the published table of the 16S rDNA copy numbers was extended using rrnDB database (Stoddard et al., 2015) (**Supplementary Table 4**). The sequencing statistics including diversity indices and rarefaction curves is summarized in **Supplementary Table 7**). The raw sequences are deposited in the NCBI Short Read Archive (BioProject accession no. PRJNA670573).

Bioinformatic Analysis

The Weighted Jaccard Similarity clustering analysis (**Figure 1**) was performed using the Python programming language (Rossum and Drake, 2010), specifically the NetworkX library (Hagberg et al., 2008). The network graph was generated based on **Supplementary Table 5**, which displays the relative per-sample abundances of all taxa, with abundances below 0.05% rounded down to zero as insignificant. To reduce the dimensionality of the data, a similarity matrix was calculated between all samples using an abundance-weighted Jaccard similarity index. This highlighted the similarities in the overall composition of the biofilm, rather than taxonomic diversity. In other words, the impact of low-abundance taxa on the similarity metric was reduced. For visualization, the aforementioned NetworkX library was used to create a network graph. The generated graph is unweighted, meaning a shorter distance

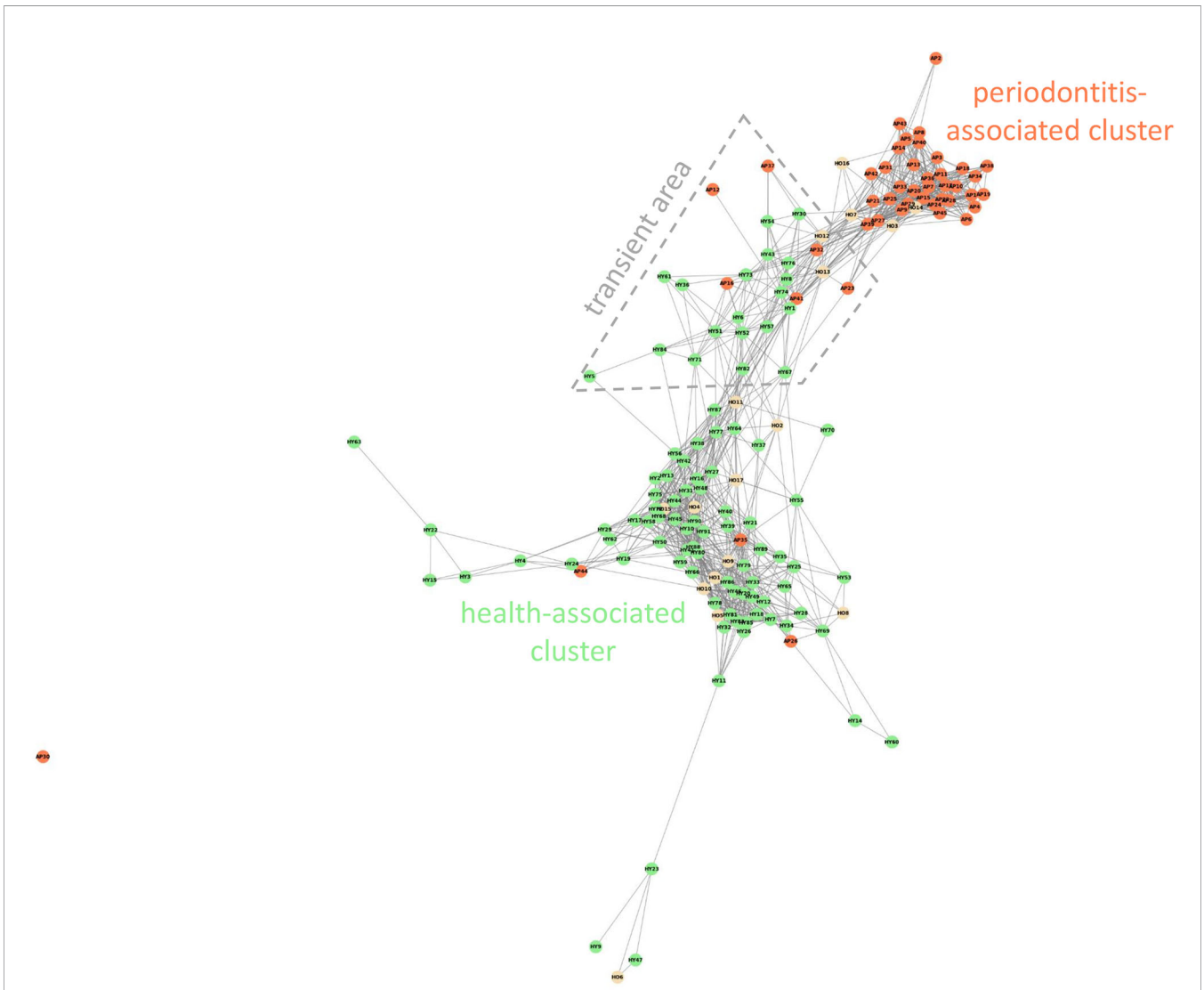


FIGURE 1 | Weighted Jaccard Similarity clustering analysis of all OM samples. Weighted Jaccard similarity index 0.3. The red-colored spots correspond to patients with AP, HY samples are green and HO samples are yellow. The transient area is delimited with the gray dashed line.

between nodes does not always reflect greater similarity. Links were displayed only between nodes with a similarity index of 0.3 or greater.

The PCA (**Figure 3**) and multivariate clustering analysis (**Figure 4**) were performed in the free statistical software PAST 3.25 (Hammer et al., 2001) using the processed sequencing data summarized in **Supplementary Table 5**. The paired group (UPGMA) algorithm and the Bray-Curtis similarity index were employed to obtain **Figure 4**. The statistical significance of identified group characteristics was further analyzed using one-way PERMANOVA test of CLR (central log ratio) transformed data in PAST 3.25. A SIMPER (Similarity Percentage) test within PAST 3.25 software was used for assessing which taxa are primarily responsible for observed differences between groups of samples and One-way ANOVA with Dunn´s *post hoc* test was employed to assess the statistical significance of individual taxa relative abundance difference among groups identified by multivariate clustering analysis (**Supplementary Table 6**).

RESULTS

The sequencing statistics is summarized in the **Supplementary Table 7**. In average we obtained 11334 reads/sample for HY group, 16709 for HO group, and 11974 for AP group with minimum 1006 reads/sample. The rarefaction curves were calculated for a random selection of 1000 sequences to reflect the minimum sample size, set up to cover the expected relevant taxonomic diversity of oral microbiome samples (the sequencing was repeated when the number of reeds per sample was < 1000 to ensure identification of all taxa exceeding average relative abundance 0.5%). The diversity, evenness, and species richness parameters were estimated for each sample (**Supplementary Table 7**) and were compared also between HY HO and AP groups using one-way ANOVA with Dunn´s *post hoc* test. No statistically significant difference was identified for any parameter except for the Chao-1 richness, statistically significantly higher in HY group when compared to AP ($p = 0.002$); for HY/HO and HO/HP p was > 0.05.

Healthy and Diseased Samples Cluster Mostly Separately but Few of Them Are Always Misclassified

All 153 samples were clustered using Weighted Jaccard Similarity analysis, index value 0.3 (**Figure 1**). For each sample, all taxa with relative abundance value >0.05% were included in the analysis. Two major distinct clusters were observed: the periodontitis-associated cluster containing mainly samples of AP patients (red spots in **Figure 1**) and the health-associated cluster covering mainly subjects without clinical signs of the disease (green and yellow spots in **Figure 1**) with several samples from all three cohorts localized in the area in between. The periodontitis-associated cluster contains 39 samples (35 AP and 4 HO), the health-associated cluster includes 86 samples (72 HY, 11 HO, and 3 AP). Twenty-seven samples are localized in the area connecting both the above-mentioned clusters (transient area, dashed line in **Figure 1**) and one sample (AP30) was an outlier, not included in further analysis. The transient area contains 6 AP and 21 H samples (19 HY + 2 HO). The one-way PERMANOVA test confirmed the difference among the healthy, periodontitis and transient groups with p=0.0001 for all mutual comparisons.

In the health-associated cluster, the most abundant and prevalent taxa are CT2 Streptococcus mitis/S. oralis, CT43 Streptococcus gordonii/S. sanguinis, CT6 Veillonella dispar/V. parvula, CT25 Neisseria flava/N. mucosa, CT27 Neisseria subflava, HMT14 Neisseria oralis, HMT718 Haemophilus parainfluenzae, CT23 Haemophilus haemolyticus, CT24 Haemophilus sputorum, CT10 Prevotella histicola, CT37 Gemella morbillorum, CT48 Rothia dentocariosa, HMT22 Lautropia mirabilis, HMT37 Stenotrophomonas nitritireducens, and CT13 Aggregatibacter aphrophilus (**Table 1**). The relative abundance and in most cases also the prevalence of these taxa remarkably decreases toward the transient area and periodontitis-associated cluster. On the other hand, the relative abundance and prevalence of taxa typical for periodontitis do not reach high values in periodontal health and transient area, but the increasing trend from the health, through the transient area to the periodontitis-associated cluster is noticeable. Several taxa exhibit the highest abundance and prevalence in subjects from the transient area, namely CT3 *Fusobacterium nucleatum*, CT8 *Porphyromonas pasteri/P. catoniae*, HMT775 *Capnocytophaga sputigena*, HMT329 *Capnocytophaga leadbetteri* and CT51 *Capnocytophaga granulosa*, HMT311 *Prevotella oris*, CT53 *Tannerella* sp. and HMT623 *Campylobacter gracilis*.

Even though no obvious age-dependent clustering pattern was observed, still 35% of samples from HO group (6 out of 17), clustered together with AP or in close proximity (yellow spots in **Figure 1**), but no HY sample clustered together with AP. We compared the OM taxonomic composition of these 6 HO samples (HO3, 7, 12-14, and 16; highlighted in **Supplementary Table 5**) with a group of remaining 102 samples from healthy individuals (HY and HO) and also a group of 45 AP samples. The OM of these 6 HO samples resembles the diseased one, just the relative abundance and prevalence values of main periodontitis-associated taxa (red-complex taxa, *Fretibacteria* CT12, and HMT 363, and *Filifactor alocis* HMT 539) are slightly lower when compared to the AP group. On the other hand, the average relative abundance of *Fusobacterium nucleatum* is higher in these six samples. The taxa typical for periodontal health (CT2 *Streptococcus mitis/S. oralis*, CT6 *Veillonella dispar/V. parvula*, CT25 *Neisseria flava/N. mucosa*) are almost absent exhibiting the average relative abundance values comparable to the AP group. Even though the 6 HO individuals did not exhibit any clinical signs of the disease, their OM taxonomic composition indicates a high risk of periodontitis development in the future.

The OM Taxonomical Composition in Aging

The relative abundance values of selected taxa in relation to the age and state of health are compared in **Figure 2** and summarized including the prevalence data in **Table 2**. Three groups of samples were compared: HY (average age 23), HO

TABLE 1 | The average relative abundance and representation of selected taxa in the health-associated cluster, transient area, and periodontitis-associated cluster.

HMT/CT	*health-associated cluster*		*transient area*		*periodontitis-associated cluster*	
	average abundance [%]	prevalence [%]	average abundance [%]	prevalence [%]	average abundance [%]	prevalence [%]
CT2 *Streptococcus mitis; S. oralis*	20.87	100.00	4.37	96.30	0.52	48.72
CT6 *Veillonella dispar; V. parvula*	15.19	100.00	5.15	100.00	0.51	53.85
CT25 *Neisseria flava; N. mucosa*	6.54	88.37	2.37	74.07	0.89	28.21
CT43 *Streptococcus gordonii; S. sanguinis*	4.97	100.00	0.65	74.07	0.12	33.33
HMT718 *Haemophilus parainfluenzae*	4.83	96.51	0.60	70.37	0.05	20.51
CT10 *Prevotella histicola*	2.48	84.88	0.71	77.78	0.17	38.46
CT27 *Neisseria subflava*	2.45	87.21	0.61	77.78	0.24	38.46
HMT14 *Neisseria oralis*	2.43	53.49	1.50	62.96	0.12	23.08
CT37 *Gemella morbillorum*	2.18	89.53	1.09	85.19	0.10	33.33
CT48 *Rothia dentocariosa*	1.95	74.42	0.31	37.04	0.17	17.95
HMT22 *Lautropia mirabilis*	1.93	80.23	0.40	62.96	0.09	25.64
CT23 *Haemophilus haemolyticus*	1.44	80.23	0.27	62.96	0.02	5.13
HMT37 *Stenotrophomonas nitritireducens*	1.13	69.77	0.05	22.22	0.09	10.26
CT24 *Haemophilus sputorum*	1.07	66.28	0.17	44.44	0.00	5.13
CT13 *Aggregatibacter aphrophilus*	1.00	56.98	0.43	59.26	0.14	17.95
CT3 *Fusobacterium nucleatum*	4.76	97.67	31.30	100.00	23.32	100.00
CT8 *Porphyromonas pasteri; P. catoniae*	2.15	82.56	13.05	92.59	0.37	41.03
HMT775 *Capnocytophaga sputigena*	1.37	63.95	2.63	77.78	0.09	23.08
HMT329 *Capnocytophaga leadbetteri*	0.56	44.19	1.91	62.96	0.12	25.64
HMT311 *Prevotella oris*	0.26	34.88	1.20	55.56	0.68	20.51
CT51 *Capnocytophaga granulosa*	0.97	34.88	1.15	66.67	0.32	43.59
CT53 *Tannerella* sp.	0.23	32.56	1.06	59.26	0.23	33.33
HMT623 *Campylobacter gracilis*	0.38	43.02	1.02	88.89	0.35	64.10
HMT619 *Porphyromonas gingivalis*	0.02	6.98	0.13	14.81	12.15	74.36
HMT613 *Tannerella forsythia*	0.01	5.81	0.21	25.93	7.29	100.00
CT12 *Fretibacterium* sp.	0.00	2.33	0.13	37.04	6.55	97.44
CT50 *Treponema denticola*	0.02	8.14	0.30	25.93	5.91	94.87
CT22 *Porphyromonas endodontalis*	0.20	13.95	1.34	37.04	4.17	94.87
CT7 *Treponema vincentii*	0.02	9.30	0.78	44.44	3.59	92.31
CT11 *Treponema socranskii*	0.04	13.95	0.90	48.15	3.25	100.00
HMT643 *Prevotella intermedia*	0.20	5.81	2.04	25.93	2.70	74.36
HMT274 *Bacteroidales [G-2]* sp.	0.13	11.63	0.29	37.04	2.38	69.23
HMT539 *Filifactor alocis*	0.05	6.98	0.03	18.52	1.57	89.74
CT56 *Campylobacter rectus*	0.17	36.05	0.99	70.37	1.49	87.18
HMT363 *Fretibacterium fastidiosum*	0.01	2.33	0.04	14.81	1.46	97.44
CT42 *Treponema maltophilum*	0.00	3.49	0.05	18.52	1.17	94.87

The oral taxa predominantly abundant and prevalent in the health-associated cluster are marked green. Taxa with the highest abundance and representation in the transient area are gray and the taxa with the highest average relative abundance and prevalence in the periodontitis-associated cluster are red. The table includes only the taxa with minimal average relative abundance >1% and minimal prevalence >50% in at least one cluster. CT stands for combined taxon (See ***Supplementary Table 3****).*

(average age 46), and patients with AP (average age 33). The one-way PERMANOVA test revealed clear difference between groups. When Bray-Curtis was used as a similarity index, the HY and HO were found close to each other (0.0059) but clearly different from AP (0.94 and 0.86, resp.). Twelve taxa were identified to be the most abundant in HY: CT6 *Veillonella dispar*/*V. parvula*, CT8 *Porphyromonas pasteri*; *P. catoniae*, CT25 *Neisseria flava*/*N. mucosa*, HMT718 *Haemophilus parainfluenzae*, CT43 *Streptococcus gordonii*/*S. sanguinis*, CT10 *Prevotella histicola*, CT27 *Neisseria subflava*, CT37 *Gemella morbillorum*, HMT14 *Neisseria oralis*, HMT22 *Lautropia mirabilis*, CT48 *Rothia dentocariosa*, and CT23 *Haemophilus haemolyticus*. A nicely visible gradual trend of diminution can be seen in the abundance and in most cases also in prevalence of these taxa toward to group of HO and patients with AP. On the other hand, in the AP group the highest abundance of periodontitis-associated taxa CT3 *Fusobacterium nucleatum*, CT7 *Treponema vincentii*, CT11 *Treponema socranskii*, CT50 *Treponema denticola*, CT12 *Fretibacterium* sp., CT22 *Porphyromonas endodontalis*, HMT619 *Porphyromonas gingivalis*, CT56 *Campylobacter rectus*, HMT274 *Bacteroidales [G-2]* sp., HMT363 *Fretibacterium fastidiosum*, HMT539 *Filifactor alocis*, HMT613 *Tannerella forsythia*, and HMT643 *Prevotella intermedia* was identified with an opposite trend of subsequent decreasing of the abundance and prevalence of these taxa toward to HO and then to HY subjects. Only a few bacterial taxa were found to prevail in the group of HO samples: CT2 *Streptococcus mitis*/*S. oralis*, CT51 *Capnocytophaga granulosa*, HMT329 *Capnocytophaga leadbetteri*, HMT775 *Capnocytophaga sputigena*, and HMT322 *Bergeyella* sp.

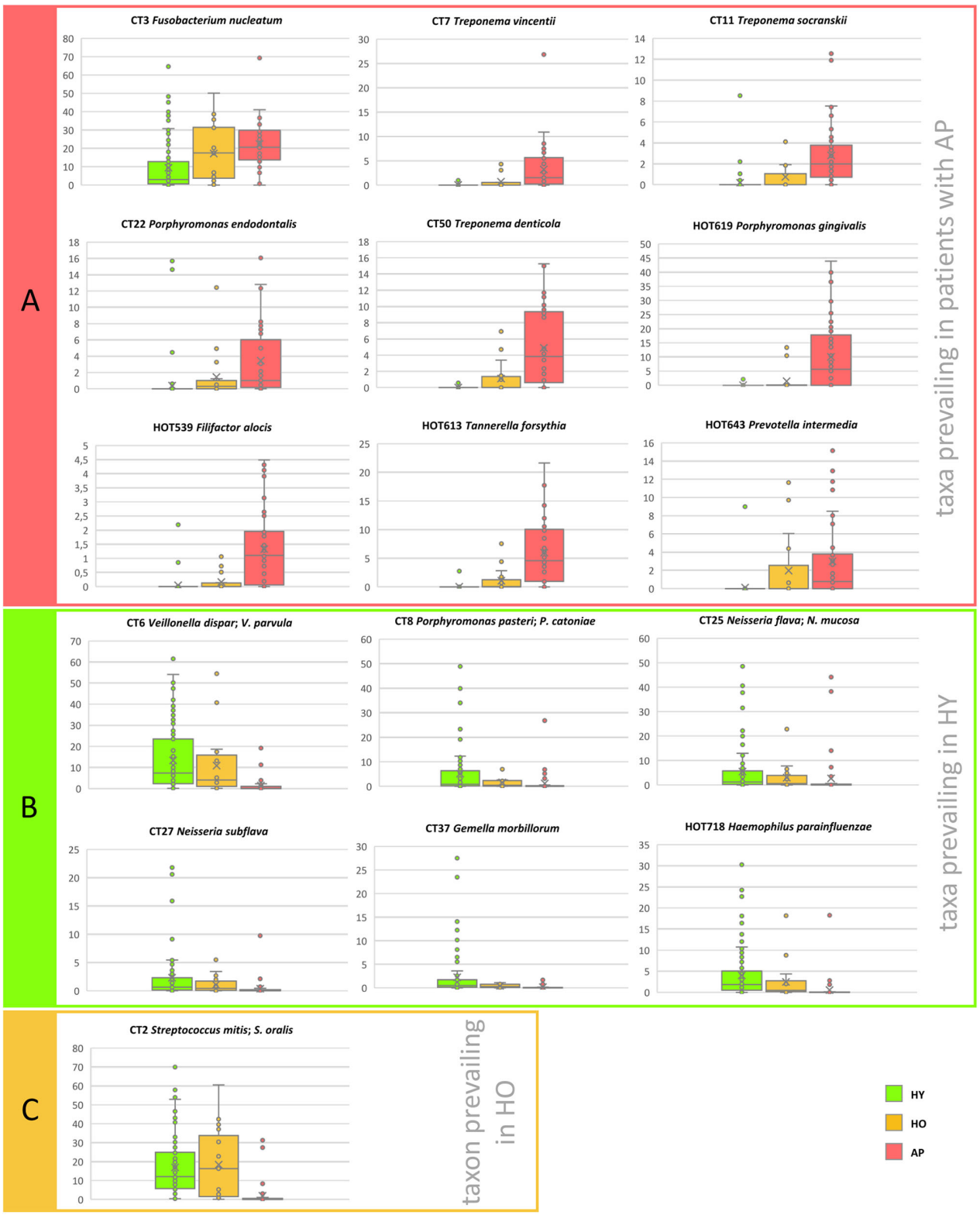


FIGURE 2 | The distribution of the relative abundances of selected taxa in relation to the age of probands and their state of health. **(A)** Taxa prevailing in periodontitis, **(B)** Taxa prevailing in HY, **(C)** Taxon prevailing in HO. The area of box plots with oral taxa dominant in patients with AP is highlighted in red, HY in green and one boxplot related to taxon most abundant in the HO group is highlighted in yellow. CT stands for combined taxon (See **Supplementary Table 3**).

TABLE 2 | The comparison of the relative abundance and prevalence of significant oral taxa according to the age (HY vs HO) and state of health (HY, HO and AP).

HMT/CT	HY		HO		AP	
	average abundance [%]	prevalence [%]	average abundance [%]	prevalence [%]	average abundance [%]	prevalence [%]
CT6 *Veillonella dispar; V. parvula*	13.37	100.00	10.87	100.00	1.46	60.00
CT8 *Porphyromonas pasteri; P. catoniae*	5.19	85.71	1.59	88.24	1.19	44.44
CT25 *Neisseria flava; N. mucosa*	5.44	83.52	3.03	94.12	2.66	35.56
HMT718 *Haemophilus parainfluenzae*	4.02	92.31	2.47	88.24	0.56	26.67
CT43 *Streptococcus gordonii; S. sanguinis*	3.98	96.70	2.35	82.35	1.27	40.00
CT10 *Prevotella histicola*	2.37	84.62	0.77	82.35	0.25	42.22
CT27 *Neisseria subflava*	2.31	86.81	1.09	76.47	0.38	44.44
CT37 *Gemella morbillorum*	2.27	89.01	0.43	94.12	0.16	37.78
HMT14 *Neisseria oralis*	2.17	54.95	0.77	64.71	1.09	26.67
HMT22 *Lautropia mirabilis*	1.82	82.42	0.23	47.06	0.25	31.11
CT48 *Rothia dentocariosa*	1.64	68.13	1.05	52.94	0.43	24.44
CT23 *Haemophilus haemolyticus*	1.37	81.32	0.38	64.71	0.01	6.67
CT2 *Streptococcus mitis; S. oralis*	16.94	100.00	18.38	100.00	1.95	53.33
CT51 *Capnocytophaga granulosa*	1.04	40.66	1.24	58.82	0.24	20.00
HMT322 *Bergeyella* sp.	0.60	70.33	1.73	70.59	0.04	13.33
HMT329 *Capnocytophaga leadbetteri*	0.68	48.35	1.63	58.82	0.32	24.44
HMT775 *Capnocytophaga sputigena*	1.74	64.84	1.39	82.35	0.23	26.67
CT3 *Fusobacterium nucleatum*	9.48	98.90	17.23	100.00	22.44	97.78
CT7 *Treponema vincentii*	0.05	12.09	0.71	35.29	3.25	88.89
CT11 *Treponema socranskii*	0.17	17.58	0.75	47.06	2.80	88.89
CT12 *Fretibacterium* sp.	0.01	5.49	1.03	35.29	5.36	88.89
CT22 *Porphyromonas endodontalis*	0.40	9.89	1.45	58.82	3.46	91.11
CT50 *Treponema denticola*	0.02	4.40	1.11	47.06	4.90	86.67
CT56 *Campylobacter rectus*	0.38	40.66	0.63	64.71	1.21	82.22
HMT274 *Bacteroidales [G-2]* sp.	0.16	13.19	0.98	47.06	1.79	60.00
HMT363 *Fretibacterium fastidiosum*	0.01	3.30	0.38	35.29	1.14	77.78
HMT539 *Filifactor alocis*	0.05	5.49	0.16	41.18	1.32	75.56
HMT613 *Tannerella forsythia*	0.04	3.30	1.07	52.94	5.99	88.89
HMT619 *Porphyromonas gingivalis*	0.03	3.30	1.44	29.41	10.04	68.89
HMT643 *Prevotella intermedia*	0.10	2.20	1.97	47.06	3.00	71.11

The taxa prevailing in each group are highlighted in respective color: green for HY, peach for HO, and red for AP. The table includes only taxa with minimal average relative abundance >1% and minimal prevalence >50% in at least one group. CT stands for combined taxon (See ***Supplementary Table 3****).*

The Stomatotypes in Oral Health and Disease

All 153 samples were analyzed using principal component analysis (PCA; **Figure 3**) and hierarchical clustering analysis based on the Bray-Curtis similarity indexes (**Figure 4**). All identified taxa in each sample were taken into account in the calculation of sample distance. In addition, in this case, the majority of AP samples form a compact cluster (red triangles in **Figure 3**) apart from a much more diffuse group of health-associated spots (green dots and yellow squares in **Figure 3**). The inner panel in **Figure 3** shows the contribution of individual taxa to the distribution of samples.

The clustering analysis based on Bray-Curtis similarity indexes revealed two superclusters (**Figure 4**), one associated mainly with periodontal health and the second with periodontitis. The health-associated supercluster comprises two main clusters of equal size (Cluster 1 and Cluster 2) and two small clusters (Cluster 3a and 3b) corresponding to outliers from **Figure 1**. The periodontitis-associated supercluster comprises also two main clusters: Cluster 4 corresponding to the transient state, and Cluster 5 corresponding to periodontitis (compare to **Figure 1**). Using one-way PERMANOVA test, the Clusters 3a and 3b were not found to differ significantly from each other ($p = 0.33$), however, when considering the Clusters 1 and 2 (health-associated), 4 (transient) and 5 (periodontits-associated), they all mutually differ significantly ($p = 0.0001$). The determining taxa for each cluster are listed in **Figure 4** and in more detail in **Supplementary Table 6**. For twenty most discriminating taxa according to Simper test also One-way ANOVA with Dunn´s *post hoc* test was employed to compare the clusters (**Supplementary Table 6**; individual lists; statistically significant results are highlighted in red).

Predictably, the periodontitis associated cluster Cluster 5 is characterized by the high relative abundance of CT3 *F. nucleatum*, CT12 *Fretibacterium* spp., and red complex taxa HMT619 *P. gingivalis*, HMT613 *T. forsythia*, and CT50 *T. denticola*. Also, the transient state-associated cluster Cluster 4 is characterized by the high relative abundance of CT3 *F. nucleatum*, but in this case, accompanied by CT8 *Porphyromonas pasteri/catoniae* and no red-complex taxa. Two

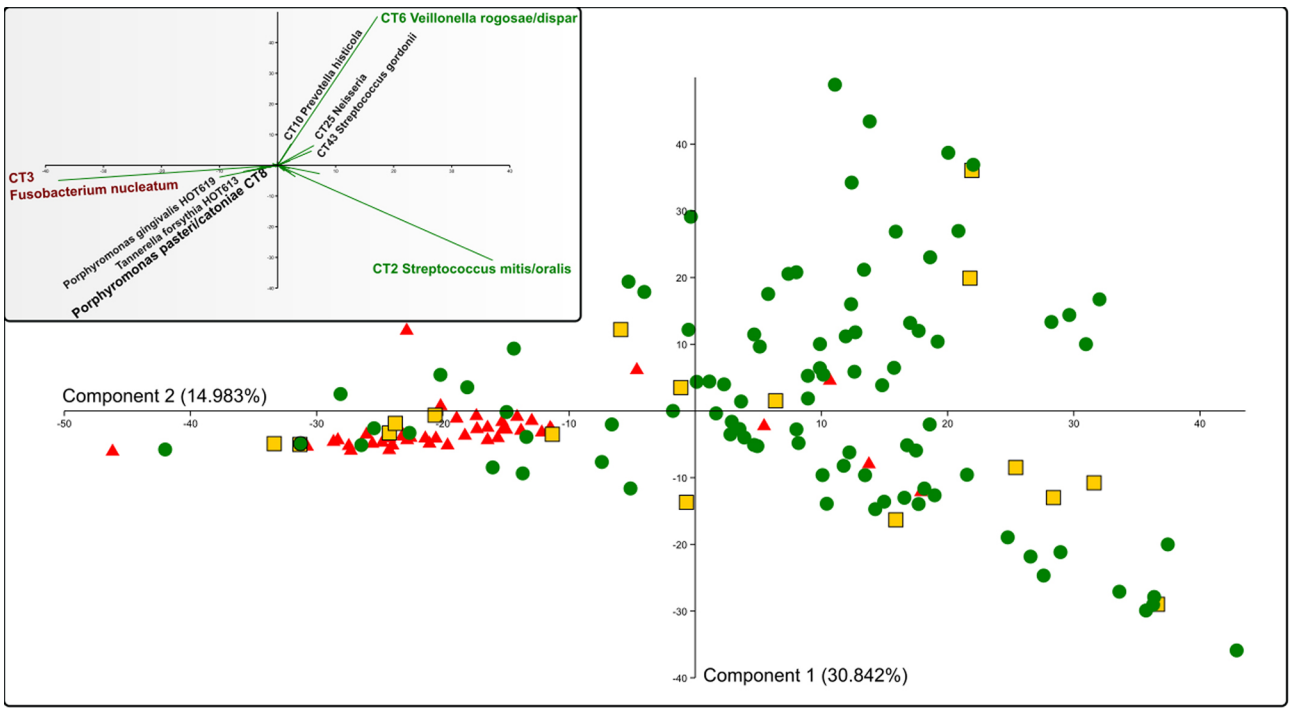


FIGURE 3 | PCA analysis of all OM samples. Red triangles represent AP samples, green spots HY, and yellow squares represent the HO samples. The inner panel shows the contribution of selected oral taxa to the distribution. The determining taxa for each quadrant are highlighted in colors (green for periodontal health and red for periodontitis), CT8 P. pasteri/catoniae determining the transient state is highlighted by a bigger font. CT stands for combined taxon (See **Supplementary Table 3**).

main health-associated clusters are Cluster 1 (characterized by CT2 *Streptococcus mitis/oralis* and *Haemophilus* spp. HMT718, CT23, and CT24) and Cluster 2 (characterized by CT6 *Veillonella rogosae/dispar*, *Neisseria* spp. CT25, CT27, and CT37 *Gemella morbillorum*). This distribution corresponds well also with the main taxa driving the PCA distribution (the inner panel in **Figure 3**). In contrast to the periodontitis-associated cluster, the two health-associated stomatotypes are not conclusively distinguished. They rather form a diffuse group covering all possible combinations of health-associated taxa.

DISCUSSION

Misclassification of Samples

The Weighted Jaccard Similarity analysis (**Figure 1**), as well as the PCA analysis (**Figure 3**), expectedly distinguished a compact cluster of AP samples from a bigger and more diffuse cluster of samples of healthy individuals. A similar trend was observed earlier (Kirst et al., 2015) and it was explained as a result of microbial succession during the onset of periodontal disease. Kirst et al. revealed that shallow sampling sites in patients with chronic periodontitis exhibited the highest species richness and diversity (containing both, the health-associated taxa as well as the taxa typical for periodontitis), while the deep periodontal pockets contained only a limited number of species that were, moreover, quite uniform among the tested individuals. The healthy sites were also less diverse, but the individuals differed more, which could result from an ambiguous diagnosis of some healthy probands, in whom the early stages of dysbiosis can occur without any clinical signs. In our study, in order to characterize a periodontally healthy oral microbiome, we included the HY group, selected on purpose from young people with precise oral hygiene, i.e. with a very low probability of dysbiosis that could otherwise bias the results. To contrast with the former, we also included a group of severe periodontitis (AP) patients who are typically characterized by unambiguous diagnosis, deep periodontal pockets, and a rapid progression of the disease. Nevertheless, even with such a well-distinguished set of individuals, still, several HY samples cluster close to the AP group. From the HO group, in which the dysbiosis preceding the periodontitis development could already be expected, four samples cluster directly within the AP group (**Figure 1**), and few AP samples, on the other hand, cluster together with the healthy ones. The PCA analysis shows an even bigger overlap (**Figure 3**).

A similar discrepancy between the microbial profile and clinical status in some percentage of samples was already shown previously (Kirst et al., 2015; Park et al., 2015; Szafranski et al., 2015). PCA or PCoA analysis frequently revealed a compact cluster of samples from periodontitis patients, a bigger and more diffuse cluster of samples from periodontally healthy individuals, and several outliers or

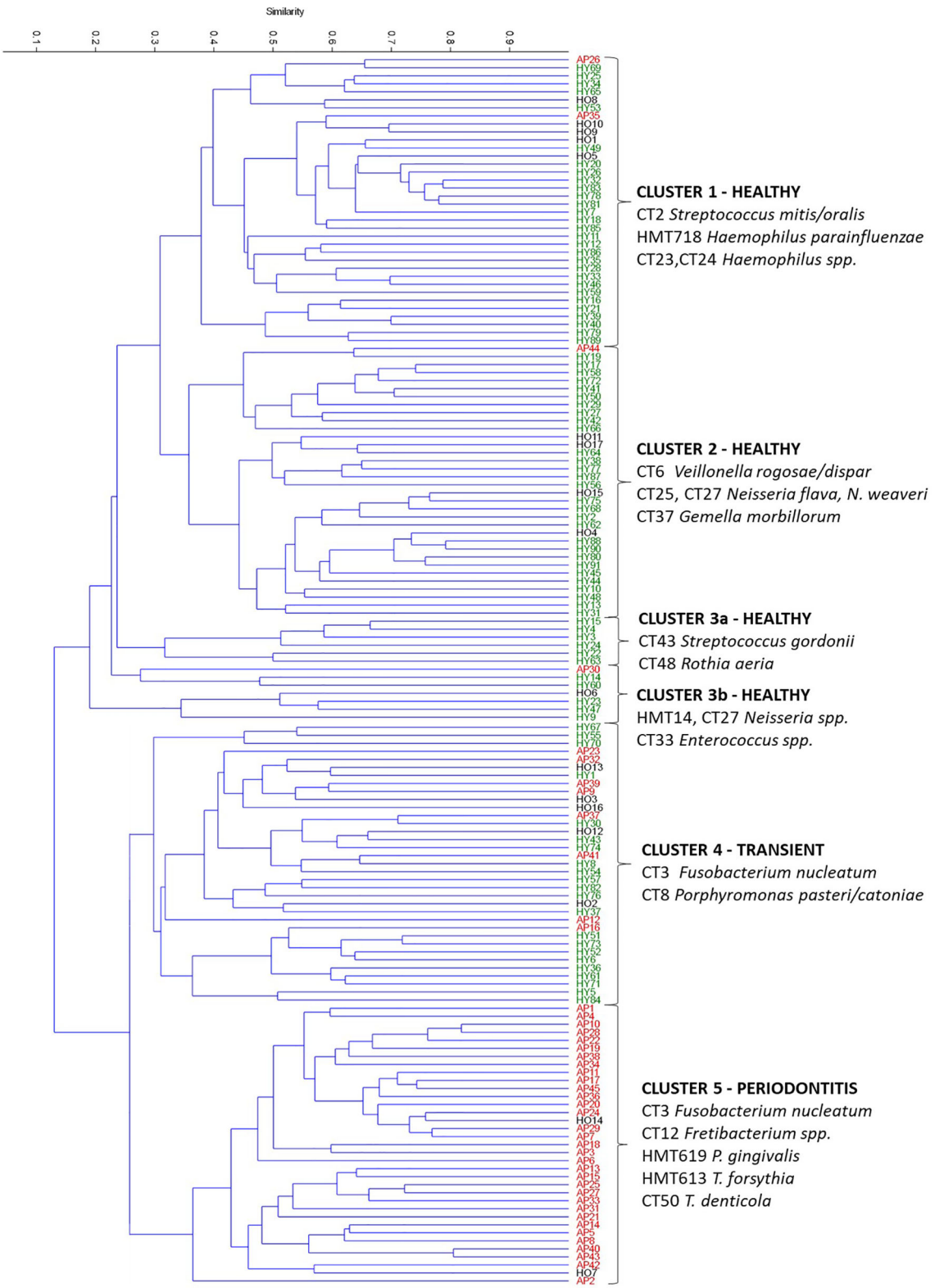


FIGURE 4 | Hierarchical clustering of all OM samples based on the Bray-Curtis similarity indexes. AP samples are marked red, HO black, and HY green. The upper measure indicates the Bray-curtis value. The determining taxa for each cluster are listed. CT stands for combined taxon (See **Supplementary Table 3**).

samples assigned to an improper group. In our study, ~18% of samples did not belong to any of the identified AP or healthy cluster but form a connection between them (**Figure 1**). We suppose that these samples represent a transient state. The clinically healthy subjects (HY and HO) from the transient area and the four HO samples clustering with AP would thus probably experience dysbiosis and consequently would be at a higher risk of periodontitis onset, while the AP samples from the transient area could correspond to a milder course of the disease or patients with better prognosis. Nevertheless, we must consider also the possibility of altered functional activity of the OM as discussed for example by Duran-Pinedo (Duran-Pinedo and Frias-Lopez, 2015) and/or an unusual host immune response, more tolerogenic in case of healthy subjects with unhealthy OM and more proinflammatory in AP subjects with transient OM (Sultan et al., 2018).

Four AP samples from our set were misclassified: One outlier (AP30), and three samples localized within the health-associated cluster in **Figure 1** (AP26, AP35, and AP44). The outlier AP30 exhibited very high relative abundance (44%) of CT98 *Propionibacterium propionicum* (**Supplementary Table 5**; list AP). The high relative abundance of *P. propionicum* is by some authors correlated with apical periodontitis and endodontal lesions, however, this finding has not been corroborated by others, and there is no consensus concerning the role of *P. propionicum* in the pathogenesis of the periodontal disease (Dioguardi et al., 2020). The three remaining samples represented typical health-associated OMs and similarly to the AP30 they did not contain any of the above-mentioned "true periopathogens". Some authors explain this phenomenon with other causes of periodontal pocket formation then periodontitis, for example, anatomical abnormalities in labial frena (Monnet-Corti et al., 2018). This, however, is not the case of our patients. Two of them (AP26 and AP35) suffer from a localized form of the disease with 5 affected teeth, and two other (AP30 and AP44) have generalized AP with even 25 and 20 periodontal pockets, respectively. Such an extent of the disease cannot be caused by labial frena abnormalities. The severe periodontitis in these individuals thus probably originated from an unusual interplay between their OM and immune system. It is important to note, that the 16S rDNA sequencing-based taxonomic characterization of the OM provides a valuable, but still incomplete picture of the oral ecosystem. The bacteria are major and very important members of the OM, but fungi and archaea species also can play their role, and additionally, the metabolic activity of individual species can differ in relation to interactions with their surroundings (Sultan et al., 2018). Consequently, more complex diagnostic tools including proteomic or metabolomic studies will be required to reveal the cause of inflammation and periodontitis development in these nonstandard cases.

The most abundant and prevalent taxa in health and periodontitis-associated cluster are consistent with previously published data (Griffen et al., 2012; Abusleme et al., 2013; Pérez-Chaparro et al., 2014; Kirst et al., 2015). Slightly surprising could be very low prevalence of *Aggregatibacter actinomycetemcomitans* in our AP group (the taxon was identified only in three samples; 0.11% in AP24, 3.6% in AP12, and 15.8% in AP37), because for a long time, this taxon has been typically associated with periodontitis, mainly with its severe (aggressive, according to former classification) form (Schacher et al., 2007; Henderson et al., 2010). Nevertheless, Henderson et al., also document, that the proportion of the population that harbors *A. actinomycetemcomitans* varies dramatically between various geographical areas and different clinical presentations of periodontitis. For example within Europe, 23% of Dutch subjects harbored *A. actinomycetemcomitans* compared to only 3% of Spanish subjects, on the other hand in Asia it was detected in 78% of healthy Vietnamese subjects. There is no general study concerning the prevalence of *A. actinomycetemcomitans* in periodontitis patients in the Czech Republic, however, with respect to the published geographical and ethnical variability, the low *A. actinomycetemcomitans* prevalence in our cohort does not put our diagnosis of severe (aggressive) periodontitis in question.

The OM of individuals in the transient area is characterized by the decreased relative abundance of typical health-associated taxa and increased relative abundance of anaerobic or facultative anaerobic taxa (genera *Fusobacterium, Porphyromonas* or *Capnocytophaga*), which are supposed to act as later colonizers, facilitating further colonization by "true periopathogens" of red complex (Socransky et al., 1998), and/or species of periodontitis-associated genera like *Treponema*, *Fretibacterium*, or *Filifactor*. The role of *F. nucleatum* in the subgingival biofilm formation probably lies in the bridging among microorganisms, allowing attachment of periodontitis-specific bacteria (Kolenbrander et al., 2006; Kolenbrander et al., 2010). The presence of *F. nucleatum* on its own does not cause periodontal disease, however, its increased abundance is undoubtedly associated with the disease (He et al., 2012; Yang et al., 2014). The second most abundant bacteria in samples from the transient area is CT8 *Porphyromonas pasteri/catoniae* (13.05%). The genus *Porphyromonas* is quite unique because some *Porphyromonas* species are frequently associated with oral health (De Lillo et al., 2004; Camelo-Castillo et al., 2015; Takeshita et al., 2016; Yasunaga et al., 2017; Rusthen et al., 2019) while another member of the genus, *Porphyromonas gingivalis*, belongs to the red-complex and it is unequivocally disease-associated. Our results show that representatives of CT8 *P. pasteri* and/or *P. catoniae* do not form the core microbiome in oral health but rather indicate the transient state with an increased risk of periodontitis development. The genus *Capnocytophaga* is represented in samples from the transient state by three species: HMT 775 *Capnocytophaga sputigena* (2.63%), HMT329 *Capnocytophaga leadbetteri* (1.91%), and CT51 *Capnocytophaga granulosa* (1.15%). It is not a highly abundant genus but its relative abundance in healthy samples is remarkably lower and in periodontitis, it almost disappears. This finding is in good agreement with Pudakalkatti et al., who described *Capnocytophaga* species to be the most prevalent in gingivitis (a transient state from a clinical point of view) rather than in healthy periodontium and periodontitis (Pudakalkatti et al., 2016). They also claimed that *Capnocytophaga* has the potential to cause periodontal disease, but as it is less competitive in the periodontal pocket, it is usually overgrown by other rapidly growing bacteria. The role of *Capnocytophaga* is also supported by experiments published by Okuda et al., who proved that biofilm

formation by *F. nucleatum* is enhanced by a soluble factor produced by *Capnocytophaga* cells (Okuda et al., 2012). Another taxon exhibiting the highest average relative abundance and prevalence in the transient group is HMT311 *Prevotella oris* (1.20%), a taxon previously proved to co-aggregate with *P. gingivalis* and thus to promote the colonization of the gingiva by *P. gingivalis* in early stage of biofilm formation (Sato and Nakazawa, 2014). Finally, CT53 *Tannerella* sp. and HMT623 *Campylobacter gracilis* were also associated with the transient state. CT53 is comprised of three *Tannerella* species, two of them (HMT808, and HMT916) associated with periodontitis (Griffen et al., 2012; Beall et al., 2018), while HMT286 having a relationship to oral health (Leys et al., 2002). The 16S rDNA region sequenced in this study does not allow us to differentiate these three species, thus preventing the meaningful discussion of their role in the development of periodontitis. The average relative abundance of HMT623 *C. gracilis* in periodontal health and periodontitis is comparable and almost negligible (<0.4%). In transient state, it increased above 1% and also the prevalence was remarkably higher (89% compared to 43% and 64% in both border states). This finding corresponds well with previous association of *C. gracilis* with shallow periodontal pockets rather than the deeper ones (Macuch and Tanner, 2000). As a microaerophilic organism, which requires an environment that contains a reduced concentration of oxygen (Guillermo et al., 1996), *C. gracilis* could be, together with the above mentioned transient state-associated taxa, another supporting indicator of the initiating dysbiosis and increased risk of periodontitis development.

OM Changes in Aging

Belibasakis in his recent review summarized the knowledge concerning the OM composition changes in relation to the aging (Belibasakis, 2018) showing that relatively simple OM in early childhood is enriched by the acquisition of new taxa at an early predentate imprinting period and later during the eruption of primary teeth. During adult life, the OM composition of healthy individuals tends to keep a dynamically balanced state called "microbial homeostasis" comprising both natural and repeated colonization of the oral cavity by novel taxa without a remarkable effect on oral health. However, aging does result in changes to the host immune system, which in turn shifts the tolerance against microbial inhabitants of the oral cavity and which could consequently cause dysbiosis and periodontal disease. Besides the increasing prevalence of *Actinomyces* spp. in samples from older individuals (in spite of the increased prevalence of exposed root surfaces in higher age), no considerable differences in the OM composition were noted with regards to dental caries or periodontitis, between younger and elderly healthy populations (Belibasakis, 2018).

Our analysis, nevertheless, revealed several oral taxa clearly more abundant and/or prevalent in HO group when compared to the HY and opposite, even though the age difference between both groups of healthy individuals is not big (40-53 years with average 46 in HO group *vs.* 19-39 years with average 23 in HY). Generally, these changes could be summarized as an aging-related gradual decrease of relative abundance of health-associated taxa and an increase of taxa associated with the transient state. The most remarkable is the increasing relative abundance of CT3 *F. nucleatum* and *Capnocytophaga* species and decreasing relative abundance of health-associated *Neisseria* species, *Lautropia mirabilis*, *Prevotella histicola* or *Gemella morbillorum* (**Table 2**). Quite specific is a case of CT8 *P. pasteri*/*P. catoniae*, which average relative abundance clearly decreases with growing age (5.19% in HY *vs.* 1.59% in HO), but according to the **Table 1** it is a typical taxon for the transient state (2.15% in health vs. 13.05% in a transient state and only 0.37% in periodontitis). When considering solely the HY group, the average relative abundance of CT8 in 19 transient samples is 15.9% while in 72 remaining samples it is only 2.35%. Similarly, in 6 AP samples in the transient area, the average relative abundance of CT8 is 6.49% while in the remaining 39 AP samples it is only 0.37%. Members of CT8 *P. pasteri*/*P. catoniae* belong to the so-called POTG (*Porphyromonas* other than *gingivalis*) group of microorganisms (Guilloux et al., 2020). Typically, they colonize lungs and lower airways and in some diseases like cystic fibrosis, the relative abundance of *P. catoniae* can serve as a marker to discriminate between various states of health (Cuthbertson et al., 2016). POTG and mainly CT8 taxa were also frequently identified in the oral cavity, but in contrast to *P. gingivalis*, they were associated with periodontal health (Abusleme et al., 2013; Camelo-Castillo et al., 2015). Our data, however, indicate, that the increased relative abundance of CT8 is rather a marker of the transient state directing to periodontitis (see small panel in **Figure 3**).

The increased average relative abundance of periodontitis associated taxa in HO samples corresponds well with frequently reported higher prevalence and severity of periodontal disease among older adults (Eke et al., 2012; Baelum and López, 2013; Eke et al., 2015; Feres et al., 2016; Ebersole et al., 2018). Generally, the aging comes with risk factors like a higher predisposition to other systemic diseases which can indirectly modulate the periodontal condition (Persson, 2018), the excessive immune response of the host to oral microbiota (Ebersole et al., 2016) resulting in aging-related moderate loss of periodontal attachment and alveolar bone (Burt, 1994) or higher incidence of the exposed root surfaces, facilitating overgrowth of opportunistic pathogens. The age was described to be a significant factor driving the OM composition dynamics (Belibasakis, 2018; Deshpande et al., 2018), however, the causality still remains unclear (Feres et al., 2016; LaMonte et al., 2019).

The Stomatotypes in Periodontal Health and Microbial Succession

The OM compositional patterns representing various global optimal equilibria of the microbial community have recently been referred to as "stomatotypes" (Willis et al., 2018; Willis and Gabaldón, 2020). The yet identified stomatotypes from systematically healthy individuals (De Filippis et al., 2014; Takeshita et al., 2016; Zaura et al., 2017; Willis et al., 2018) are summarized in **Table 3** together with stomatotypes identified in this study.

The comparability of the data is slightly limited by inconsistent or even missing subject characterization and examination of periodontal health before sampling.

TABLE 3 | OM stomatotypes in periodontal health.

Authors	Age of subjects	No. of subjects	Country	Sample type	Medical/dental examination	Stomatotype designation	The determining taxa	Method
Current study	**19-53**	**108**	**Czech Republic**	**Dental plaque from the gingival sulcus**	**All probands examined by periodontologist: no periodontal pocket > 3mm**	**Cluster 1**	**CT2 *Streptococcus mitis/oralis* HMT718, CT23, and CT24 *Haemophilus* spp.**	**Illumina MiSeq**
						Cluster 2	**CT6 *Veillonella rogosae/dispar* CT25, CT27 *Neisseria* CT37 *Gemella morbillorum***	
						Cluster 3a	**CT43 *Streptococcus gordonii* CT48 *Rothia aeria***	
						Cluster 3b	**HMT14, CT27 *Neisseria* spp. CT33 Class Bacilli**	
Willis et al. (2018)	13-15	1319	Iberian Peninsula and Balearic Islands	Saliva- mouth wash	Without dental and medical exclusion criteria	stomatotype 1	*Neisseria, Haemophilus*	Illumina
						stomatotype 2	*Veillonella, Prevotella, Streptococcus,*	MiSeq
Zaura et al. (2017)	18-32	268	The Netherlands	Unstimulated saliva	Systemically healthy individuals, sampling during a regular check-up at the dentist.	MIC3	*Veillonella atypica/Veillonella dispar and Prevotella*	Illumina
						MIC2	*Streptococcus mitis group Streptococcus gordonii Rothia mucilaginosa*	MiSeq
						MIC1.3	*Neisseria flavescens Neisseria subflava and Haemophilus parainfluenzae*	
						MIC1.2	*Streptococcus salivarius, Streptococcus vestibularis, Streptococcus australis Streptococcus parasanguinis* and *Granulicatella adiacens*	
						MIC1.1	*Prevotella* sp. HMT313 and *Paraprevotella/Alloprevotella* sp. HMT308	
De Filippis et al. (2014)	18–55	161	Italy	Unstimulated saliva	Systemically healthy individuals. No information about dental examination.	cluster I	*Neisseria, Fusobacterium, Porphyromonas*	454
						cluster II	*Prevotella*	
						cluster III	*Streptococcus, Gemella, Porphyromonas*	
Takeshita et al. (2016)	> 40	2343	Japan	Stimulated saliva	Dental and medical examinations were performed on 68.2% individuals	type I	*Veillonella, Prevotella, Actinomyces, Rothia, S. salivarius*, and *S. parasanguinis*	Ion PGM
						type II	*Streptococcus mitis, Haemophilus, Porphyromonas, Gemella*, and *Neisseiria*	

*Mutually corresponding stomatotypes are color-coded. CT stands for combined taxon (See **Supplementary Table 3**).*

Additionally, all the other OM samples were isolated from saliva, which differs in microbial composition from subgingival plaque. Certain variability in the taxonomic composition of the identified stomatotypes can also be given by demographic differences such as the drinking water source (Willis et al., 2018) or the prevailing diet (Lassalle et al., 2018). Nevertheless, the separate clustering of *Streptococcus* and *Veillonella* based OMs was generally observed in other studies as well (Takeshita et al., 2016; Zaura et al., 2017).

Considering the current knowledge of the microbial succession in the oral cavity during the onset of periodontal disease, and the characteristics of identified genera, we could hypothesize, that only the *Streptococcus*-based Cluster 1 represents healthy OM, while the Cluster 2 could already represent the initial dysbiotic state. The Cluster 1 stomatotype, and also the outliers Cluster 3a and Cluster 3b, are characterized by the predominant presence of early colonizers like *Streptococcus*, *Neisseria* (Mahajan et al., 2013), *Haemophilus* (Kolenbrander et al., 1993), and *Rothia* (Sulyanto et al., 2019) involved in the initial plaque formation. Samples from the stomatotype Cluster 2 also contain early colonizers of genera *Neisseria* and *Gemella* (Mahajan et al., 2013), but most remarkably they are exceptionally rich in *Veillonella* species. The genus *Veillonella* is considered to be a pioneer colonizer as well (Sulyanto et al., 2019), but among others, it is the only highly abundant anaerobic taxon assigned generally to periodontal health. *Veillonella* species possess two characteristics that rank them among the most important bridging taxa in the oral biofilm community. They can utilize the lactate generated mainly by streptococci as their primary energy and carbon source, and they produce catalase protecting *F. nucleatum* and other more fastidious anaerobes against hydrogen peroxide (Rogosa and Bishop, 1964). *Veillonella* also produces nutrients for the survival and growth of periodontal pathogens (Zhou et al., 2017). Therefore, we hypothesize, that the stomatotype Cluster 2 still represents clinically healthy individuals but already with an increased risk of periodontitis development. A further stage in the disease onset and progression could be represented by Cluster 4 (the transient state) with the increased relative abundance of anaerobic CT3 *F. nucleatum* and CT8 *Porphyromonas pasteri/catoniae* but still no or a negligible amount of true periopathogens and mostly no clinical signs of the disease. *F. nucleatum* forms a coaggregation bridge between early aerobic colonizers and other bacteria including anaerobic members of the red complex (*P. gingivalis*, *T. forsythia*, and *T. denticola*) (Bradshaw et al., 1998; Mahajan et al., 2013). This ability of *F. nucleatum* to coaggregate with a wide variety of partner strains is highly unusual (Kolenbrander et al., 2002; Kolenbrander et al., 2006). It has been shown that fusobacteria play a role in protecting against atmospheric oxygen and hydrogen peroxide in the oral biofilm and even support the growth of anaerobes, such as *Porphyromonas gingivalis*, under aerated conditions (Diaz et al., 2002). The presence of *F. nucleatum* in a higher amount thus would enable the periodontitis-associated bacteria to overgrowth the first colonizers.

Nevertheless, it is not only the taxonomic composition of the OM but the overall metabolic activity in the oral habitat including the host response to microbial production, which are the critical factors distinguishing between oral health and dysbiosis resulting in any type of oral pathogenesis. In the majority (~ 90%) of samples, the OM taxonomic composition corresponds well to the state of health and can serve as a fast diagnostic tool, however, still, there are individuals with atypical OM taxonomic composition, the atypical metabolic activity of typical OM or unusual immune reaction toward usual OM – in all these cases, the evaluation of proteome and/or metabolome could provide a more accurate image.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/bioproject, PRJNA670573.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the First Faculty of Medicine of Charles University and General University Hospital in Prague as a part of project No. 17-30753A of the Czech Health Research Council. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JJ and LN postulated the hypothesis and designed the experiments. ML participated on experiment design, evaluated sequencing results and prepared first draft of the manuscript. JM diagnosed the patients and performed sampling. TJ assisted in sampling and performed DNA isolation and primary PCR. BT further processed samples for sequencing and participated on evaluation of sequencing results. OH did the clustering analysis and participated on evaluation of results and figure preparation. LN participated on the evaluation of results and finalized manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by grant 486417 from the Grant Agency of Charles University and project 17-30753A from Czech health research council.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2021.629723/full#supplementary-material

REFERENCES

Abusleme, L., Dupuy, A. K., Dutzan, N., Silva, N., Burleson, J. A., Strausbaugh, L. D., et al. (2013). The subgingival microbiome in health and periodontitis and its relationship with community biomass and inflammation. *Isme J.* 7, 1016–1025. doi: 10.1038/ismej.2012.174

Armitage, G. C., and Cullinan, M. P. (2010). Comparison of the clinical features of chronic and aggressive periodontitis. *Periodontol. 2000* 53, 12–27. doi: 10.1111/j.1600-0757.2010.00353.x

Aronesty, E. (2013). Comparison of sequencing utility programs. *Open Bioinform. J.* 7 (1), 1–8.

Baelum, V., and López, R. (2013). Periodontal disease epidemiology - learned and unlearned? *Periodontol. 2000* 62, 37–58. doi: 10.1111/j.1600-0757.2012.00449.x

Baldrian, P., Kolaiřík, M., Štursová, M., Kopecký, J., Valášková, V., Větrovský, T., et al. (2012). Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *ISME J.* 6, 248–258. doi: 10.1038/ismej.2011.95

Beall, C. J., Campbell, A. G., Griffen, A. L., Podar, M., and Leys, E. J. (2018). Genomics of the uncultivated, periodontitis-associated bacterium Tannerella sp. BU045 (Oral Taxon 808). *mSystems* 3 (3), e00018–18. doi: 10.1128/msystems.00018-18

Belibasakis, G. N. (2018). Microbiological changes of the ageing oral cavity. *Arch. Oral. Biol.* 96, 230–232. doi: 10.1016/j.archoralbio.2018.10.001

Bradshaw, D. J., Marsh, P. D., Watson, K. G., and Allison, C. (1998). Role of Fusobacterium nucleatum and coaggregation in anaerobe survival in planktonic and biofilm oral microbial communities during aeration. *Infect. Immun.* 66, 4729–4732. doi: 10.1128/iai.66.10.4729-4732.1998

Burt, B. A. (1994). Periodontitis and aging: reviewing recent evidence. *J. Am. Dent Assoc.* 125, 273–279. doi: 10.14219/jada.archive.1994.0034

Camelo-Castillo, A. J., Mira, A., Pico, A., Nibali, L., Henderson, B., Donos, N., et al. (2015). Subgingival microbiota in health compared to periodontitis and the influence of smoking. *Front. Microbiol.* 6, 119. doi: 10.3389/fmicb.2015.00119

Caton, J. G., Armitage, G., Berglundh, T., Chapple, I. L., Jepsen, S., Kornman, K. S., et al. (2018). A new classification scheme for periodontal and peri-implant diseases and conditions–Introduction and key changes from the 1999 classification. *J. Periodontol* 89, S1–S8. doi: 10.1002/JPER.18-0157

Chen, T., Yu, W. H., Izard, J., Baranova, O. V., Lakshmanan, A., and Dewhirst, F. E. (2010). The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database* 2010. doi: 10.1093/database/baq013

Cuthbertson, L., Rogers, G. B., Walker, A. W., Oliver, A., Green, L. E., Daniels, T. W. V., et al. (2016). Respiratory microbiota resistance and resilience to pulmonary exacerbation and subsequent antimicrobial intervention. *ISME J.* 10, 1081–1091. doi: 10.1038/ismej.2015.198

De Filippis, F., Vannini, L., La Storia, A., Laghi, L., Piombino, P., Stellato, G., et al. (2014). The same microbiota and a potentially discriminant metabolome in the saliva of omnivore, ovo-lacto-vegetarian and vegan individuals. *PloS One* 9 (11), e112373. doi: 10.1371/journal.pone.0112373

De Lillo, A., Booth, V., Kyriacou, L., Weightman, A. J., and Wade, W. G. (2004). Culture-independent identification of periodontitis-associated Porphyromonas and Tannerella populations by targeted molecular analysis. *J. Clin. Microbiol.* 42, 5523–5527. doi: 10.1128/JCM.42.12.5523-5527.2004

Deo, P. N., and Deshmukh, R. (2019). Oral microbiome: Unveiling the fundamentals. *J. Oral. Maxillofac Pathol. JOMFP* 23, 122–128. doi: 10.4103/jomfp.JOMFP

Deshpande, N. P., Riordan, S. M., Castaño-Rodríguez, N., Wilkins, M. R., and Kaakoush, N. O. (2018). Signatures within the esophageal microbiome are associated with host genetics, age, and disease. *Microbiome* 6, 1–14. doi: 10.1186/s40168-018-0611-4

Diaz, P. I., Zilm, P. S., and Rogers, A. H. (2002). Fusobacterium nucleatum supports the growth of Porphyromonas gingivalis in oxygenated and carbon-dioxide-depleted environments. *Microbiology* 148, 467–472. doi: 10.1099/00221287-148-2-467

Dioguardi, M., Alovisi, M., Crincoli, V., Aiuto, R., Malagnino, G., Quarta, C., et al. (2020). Prevalence of the genus Propionibacterium in primary and persistent endodontic lesions: A systematic review. *J. Clin. Med.* 9 (3), 739. doi: 10.3390/jcm9030739

Dowd, S. E., Callaway, T. R., Wolcott, R. D., Sun, Y., Mckeehan, T., Hagevoort, R. G., et al. (2008). Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). *BMC Microbiol.* 8. doi: 10.1186/1471-2180-8-125

Duran-Pinedo, A. E., and Frias-Lopez, J. (2015). Beyond microbial community composition: functional activities of the oral microbiome in health and disease. *Microbes Infect.* 17, 505–516. doi: 10.1016/j.micinf.2015.03.014

Ebersole, J. L., Graves, C. L., Gonzalez, O. A., Dawson, D., Morford, L. A., Huja, P. E., et al. (2016). Aging, inflammation, immunity and periodontal disease. *Periodontol. 2000* 72, 54–75. doi: 10.1111/prd.12135

Ebersole, J. L., Al-Sabbagh, M., Gonzalez, O. A., and Dawson, D. R. (2018). Ageing effects on humoral immune responses in chronic periodontitis. *J. Clin. Periodontol.* 45, 680–692. doi: 10.1111/jcpe.12881

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Eke, P. I., Dye, B. A., Wei, L., Thornton-Evans, G. O., and Genco, R. J. (2012). Prevalence of periodontitis in adults in the United States: 2009 and 2010. *J. Dent Res.* 91, 914–920. doi: 10.1177/0022034512457373

Eke, P. I., Dye, B. A., Wei, L., Slade, G. D., Thornton-Evans, G. O., Borgnakke, W. S., et al. (2015). Update on prevalence of periodontitis in adults in the United States: NHANES 2009 – 2012. *J. Periodontol.* 86, 611–622. doi: 10.1016/j.physbeh.2017.03.040

Eke, P. I., Wei, L., Borgnakke, W. S., Thornton-Evans, G., Zhang, X., Lu, H., et al. (2016). Periodontitis prevalence in adults ≥ 65 years of age, in the USA. *Periodontol. 2000* 72, 76–95. doi: 10.1111/prd.12145

Feres, M., Teles, F., Teles, R., Figueiredo, L. C., and Faveri, M. (2016). The subgingival periodontal microbiota of the aging mouth. *Periodontol. 2000* 72, 30–53. doi: 10.1111/prd.12136.The

Frencken, J. E., Sharma, P., Stenhouse, L., Green, D., Laverty, D., and Dietrich, T. (2017). Global epidemiology of dental caries and severe periodontitis – a comprehensive review. *J. Clin. Periodontol.* 44, S94–S105. doi: 10.1111/jcpe.12677

Griffen, A. L., Beall, C. J., Campbell, J. H., Firestone, N. D., Kumar, P. S., Yang, Z. K., et al. (2012). Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing. *ISME J.* 6, 1176–1185. doi: 10.1038/ismej.2011.191

Guillermo, I., Perez-Perez, and Blaser, M. J. (1996). "Chapter 23 Campylobacter and Helicobacter," In: S. Baron *Medical Microbiology*, *4th edition* (University of Texas Medical Branch at Galveston).

Guilloux, C., Lamoureux, C., Beauruelle, C., and Héry-Arnaud, G. (2020). Porphyromonas: A neglected potential key genus in human microbiomes. *Anaerobe*. doi: 10.1016/j.anaerobe.2020.102230

Hagberg, A., Swart, P., and Chult, D. (2008). "Exploring network structure, dynamics, and function using NetworkX. @," in *Proceedings of the 7th Python in Science Conference*. Eds. J. M. Gäel Varoquaux and T. Vaught, 11–15.

Hajishengallis, G. (2015). Periodontitis: from microbial immune subversion to systemic inflammation. *Nat. Rev. Immunol.* 15, 30–44. doi: 10.1038/nri3785

Hammer, Ø., Harper, D. A. T., and Ryan, P. D. (2001). Past: Paleontological statistics software package for education and data analysis. *Palaeontol Electron.* 4.1.4, 1–9.

He, J., Huang, W., Pan, Z., Cui, H., Qi, G., Zhou, X., et al. (2012). Quantitative analysis of microbiota in saliva, supragingival, and subgingival plaque of Chinese adults with chronic periodontitis. *Clin. Oral. Investig.* 16, 1579–1588. doi: 10.1007/s00784-011-0654-4

Henderson, B., Ward, J. M., and Ready, D. (2010). Aggregatibacter (Actinobacillus) actinomycetemcomitans: A triple A* periodontopathogen? *Periodontol. 2000* 54, 78–105. doi: 10.1111/j.1600-0757.2009.00331.x

Hernández, M., Dutzan, N., García-Sesnich, J., Abusleme, L., Dezerega, A., Silva, N., et al. (2011). Host-pathogen interactions in progressive chronic periodontitis. *J. Dent Res.* 90, 1164–1170. doi: 10.1177/0022034511401405

Kirst, M. E., Li, E. C., Alfant, B., Chi, Y. Y., Walker, C., Magnusson, I., et al. (2015). Dysbiosis and alterations in predicted functions of the subgingival microbiome in chronic periodontitis. *Appl. Environ. Microbiol.* 81, 783–793. doi: 10.1128/AEM.02712-14

Kolenbrander, P. E., Ganeshkumar, N., Cassels, F. J., and Hughes, C. V. (1993). Coaggregation: specific adherence among human oral plaque bacteria. *FASEB J.* 7, 406–413. doi: 10.1096/fasebj.7.5.8462782

Kolenbrander, P. E., Andersen, R. N., Blehert, D. S., Egland, P. G., Foster, J. S., and Palmer, R. J. (2002). Communication among oral bacteria. *Microbiol. Mol. Biol. Rev.* 66, 486–505. doi: 10.1128/mmbr.66.3.486-505.2002

Kolenbrander, P. E., Palmer, R. J., Rickard, A. H., Jakubovics, N. S., Chalmers, N. I., and Diaz, P. I. (2006). Bacterial interactions and successions during plaque development. *Periodontol. 2000* 42, 47–79. doi: 10.1111/j.1600-0757.2006.00187.x

Kolenbrander, P. E., Palmer, R. J., Periasamy, S., and Jakubovics, N. S. (2010). Oral multispecies biofilm development and the key role of cell-cell distance. *Nat. Rev. Microbiol.* 8, 471–480. doi: 10.1038/nrmicro2381

LaMonte, M. J., Genco, R. J., Buck, M. J., McSkimming, D. I., Li, L., Hovey, K. M., et al. (2019). Composition and diversity of the subgingival microbiome and its relationship with age in postmenopausal women: An epidemiologic investigation. *BMC Oral. Health* 19, 246. doi: 10.1186/s12903-019-0906-2

Lane, D. J., Pace, B., Olsen, G. J., Stahlt, D. A., Sogin, M. L., and Pace, N. R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc. Natl. Acad. Sci. U S A* 82, 6955–6959. doi: 10.1073/pnas.82.20.6955

Lassalle, F., Spagnoletti, M., Fumagalli, M., Shaw, L., Dyble, M., Walker, C., et al. (2018). Oral microbiomes from hunter-gatherers and traditional farmers reveal shifts in commensal balance and pathogen load linked to diet. *Mol. Ecol.* 27, 182–195. doi: 10.1111/mec.14435

Leys, E. J., Lyons, S. R., Moeschberger, M. L., Rumpf, R. W., and Griffen, A. L. (2002). Association of Bacteroides forsythus and a novel Bacteroides phylotype with periodontitis. *J. Clin. Microbiol.* 40, 821–825. doi: 10.1128/JCM.40.3.821-825.2002

Liu, B., Faller, L. L., Klitgord, N., Mazumdar, V., Ghodsi, M., Sommer, D. D., et al. (2012). Deep sequencing of the oral microbiome reveals signatures of periodontal disease. *PloS One* 7 (6), e37919. doi: 10.1371/journal.pone.0037919

López, R., Smith, P., Göstemeyer, G., and Schwendicke, F. (2017). Ageing, dental caries and periodontal diseases. *J. Clin. Periodontol.* 44, S145–S152. doi: 10.1111/jcpe.12683

Macuch, P. J., and Tanner, A. C. (2000). Campylobacter species in health, gingivitis, and periodontitis. *J. Dent Res.* 79, 785–792. doi: 10.1177/00220345000790021301

Mahajan, A., Singh, B., Kashyap, D., Kumar, A., and Mahajan, P. (2013). Interspecies communication and periodontal disease. *Sci. World J.* 2013. doi: 10.1155/2013/765434

Meuric, V., Gall-David, S., Boyer, E., Acuña-Amador, L., Martin, B., Fong, S. B., et al. (2017). Signature of microbial dysbiosis in periodontitis. *Appl. Environ. Microbiol.* 83, e00462–17. doi: 10.1128/AEM.00462-17

Monnet-Corti, V., Antezack, A., and Moll, V. (2018). Vestibular frenectomy in periodontal plastic surgery. *J. Dentofac Anomalies Orthod.* 21 (2), 205. doi: 10.1051/odfen/2018054

Najmanova, L., Sabova, L., Lenartova, M., Janatova, T., Mysak, J., Vetrovsky, T., et al. (2021). R/G value – a numeric index of periodontal health. *Front. Cell. Inf. Microbiol* 11, 602643. doi: 10.3389/fcimb.2021.602643

Okuda, T., Okuda, K., Kokubu, E., Kawana, T., Saito, A., and Ishihara, K. (2012). Synergistic effect on biofilm formation between Fusobacterium nucleatum and Capnocytophaga ochracea. *Anaerobe* 18, 157–161. doi: 10.1016/j.anaerobe.2012.01.001

Park, O.-J., Yi, H., Jeon, J. H., Kang, S.-S., Koo, K.-T., Kum, K.-Y., et al. (2015). Pyrosequencing analysis of subgingival microbiota in distinct periodontal conditions. *J. Dent Res.* 94, 921–927. doi: 10.1177/0022034515583531

Pérez-Chaparro, P., Gonçalves, C., Figueiredo, L., Faveri, M., Lobão, E., Tamashiro, N., et al. (2014). Newly identified pathogens associated with periodontitis: a systematic review. *J. Dent Res.* 93, 846–858. doi: 10.1177/0022034514542468

Persson, G. R. (2018). Periodontal complications with age. *Periodontol. 2000* 78, 185–194. doi: 10.1111/prd.12227

Pudakalkatti, P. S., Baheti, A. S., Hattarki, S. A., Kambali, S. S., and Naik, R. M. (2016). Detection and prevalence of Capnocytophaga in periodontal health and disease. *J. Orofac Sci.* 8, 92–95. doi: 10.4103/0975-8844.195911

Rogosa, M., and Bishop, F. S. (1964). The genus Veillonella. *J. Bacteriol.* 88, 37–41. doi: 10.1128/jb.88.1.37-41.1964

Rossum, G. V., and Drake, F. *The Python Language Reference*. Amsterdam, Netherlands: Python Software Foundation (2010).

Rusthen, S., Kristoffersen, A. K., Young, A., Galtung, H. K., Petrovski, B. É., Palm, Ø., et al. (2019). Dysbiotic salivary microbiota in dry mouth and primary Sjögren's syndrome patients. *PloS One* 14 (6), e0218319. doi: 10.1371/journal.pone.0218319

Sato, T., and Nakazawa, F. (2014). Coaggregation between Prevotella oris and Porphyromonas gingivalis. *J. Microbiol. Immunol. Infect.* 47, 182–186. doi: 10.1016/j.jmii.2012.09.005

Schacher, B., Baron, F., Roßberg, M., Wohlfeil, M., Arndt, R., and Eickholz, P. (2007). Aggregatibacter actinomycetemcomitans as indicator for aggressive periodontitis by two analysing strategies. *J. Clin. Periodontol.* 34, 566–573. doi: 10.1111/j.1600-051X.2007.01080.x

Socransky, S. S., Haffajee, A. D., Cugini, M. A., Smith, C., and Kent, R. L. (1998). Microbial complexes in subgingival plaque. *J. Clin. Periodontol.* 25, 134–144. doi: 10.1111/j.1600-051X.1998.tb02419.x

Stoddard, S. F., Smith, B. J., Hein, R., Roller, B. R. K., and Schmidt, T. M. (2015). rrnDB: Improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res.* 43, D593–D598. doi: 10.1093/nar/gku1201

Sultan, A. S., Kong, E. F., Rizk, A. M., and Jabra-Rizk, M. A. (2018). The oral microbiome: A Lesson in coexistence. *PloS Pathog.* 14, e1006719. doi: 10.1371/journal.ppat.1006719

Sulyanto, R. M., Thompson, Z. A., Beall, C. J., Leys, E. J., and Griffen, A. L. (2019). The Predominant Oral Microbiota Is Acquired Early in an Organized Pattern. *Sci. Rep.* 9, 1–8. doi: 10.1038/s41598-019-46923-0

Szafranski, S. P., Wos-Oxley, M. L., Vilchez-Vargas, R., Jáuregui, R., Plumeier, I., Klawonn, F., et al. (2015). High-resolution taxonomic profiling of the subgingival microbiome for biomarker discovery and periodontitis diagnosis. *Appl. Environ. Microbiol.* 81, 1047–1058. doi: 10.1128/AEM.03534-14

Takeshita, T., Kageyama, S., Furuta, M., Tsuboi, H., Takeuchi, K., Shibata, Y., et al. (2016). Bacterial diversity in saliva and oral health-related conditions: the Hisayama Study. *Sci. Rep.* 6, 22164. doi: 10.1038/srep22164

Tsai, C. Y., Tang, C. Y., Tan, T. S., Chen, K. H., Liao, K. H., and Liou, M. L. (2018). Subgingival microbiota in individuals with severe chronic periodontitis. *J. Microbiol. Immunol. Infect.* 51, 226–234. doi: 10.1016/j.jmii.2016.04.007

Van der Velden, U. (2017). What exactly distinguishes aggressive from chronic periodontitis: is it mainly a difference in the degree of bacterial invasiveness? *Periodontol. 2000* 75, 24–44. doi: 10.1111/prd.12202

Větrovský, T., Baldrian, P., and Morais, D. (2018). SEED 2: A user-friendly platform for amplicon high-throughput sequencing data analyses. *Bioinformatics* 34, 2292–2294. doi: 10.1093/bioinformatics/bty071

Větrovský, T., and Baldrian, P. (2013). The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PloS One* 8. doi: 10.1371/journal.pone.0057923

Willis, J. R., González-Torres, P., Pittis, A. A., Bejarano, L. A., Cozzuto, L., Andreu-Somavilla, N., et al. (2018). Citizen science charts two major "stomatotypes" in the oral microbiome of adolescents and reveals links with habits and drinking water composition. *Microbiome* 6, 218. doi: 10.1186/s40168-018-0592-3

Willis, J. R., and Gabaldón, T. (2020). The human oral microbiome in health and disease: From sequences to ecosystems. *Microorganisms* 8, 308. doi: 10.3390/microorganisms8020308

Yang, N. Y., Zhang, Q., Li, J. L., Yang, S. H., and Shi, Q. (2014). Progression of periodontal inflammation in adolescents is associated with increased number of Porphyromonas gingivalis, Prevotella intermedia, Tannerella forsythensis, and Fusobacterium nucleatum. *Int. J. Paediatr. Dent* 24, 226–233. doi: 10.1111/ipd.12065

Yasunaga, H., Takeshita, T., Shibata, Y., Furuta, M., Shimazaki, Y., Akifusa, S., et al. (2017). Exploration of bacterial species associated with the salivary microbiome of individuals with a low susceptibility to dental caries. *Clin. Oral. Investig.* 21, 2399–2406. doi: 10.1007/s00784-016-2035-5

Zaura, E., Brandt, B. W., Prodan, A., Teixeira De Mattos, M. J., Imangaliyev, S., Kool, J., et al. (2017). On the ecosystemic network of saliva in healthy young adults. *ISME J.* 11, 1218–1231. doi: 10.1038/ismej.2016.199

Zhou, P., Li, X., Huang, I.-H., and Qi, F. (2017). Veillonella catalase protects the growth of Fusobacterium nucleatum in microaerophilic and Streptococcus gordonii-resident environments. *Appl. Environ. Microbiol.* 83, e01079–17. doi: 10.1128/AEM.01079-17

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cellular and Infection Microbiology

REVIEW
published: 31 May 2021
doi: 10.3389/fcimb.2021.689197

Low-Abundant Microorganisms: The Human Microbiome's Dark Matter, a Scoping Review

Jéssica Alves de Cena[1], *Jianying Zhang*[2,3], *Dongmei Deng*[3], *Nailê Damé-Teixeira*[1,4]* *and Thuy Do*[4]

[1] Department of Dentistry, School of Health Sciences, University of Brasília, Brasilia, Brazil, [2] Department of Preventive Dentistry, Academic Center for Dentistry Amsterdam (ACTA), University of Amsterdam and VU University Amsterdam, Amsterdam, Netherlands, [3] Xiangya School of Stomatology, Xiangya Stomatological Hospital, Central South University, Changsha, China, [4] Division of Oral Biology, School of Dentistry, University of Leeds, Leeds, United Kingdom

Research on the human microbiome has mainly been restricted to the identification of most abundant microbiota associated with health or disease. Their abundance may reflect their capacity to exploit their niche, however, metabolic functions exerted by low-abundant microrganisms can impact the dysbiotic signature of local microbial habitats. This scoping review aims to map the literature regarding the management of low-abundant microorganisms in studies investigating human microbiome samples. A systematic literature search was performed in 5 electronic databases, as well as grey literature. We selected clinical microbiome studies targeting human participants of any age, from any body site. We also included studies with secondary data which originated from human biofilm samples. All of the papers used next-generation sequencing (NGS) techniques in their methodology. A total of 826 manuscripts were retrieved, of which 42 were included in this review and 22 reported low-abundant bacteria (LB) in samples taken from 7 body sites (breast, gut, oral cavity, skin, stomach, upper respiratory tract (URT), and vagina). Four studies reported microbes at abundance levels between 5 and 20%, 8 studies reported between 1 and 5%, and 18 studies reported below 1%. Fifteen papers mentioned fungi and/or archaea, and from those only 4 (fungi) and 2 (archaea) produced data regarding the abundance of these domains. While most studies were directed towards describing the taxonomy, diversity and abundance of the highly abundant species, low-abundant species have largely been overlooked. Indeed, most studies select a cut-off value at <1% for low-abundant organisms to be excluded in their analyses. This practice may compromise the true diversity and influence of all members of the human microbiota. Despite their low abundance and signature in biofilms, they may generate important markers contributing to dysbiosis, in a sort of 'butterfly effect'. A detailed snapshot of the physiological, biological mechanisms at play, including virulence determinants in the context of a dysbiotic community, may help better understand the health-disease transition.

Keywords: next-generation sequencing, human microbiome, low-abundant microrganisms, scoping review, minority microbiota

OPEN ACCESS

Edited by:
Georgios N. Belibasakis, Karolinska Institutet (KI), Sweden

Reviewed by:
Maribasappa Karched, Kuwait University, Kuwait
Ashu Sharma, University at Buffalo, United States

***Correspondence:**
Nailê Damé-Teixeira
nailedame@hotmail.com; nailedame@unb.br

Specialty section:
This article was submitted to Microbiome in Health and Disease, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: *31 March 2021*
Accepted: *13 May 2021*
Published: *31 May 2021*

Citation:
Cena JA, Zhang J, Deng D, Damé-Teixeira N and Do T (2021) Low-Abundant Microorganisms: The Human Microbiome's Dark Matter, a Scoping Review. Front. Cell. Infect. Microbiol. 11:689197. doi: 10.3389/fcimb.2021.689197

INTRODUCTION

Advances in high-throughput sequencing approaches have revolutionised microbiology and enabled the characterization of the complex ecological contents of microbial communities, however, our understanding of the mechanisms impacting host-microbial homeostasis remains limited (Hajishengallis et al., 2012). Changes to the human gut microbial composition, for example, can influence host health and diseases, and may affect the microbiota at other body sites (Banerjee et al., 2018). A concept of pathogenicity influenced by both microorganisms and the host has been proposed in the damage-response framework (Casadevall and Pirofski, 2003).

Research on the human microbiome has mainly been restricted to comparisons of the most abundant organisms and the identification of a "core" microbiota associated with health or disease. Indeed, the core microbiome may reflect their capacity to exploit their niche, being favoured by nutrients, O_2 concentrations, etc. to allow surface colonisation. However, opportunistic pathogens may contribute to the compositional and or functional shift towards dysbiosis and could be among the minority taxa. Key species could therefore easily be overlooked in next generation sequencing (NGS) analyses (Turnbaugh et al., 2007; Zerón, 2014).

Furthermore, studies using a 16S rRNA metagenomic approach are limited to the identification of bacteria and archaeae (arguably accurately to the genus level), leaving the view of the richness and diversity of the whole microbiome incomplete and underestimated (Brooks et al., 2015). This is certainly true for *Methanobrevibacter smithii*, a member of the *Archaea* domain in a relatively minor constituent of the gut microbiome that contributes to bacterial metabolism in ways that promote host dysbiosis (Hajishengallis et al., 2012). This species and its methanogenic relatives, though in low abundance, have been demonstrated to be capable of providing conditions for the growth of pathogenic bacteria in periodontal sites, driving to periodontitis (Lepp et al., 2004). The composition of the microbial communities can be misinterpreted regarding the presence of virus, archaea, and fungi, making it a challenge to gain a holistic view.

Subsequently, low-abundant microrganisms could be considered the "dark matter" of the human microbiome. Recent studies (Hajishengallis and Lamont, 2016; Wang et al., 2017; Banerjee et al., 2018; Stobernack, 2019; Berg et al., 2020; Xiao et al., 2020) are paying more attention to these organisms, and increasingly taking into account the "keystone species" concept, corresponding to organisms which effect on the community is disproportionately large compared to their relative abundance (Power et al., 1996). A similar concept in macroecology suggests species in low abundance have a major role in their respective community (Hajishengallis et al., 2012). Abundance is the factor differentiating keystone microorganisms from those that are dominant. A dominant species might affect the environment exclusively by its sheer abundance, while a keystone microorganism may influence metabolic functions of the microbiome, despite its low abundance. Examples of keystone pathogens are: *Porphyromonas gingivalis* associated with periodontitis (Holt and Ebersole, 2005; Perez-Chaparro et al., 2014; Burmistrz et al., 2015; Camelo-Castillo et al., 2015; Ai et al., 2017; Stobernack, 2019), *Klebsiella pneumonia*, *Proteus mirabilis* (Garrett et al., 2010), and *Citrobacter rodentium* (Bry et al., 2006) associated with intestinal inflammatory diseases; and *Fusobacterium nucleatum* (Kostic et al., 2013; Rubinstein et al., 2013) associated with colon cancer (Banerjee et al., 2018). Furthermore, studies investigating *Bacteroides fragilis*, a pro-oncogenic bacterium, have found it to be a minor constituent of the colon microbiota in terms of relative abundance. Its unique virulence characteristics, such as secretion of a zinc-dependent metalloprotease toxin, alter colonic epithelial cells and mucosal immune function to promote oncogenic mucosal events, in which in addition to the intraluminal environment, enhance the oncogenic process. This gave rise to the concept of "alpha-bugs", due to its ability to be directly pro-oncogenic but also to be capable of remodeling the entire healthy microbiota (Sears and Pardoll, 2011; Hajishengallis et al., 2012). Thus, the identification of low-abundant organisms within a microbial population associated with disease could be crucial. Unless we have a more "complete" view of the microbiota, including an accurate detection of low-abundant species, our understanding of the microbiology remains limited, as well as our strategy to improve therapy designs/interventions in diseases with polymicrobial cause.

Studies of the minority microrganisms may reveal unique signatures, which could lead to diseases. Hence, a much deeper characterization of their presence in the microbiome in which they are involved is desirable. This scoping review aims to map the literature regarding the management of low-abundant organisms in studies investigating human samples. We aimed to determine: 1) How researchers classify organisms as low-abundant; 2) How they handled and processed NGS data of low-abundant organisms bioinformatically and 3) The distibution of low-abundant microorganisms among various body sites.

METHODS

Study Design

This is a scoping review to map the literature on low-abundant organisms in the human microbiome, conducted using the PRISMA Extension for Scoping Reviews (PRISMA-ScR) checklist (Tricco et al., 2018).

Search Strategy

Systematic literature wide opened search was performed in electronic databases, also including the grey literature (**Figure 1**). General controlled vocabulary (MeSH Terms) and keywords were used and the searches had no language, year, or publication type restriction. The main terms included "microbiota", "microbiome", "human microbiota", "low abundant", "minority species", "keystone". The search strategy and the results retrieved in each electronic database are shown in **Appendix 1**. Duplicated references were removed by the reference manager EndNoteWeb (Clarivate Analytics, Mumbai) and then manually.

Eligibility Criteria

Studies were included if they satisfied all the following criteria: (1) clinical studies where the target population consisted of humans of any age who were donors of samples from any site; (2) the study design was either a observational study, case series, or any other type of clinical study or studies with secondary data originated from humans; and (3) studies with any term related to low-abundant organisms (e.g. keystones, minority species) in title or abstract.

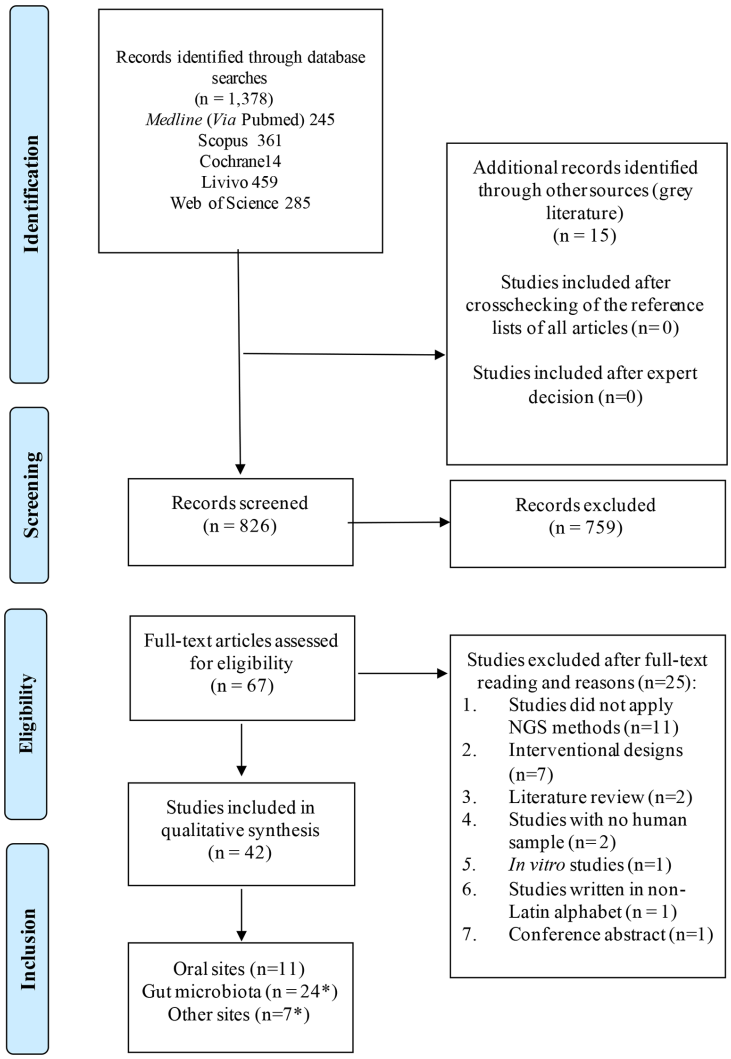


FIGURE 1 | Flow diagram for study selection according to PRISMA guidelines. *Some studies sampled multiple sites in one study.

Studies were excluded if: 1) Studies did not apply next-generation sequencing (NGS) methods to evaluate the microbiota; 2) They were designed as intervention studies; 3) They were literature review, conference abstracts, *in vitro* or animal studies, or any other kind of study carried out without human samples in a primary or secondary analysis; and 4) They were written in a non-Latin alphabet.

Selection of the Manuscripts

Two reviewers, JAC and JYZ, independently screened the eligibility of all identified titles and abstracts for inclusion in the full-text review at the Rayyan QCRI® (Qatar Computer Research Institute, Qatar). Any conflict that arose were resolved by a third reviewer. The same reviewers evaluated full-text articles for inclusion using the same inclusion and exclusion criteria. The list of selected articles was analysed to identify manuscripts that could have been lost during searches in the electronic database.

Data Extraction

Data extraction was performed by the two reviewers independently, and included the following information: Author (year), country, design of the study, range of age of patients, sampling site, type of sample, the platform of sequencing; method of sequencing (16S rRNA or metagenomics or metranscriptomics), method of data analysis and bioinformatics; and abundance of species considered as low-abundant/minority microrganisms. All extracted data was checked by a third reviewer.

RESULTS AND DISCUSSION

Characteristics of the Selected Studies

The systematic literature search resulted in 826 manuscripts of which 67 were considered for full-text review after removing duplicates and applying the eligibility criteria. Following full text reading, 42 studies remained (**Figure 1**; **Table 1**). **Figure 2** shows

TABLE 1 | Qualitative Data Synthesis of the Included Studies (n = 42).

Reference	Sampling site	N	Platform of sequencing	Method of sequencing	Method of data analysis and bioinformatics	Proportion considered low abundant	Low-abundant microbiota
(Ai et al., 2017)	Oral	43	Illumina sequencing	Secondary data from metagenomics (Duran-Pinedo and Yost)	TagCleaner, PRINSEQ, Deconseq e FLASH, MetaPhlAn, GRAMMy, Network analysis	NA	*Porphyromonas gingivalis*, *Haemophilus haemolyticus*, *Prevotella melaninogenica*, and *Capnocytophaga ochracea* were considered potential keystones.
(Albert et al., 2015)	Vagina	310	454 GS Junior pyrosequencer	cpn60 PCR amplicon	Reads were mapped using Bowtie 2, microbial Profiling Using Metagenomic Assembly pipeline (mPUMA)	4 species represented less than 0.3% of the overall reads mapped.	*Atopobium vaginae*, *Streptococcus devriesei*, *Lactobacillus acidophilus, Weissella viridescens.*
(Al-hebshi et al., 2016)	Oral	12	454 GS FLX pyrosequencer	16S rRNA (V1-V3)	Uchime, SILVA-HOMD database, ChimeraSlayer, BLASTN identity ≥98%	Together making up 0.77%	Saccharibacteria (TM7) and SR1.
(Balan et al., 2018)	Oral	24	Illumina MiSeq	16s rRNA (V4)	UPARSE (97%), Uchime, RDP Classifier v.2.2 against the Greengenes database, alignment at SILVA 108, Identification of keystone species was done using the CytoHubba plugin.	>0.5%<5.28% (keystones identification)	*Porphyromonas gingivalis* (2.22%), *Treponema denticola* (1.10%) and *Fretibacterium* sp. OT 361 (0.67%) in supragingival plaque; *Prevotella_intermedia* (0.56%) in the saliva; *Porphyromonas endodontalis* (plaque 0.89%; saliva 0.91%).
(Brawner et al., 2017)	Stomach	86	Illumina platform	16S rRNA (V4)	UCLUST (100%), filter >10 reads, RDB classifier, Multiple sequence, alignment with PyNAST.	OTUs <1% were not analysed.	*H. pylori*-infected children harboured significantly reduced proportions of three bacterial classes (Actinobacteria, Bacilli, and Gammaproteobacteria), three orders (Pseudomonadales, Actinomycetales, and Lactobacillales) and four families (Streptococcaceae, Moraxellaceae, Actinomycetaceae, and Carneobacteriaceae) compared with fluids from non-infected children, but all with proportion >1%.
(Camelo-Castillo et al., 2015)	Oral	60	454 GS FLX pyrosequencer	16S rRNA	MG-RAST, Monthur, RDB classifier, BLASTN (>97%).	<1%	Desulfobulbus (especially D. propionicus) and Filifactor (F. alocis) with the periodontal inflammation severity, and a negative association of Anaeroglobus (especially A. geminatus) and TM7.
(Camelo-Castillo et al., 2019)	URT	56	454 GS FLX pyrosequencer	16S rRNA (V1-V4)	Prinseq, RDP database (80%), OUT>97% identify.	OTUs <0.1% were not analysed; Low abundance at <1%.	Corynebacterium, Neisseria, Actinomyces, or Rothia, among others, accounting for 9% of the reads.
(Claussen et al., 2017)	Gut	822	454 GS FLX pyrosequencer	16S rRNA (V1-V2)	Entropy Shifts of Abundance Vectors under Boolean Operations (ESABO).	0.1%-0.4%	At phylum level: Chlorobi, Chloroflexi, Deferribacteres, Deinococcus-Thermus, Gemmatimonadetes, OP10, Planctomycetes, Thermodesulfobacteria, WS3.
(Dame-Teixeira et al., 2020)	Oral		Ion PGM	16S rRNA	Prinseq, USEARCH, UCLUST (97%); RDP, SILVA 132.	≤0.035%	Thaumarchaeota.
(Das et al., 2018)	Gut	84	454 GS FLX pyrosequencer	16S rRNA (V1-V5)	UCLUST (97%); RDP; SILVA.	0.01%-0.05% in at least 50% of the samples.	Phylum (Verrucomicrobia, Tenericutes and Fusobacteria); Class (Verrucomicrobia, Mollicutes and Fusobacteria); Order (Verrucomicrobiales, Bifidobacteriales, Desulfovibrionales, Anaeroplasmatales, Fusobacteriales, Rhizobiales, and Caulobacterales).
(de Goffau et al., 2013)	Gut	18	454 GS FLX pyrosequencer	16S rRNA (V1-V3)	RDP classifier, SILVA	OTUs <0.005% were not analysed; Low abundance at <12%.	Low abundance of Bifidobacteria and butyrate-producing species in children with β-cell autoimmunity.

(Continued)

TABLE 1 | Continued

Reference	Sampling site	N	Platform of sequencing	Method of sequencing	Method of data analysis and bioinformatics	Proportion considered low abundant	Low-abundant microbiota
(Dobbler et al., 2017)	Gut	132	PGM Ion Torrent;	16s rRNA (V4);	BMP Operating System (BMPOS), UPARSE, UCLUST (97%) method	>0.5 (16S rRNA)	Low abundance of *Lactobacillus* sp. in Necrotizing Enterocolitis (NEC);
			Oxford Nanopore MinION.	Metagenomics.	against the Greengenes 13.5 database	>0.38% (metagenomics)	4 day of life (without NEC) = Firmicutes (13.14%) and Actinobacteria (2.47%); 5-7 day of life (without NEC) = Bacteroidetes (13.47%) and Actinobacteria (0.54%); Fungi (*Saccharomyceta* class) = 0.38%, no virus or archaea detected.
(Feng et al., 2015)	Ascites	7	Illumina Miseq	16S rDNA (v3)	BLAST NCBI (98.5% similarity);	NA	Cyanothece, Bacillus, Streptococcus; Salmonella, Pantoea, Cupriavidus; Rothia, Faecalibacterium, Acinetobacter.
			Illumina Hiseq	Whole Genome Sequencing (WGS)	NCBI mega-blast (90% identity).		The WGS approach was better at identifying microbes with a low abundance.
(Fisher and Mehta, 2014)	Gut	NA	NA	Secondary data from metagenomics (Caporaso et al.)	Data were obtained from the MGRAST database; Learning Interactions from MIcrobial Time Series (LIMITS).	NA	*Bacteroides fragilis* and *Bacteroides stercosis* act as keystone species.
(Ghannoum et al., 2010)	Oral	20	454 GS FLX pyrosequencer	ITS1F-ITS4A (mycobiome)	BLAST Genbank (98%), Fungal ITS sequences were compared with the Assembling Fungal Tree of Life (AFTOL).	OTUs <1% were not analysed.	74 fungi genera detected (7 in high abundance); Authors declare that low-abundance genera may represent environmental fungi present in the oral cavity and could simply be spores inhaled from the air or material ingested with food.
(Hauser et al., 2015)	URT	54	454 GS FLX pyrosequencer	16S rRNA (V1-V3)	Uchime, BLAST SILVA, 111NR (95%)	OTUs <1% were not analysed.	20 minor bacterial species in one subject with completely negative culture; Low-abundance taxa were detected in 4.5% of cultures.
(Heisel et al., 2015)	Gut	11	Illumina MiSeq	ITS2 of the 18S rDNA fungal locus Validation with qPCR	UCLUST, USEARCH, alignment using MUSCLE, Mothur (hash.txt and fungalITSdatabaseID)	Present at <1.5% mean abundance across all samples.	*Candida krusei* and *Candida parapsilosis*.
(Iebba et al., 2018)	Gut, blood	60	Illumina MiSeq	16S rRNA (V3-V4)	Python v.2.7.11, Mothur v.1.38.1, SILVA v.1.1961	≥0.5%	*Gemmiger formicilis, Oscillibacter ruminantium, Roseburia faecis and Faecalibacterium prausnitzii* were significantly higher in the controls than in cirrhotic patients, being classified as keystone species.
(Kang et al., 2017)	Gut	1463	Illumina MiSeq	16S rRNA (V3-V4)	USERCH 6.1 within the QIIME (97% similarity).	OTUs <0.005% were not analysed; Low abundance = 4.1%.	Unclassified Clostridiales (associated with the group with focal or intense FDG uptake in the intestine).
(Kowalska-Duplaga et al., 2019)	Gut	82	Illumina MiSeq	16S rRNA (V3-V4)	QIIME2, DADA2, Greengenes database (99% similarity).	Low abundance in Crohn's disease: 0.67%, 0.27%, 0.49%, 3.89%, 0.62%, and 0.35%, respectively.	Butyrate-producing bacteria, including Bifidobacterium (*B. adolescentis*), Roseburia (*R. faecis*), Faecalibacterium (*F. prausnitzii*), Gemmiger (*G. formicilis*), Ruminococcus (*R. bromii*) and Veillonellaceae (Dialister).
(Lau et al., 2016)	Gut	5	Illumina MiSeq	16S rRNA (v3)	Cutadapt, PANDAseq, AbundantOTU, QIIME, Greengenes database (97% similarity).	OTUs <0.01% were not analysed; Low abundance = <1%.	Uncultured OTUs were of low abundance (<0.8% relative abundance) in the culture-independent sequencing; 12 OTUs with relative abundances >0.1% were not cultured from the donor samples and included

(Continued)

TABLE 1 | Continued

Reference	Sampling site	N	Platform of sequencing	Method of sequencing	Method of data analysis and bioinformatics	Proportion considered low abundant	Low-abundant microbiota
							Cyanobacteria, Clostridia, Mollicutes, and Bacteroidetes.
(Li et al., 2013)	Oral, skin, distal gut, and vaginal	200	454 GS FLX pyrosequencer	16S rRNA	Taxonomic Variance, Binomial Distribution qualify Presence and Absence, Ubiquity vs. Abundance (Ub-Ab) Plots, Ubiquity-Ubiquity Plots (U-U Plots), HMP Consortium.	OTUs <0.01%/90% ubiquity were not analysed; Definition of low abundance <1% (minor core taxa).	Buccal mucosa (Coriobacteriaceae Atopobium, Prevotellaceae unclassified, Bacilli unclassified, Lachnospiraceae Catonella). Hard palate (Clostridiales Family XIII. Mogibacterium, Lachnospiraceae Catonella). Keratinized gingiva (Bacilli unclassified). Palatine tonsils (Clostridiales Family XIII. Mogibacterium, Firmicutes unclassified). Saliva (Actinomycetales unclassified, Porphyromonadaceae Tannerella, Neisseriaceae Kingella). Subgingival plaque (Firmicutes unclassified). Supragingival plaque (Betaproteobacteria unclassified). Throat (Clostridiales Family XIII. Mogibacterium, Firmicutes unclassified). Tongue dorsum (Actinomycetales unclassified, Bacilli unclassified, Peptostreptococcaceae Peptostreptococcus). Anterior nares (Pseudomonadaceae Pseudomonas). Stool (Streptococcaceae Streptococcus).
(Li et al., 2019)	Oral	35	Illumina MiSeq	16 rRNA and ITS2	UPARSE (>97% identity), RDB classifier, UNITE database.	Core mycobiome: OTUs <0.1% were not analysed; Key oral fungal microbiota: OTUs with frequencies of at least 50% and relative abundances of ≥0.5% were analysed.	Keystone fungal genera (Bovista, Erysiphe, Psathyrella, etc.)
(Ling et al., 2010)	Vaginal	100	454 GS FLX pyrosequencer	16s rRNA (V3)	MOTHUR (versão 1.5.0), RDP Classifier (80%), MEGA.	0.1-1.0% of total sequences.	Chloroflexi, Tenericutes, Proteobacteria and candidate division TM7; Mobiluncus.in low abundance (not described the %).
(Liu et al., 2019)	Gut	119	Illumina MiSeq	16S rRNA (V4-V5)	BIPES pipeline, AUCHIME, QIIME (1.9.1) USEARCH, PyNAST, Greengenes database, RDP Classifier.	OTUs with median in any group <0.3% were not analysed.	*Collinsella aerofaciens* and *P. copri* is a possible keystones for cardiac valve calcification and coronary artery disease.
(Nakayama et al., 2017)	Gut	43	454 GS FLX pyrosequencer	16S rRNA (V6-V8)	QIIME, USEARCH (97% identity), RDP classifier.	<1%.	Prevotellaceae in one of the groups of children.
(Nash et al., 2017)	Gut	147	Illumina MiSeq	16S rRNA (V3-V5) 18S rRNA (ITS2) Metagenomics(fungi)	USEARCH, UCHIME, NCBI GenBank Plant (including fungi) and Environmental databases, SILVA (bacteria), UPARSE, DIAMOND (metagenomics).	–	Mycobiome is relatively low abundant; ITS2 sequencing provided greater resolution of the relatively low abundance mycobiome constituents.
(Ozkan et al., 2019)	Skin ocular	104	Illumina MiSeq	16S rRNA (V4)	UNOISE, USEARCH, Silva 128.	OTUs <1% across all samples were not analysed.	*Corynebacterium, Staphylococcus* in some sites.
(Rocas et al., 2016)	Oral	10	Illumina MiSeq	16s rRNA (V4)	Mothur v.1.36.1, Silva, UCHIME, RDP classifier (80%)	<0.1% not shown	Phylum level: Tenericutes, Synergistes

(Continued)

TABLE 1 | Continued

Reference	Sampling site	N	Platform of sequencing	Method of sequencing	Method of data analysis and bioinformatics	Proportion considered low abundant	Low-abundant microbiota
							Genus level: Megasphaera, Hawardela, Slakia, Filifactor, Parviromonas, Tannarella, Scardovia, others.
(Sakwinska et al., 2016)	Breast milk	90	Illumina MiSeq	16s rRNA (V4); Confirmation by qPCR.	Mothur, Silva, RDP classifier (80%).	0.03%-0.5%.	Bifidobacteria and lactobacilli in low abundance in few samples.
(Simón-Soro et al., 2014)	Oral	13	454 GS FLX pyrosequencer	16S rRNA	Uchime, assigned to Ribosomal Database Project with 97% identity; RDP pyrosequencing pipeline; BLASTN>99%	0.02%- 1%	Tannerella, Olsenella, Filifactor, and Treponema (dentin carious lesions); *Streptococcus mutans* (enamel and dentin carious lesions); *Porphiromonas* (enamel carious lesions).
(Singh and Manning, 2016)	Gut	200	454 GS FLX pyrosequencer	16S rRNA (V3-V5)	QUIIME, USEARCH, Greengenes database,	No cuttoff defined in the methods, but OTUs with 0.03% were described.	*Bacterioides;* *Prevotella* in the group >70 years-old.
(Son et al., 2015)	Gut	59	Illumina MiSeq	16S rRNA (V1-V2 and V1-V3)	Uchime, Silva.	OTUs with a maximal relative abundance <0.0001 and with a prevalence <0.01 were culled; Low abundance (at the genera level) threshold of significance FDR<0.1	Cyanobacterial, Chloroplast, Firmicutes, Asteroleplasma, Proteobacterial, Thalassospira, Burkholderia, Comanonadaceae, Bacteroidetes, Prevolellaceae, Actinobacteria, Mobiluncus, Sutterella, Bacteriodetes, Prevotella, Fusobacteria, Fusobacteriales.
(Wan et al., 2018)	Gut	30	Illumina MiSeq	16S rRNA (V3-V4)	QIIME, Monthur.	No cuttoff defined in the methods, but OTUs with 0.12% were described; Low abundance described as 3.53%, 0.12%.	A high abundance of Proteobacteria and Fusobacteria was observed in most septic shock patients, whereas low abundance was observed in healthy subjects.
(Wang et al., 2017)	Oral	41	PacBio RS II	16S rRNA (V1-V9)	Pacbio circular consensus sequencing, Mothur v.1.36.1, UCHIME, QIIME (97% similarity).	OTUs with a median relative abundance <0.01% were not analysed.	*Haemophilus* spp., *Neisseria* spp., *Rothia* sp. *A. aphrophilus*, *Bergeyella* sp. clone oral AK152, and *S. rubneri* were in low abundance in both the caries group and the transitional group after the 6 month follow-up.
(Zhang et al., 2018)	URT	98	Illumina MiSeq	16S rRNA	QIIME.	No cuttoff defined in the methods, but OTUs with 0.42% were described.	Streptococcus and Rothia (0.68%) keep low abundance in orofarynx microbiota of children ≤1 year old; Oropharynx: Atopobium, Moraxella (0.42, 0.51%).
(Wu et al., 2019)	Gut	59	Illumina HiSeq	Paired-end metagenomic sequencing.	MetaPhlAn2.	Relative abundance lower than 5 in the centenarians;	A lower relative abundance for Faecalibacterium (*Faecalibacterium prausnitzii*), Ruminococcus (*Ruminococcus* sp_5_1_39BFAA), Corprococcus, *Eubacterium rectale*, and Dorea was observed in the centenarians;
						Low-abundant genera were summed into one group to plot.	Description of the *Archaea* domain; *Methanobrevibacter* was enriched.
(Zakrzewski et al., 2019)	Colon	73	Illumina MiSeq	16S rRNA (V3-V4)	QUIIME, Greengenes database (97% identity), UCLUST, UCHIME.	NA	On family level (Ruminococcaceae and Christensenellaceae abundance lower in ileal Crohn's disease group).
(Zeng et al., 2019)	Gut	141	Illumina HiSeq 2500	16S rRNA (V4)	QIIME, Greengenes database (97% identity), DADA2.	<0.1%	Low abundance of butyrate-producing bacteria (Lachnospiraceae, Ruminococcaceae,

(Continued)

TABLE 1 | Continued

Reference	Sampling site	N	Platform of sequencing	Method of sequencing	Method of data analysis and bioinformatics	Proportion considered low abundant	Low-abundant microbiota
							Faecalibacterium, Roseburia, Lachnospira, and Oscillospira) with a higher risk of stroke.
(Zhang et al., 2015)	Antrum, proximal body and fundus	27	Illumina HiSeq 2500	Whole genome sequencing, confirmed by qPCR	PathSeq, Burrows-Wheeler Aligner, Bowtie2, PathoScope	No cuttoff defined in the methods, but OTUs with 0.001% were described.	The pipeline from the authors and Kraken identified high levels of *H. influenzae* (82.9 % and 75.4 %, respectively) as well as *P. acnes* (17.1 % and 24.6 %, respectively), whereas no bacteria were identified by MetaPhlAn; Main advantage of the approach from the authors over MetaPhlAn and Kraken is in samples with low levels of bacteria where the abundance of human DNA confounds bacteria detection.
(Zhang et al., 2019)	Gastric and esophageal sites	12	Illumina HiSeq 2500	Whole Genome Sequencing, confirmed by qPCR	Four different aligners: BWA, RepeatMasker, BLAST, MegaBlast.	NA	*H. pylori* in homogenization method, *Bifidobacterium* sp. and *Pantoea* sp. in lysis method.
(Zhu et al., 2018)	Gut	54	Illumina MiSeq	16s rRNA (V3-V4)	UPARSE, UCHIME, RDP classifier, Silva	53.36±21.44, 3.47±5.41, 1.93±2.71, 2.7±4.89, 1.04 ±1.92, respectively	Five bacterial families: Lachnospiraceae, Peptostreptococcaceae, Erysipelotrichaceae, Coriobacteriaceae, and Clostridiaceae_1 negatively associated with lipopolysaccharide level.

NA, Not available.

FIGURE 2 | Distribution of the literature papers of low abundant organisms by sample's sites.

the distribution of the papers by sampling site. Within them, the gastrointestinal tract and the oral cavity were the most studied ones. It may be due to the higher number of dysbiosis-related diseases or higher bacterial diversity in those sites, since only 10 out of the 42 articles exclusively analyzed samples from healthy individuals, and another 2 did not describe the status of health or disease, as they involved analysis of secondary data. The other sites included the vagina, respiratory system, skin, and blood. According to Hamady et al. (Hamady and Knight, 2009), the majority of microbiome studies describe the use of 16S rRNA gene sequencing for archaea and bacteria, and 18S rRNA gene sequencing for eukaryotes, which have limitations for the accurate identification to the species level.

Figure 3 shows the distribution of sequencing platforms used in the 42 selected articles. The most routinely used sequencing platforms were Illumina, followed by 454/Roche. Although these platforms are different in terms of biochemistry and in the way the matrix is generated, their workflows are conceptually similar (Shendure and Ji, 2008). A study of gut, mouth and skin samples from two subjects found that the composition of the gut and oral communities were not significantly dissimilar when either 454/Roche or Illumina

FIGURE 3 | Distribution of studies by platforms of sequencing.

(**Figure 3**) were used, albeit the communities of the skin were significantly different. This difference was attributed to bias associated with the primers (Caporaso et al., 2011).

Low-Abundant Bacteria (LB)

Out of 42 articles, 20 were excluded from the summary of sample site-related low abundant bacterial species, because the data on microbial abundance were unavailable or no information on low abundance rate was provided. In the remaining 22 studies, low-abundant bacteria (LB) have been reported in the biofilm samples taken from 7 body sites (breast, gut, oral cavity, skin, stomach, upper respiratory tract (URT), and vagina). LB were determined and displayed as the relative abundance of a given operational taxonomic unit (OTU), relative to the total sequencing reads. In total, 4 studies reported LB at abundance levels between 5 and 20%, 6 studies reported between 1 and 5%, and 16 studies reported below 1%. Here we summarized the information of those LB detected at abundance levels below 1%. The information on bacterial phyla can be extracted from all 22 studies, hence it is possible to summarize the major phyla of LB per sample site.

Table 2 summarizes how frequent a phylum was reported as LB (<1%) per site in the 22 studies. The frequency is indicated by the number of studies which have reported LB. In total, 6 different phyla have been reported as LB in more than 2 different studies or in more than 2 different body sites. Gut and oral cavity are the most examined body sites. Out of 6 different phyla, 5 phyla were reported in gut and 6 were reported in oral cavity. Actinobacteria and Firmicutes were the most frequently reported LB among various body sites. Actinobacteria has been found as LB in 6 different body sites. Firmicutes and Proteobacteria were found as LB in 5 different body sites. Compared to the gut, the oral cavity contains a site-specific LB phyla, Spirochaetes.

Table 3 shows the bacterial taxa at the genus level within the major LB phyla (Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria) (<1% abundance). The oral cavity and gut were the most studied body sites, where a low-abundant genus was detected in more than two studies. The reported LB at the genus level in gut was generally different from those of the oral cavity. Only 3 LB genera have been found in both gut and oral cavity, namely, Bifidobacterium, Prevotella and Streptococcus. No LB genus can be reliably identified either in the gut or the oral cavity, since the listed genera were only reported by 1 or 2 studies, which may infer on the diversity of the LB in the human body, or could be biased by sequencing/analysis methods employed.

Actinobacteria were most often reported as a low-abundant phylum among all body sites. In the gut, Actinobacteria are relatively scarce, but have a high degree of ecological connection and are positively correlated with the diversity of the intestinal microbiome, playing an important role in the biodegradation of complex starch. It may be involved in the prevention of dysbiosis in patients with inflammatory bowel disease (Trosvik and de Muinck, 2015). When very abundant, Actinobacteria are associated with obesity (White et al., 2009). In the oral cavity, members of this phylum are part of the healthy microbiota and their abundance varies at each oral sites, however in dental plaque, for example, their abundance is less than 1% (Peterson et al., 2013; Palmer, 2014).

Low-Abundance of Other Organisms

Archaea and fungi (eukaryotes) are usually reported in low abundance, however, this detection should be viewed with caution and further studies are always encouraged to validate and confirm the data. From the 42 selected articles, only 15 mentioned fungi and/or archaea, and from those only 4 (fungi) and 2 (archaea) showed data regarding the abundance of these domains. Ghannoum et al. (2010) described that low-abundance genera may be transient, and represent environmental fungi present in the oral cavity and could simply be spores inhaled from the air or material ingested with food (Ghannoum et al., 2010). They have shown several species not described before in the oral cavity. Heisel et al. showed *Candida krusei* and *Candida parapsilosis* in >1.5% mean abundance in all analysed faecal samples (Heisel et al., 2015). Wu et al., 2019, using shotgun metagenomics, identified methanogenic archaea within the core microbiota, enriched in individuals aged >100 years old (Wu et al., 2019). This technique may therefore be preferrable to 16S rRNA to identify this domain of microrganisms.

The low abundance related to these domains in other studies may be linked to the sample collection method, detection probe, pair of primers used, sequencing technique, and low number of sequences registered in current databases (Ghannoum et al., 2010; Heisel et al., 2015; Dame-Teixeira et al., 2020). Furthermore, the study of the microbial community through the use of 16S rRNA sequencing and shotgun metagenomic methods allows analysis of the composition and genetic capabilities of the microbiota, but not the particularities of the role of low abundance in the microbial community, and of microbial community interactions (Centanni et al., 2018). Microbial communities are complex and constantly changing in response to their environment, influenced by various factors such as diet, use of antibiotics, exposure to transient microorganisms. In this case, other OMICS techniques can be used to understand how microbes react to the environment, including metatranscriptomics, proteomics and metabolomics. Those approaches give a holistic view of the sample content, and a clearer idea of inter-domain interactions within the human microbiome.

Bioinformatics and Data Analysis on Low-Abundant Organisms

Since 1977, DNA-sequencing technology has evolved at a fast pace, and is reshaping our understanding of biology (Srivastava, 2011). Next generation sequencing (NGS) was introduced for the first time in 2005, extending the previous advantages achieved by Sanger sequencing, and facilitated the increase in generated data, while decreasing the cost of sequencing (Buermans and Den Dunnen, 2014). NGS is marked by the construction of libraries, enabling massively parallel sequencing, which has been increasingly simplified, and a higher throughput compared to Sanger sequencing (Ekblom and Galindo, 2011; Muzzey et al., 2015).

TABLE 2 | Number of studies[a] reported low abundant taxa (relative abundance <1%) at the level of phylum.

Phylum	Number of studies per site (n)							Total (n)	References
	Breast	Gut	Oral cavity	Skin	Stomach	URT	Vagina		
Actinobacteria	1	3	3	0	1	1	1	10	(Li et al., 2013; Simón-Soro et al., 2014; Albert et al., 2015; Son et al., 2015; Rocas et al., 2016; Sakwinska et al., 2016; Brawner et al., 2017; Das et al., 2018; Camelo-Castillo et al., 2019; Kowalska-Duplaga et al., 2019)
Bacteroidetes	0	2	4	0	0	0	0	6	(Li et al., 2013; Simón-Soro et al., 2014; Son et al., 2015; Rocas et al., 2016; Nakayama et al., 2017; Balan et al., 2018)
Firmicutes	1	4	3	0	1	0	1	10	(Li et al., 2013; Simón-Soro et al., 2014; Albert et al., 2015; Son et al., 2015; Rocas et al., 2016; Sakwinska et al., 2016; Brawner et al., 2017; Kowalska-Duplaga et al., 2019; Zeng et al., 2019)
Fusobacteria	0	3	1	0	0	0	0	4	(Son et al., 2015; Rocas et al., 2016; Das et al., 2018; Wan et al., 2018)
Proteobacteria	0	2	2	1	1	1	0	7	(Li et al., 2013; Son et al., 2015; Rocas et al., 2016; Brawner et al., 2017; Das et al., 2018; Wan et al., 2018; Camelo-Castillo et al., 2019)
Spirochaetes	0	0	2	0	0	0	0	2	(Simón-Soro et al., 2014; Rocas et al., 2016)

[a]*the phylum reported by at least 2 different studies or found in at least 2 different body sites was included.*

TABLE 3 | Number of studies which reported low abundant taxa (relative abundance <1%) collected from gut and oral cavity.

Taxa identified		Number of studies per site (n)		References
Phylum	Genus	Gut	Oral cavity	
Actinobacteria	Actinomyces	—	1	(Simón-Soro et al., 2014)
	Atopobium	—	1	(Li et al., 2013)
	Bifidobacterium	1	2	(Simón-Soro et al., 2014; Rocas et al., 2016; Sakwinska et al., 2016; Kowalska-Duplaga et al., 2019)
	Mobiluncus	1	—	(Son et al., 2015)
	Olsenella	—	1	(Simón-Soro et al., 2014)
	Unclassified	1	1	(Li et al., 2013; Das et al., 2018)
Bacteroidetes	**Prevotella**	2	2	(Son et al., 2015; Rocas et al., 2016; Nakayama et al., 2017; Balan et al., 2018)
	Tannerella	—	2	(Li et al., 2013; Simón-Soro et al., 2014)
	Unclassified	1	1	(Li et al., 2013; Son et al., 2015)
Firmicutes	Catonella	—	1	(Li et al., 2013)
	Dialister	1	—	(Kowalska-Duplaga et al., 2019)
	Faecalibacterium	2	—	(Kowalska-Duplaga et al., 2019; Zeng et al., 2019)
	Filifactor	—	1	(Simón-Soro et al., 2014)
	Lachnospira	1	—	(Zeng et al., 2019)
	Oscillospira	1	—	(Zeng et al., 2019)
	Peptostreptococcus	—	1	(Li et al., 2013)
	Roseburia	2	—	(Kowalska-Duplaga et al., 2019; Zeng et al., 2019)
	Ruminococcus	1	—	(Kowalska-Duplaga et al., 2019)
	Staphylococcus	—	1	(Rocas et al., 2016)
	Streptococcus	1	1	(Li et al., 2013; Simón-Soro et al., 2014)
	Unclassified	2	1	(Li et al., 2013; Son et al., 2015; Zeng et al., 2019)
Proteobacteria	Burkholderia	1	—	(Son et al., 2015)
	Kingella	—	1	(Li et al., 2013)
	Ochrobactrum	—	1	(Rocas et al., 2016)
	Pseudomonas	—	1	(Rocas et al., 2016)
	Sutterella	1	—	(Son et al., 2015)
	Thalassospira	1	—	(Son et al., 2015)
	Unclassified	2	1	(Li et al., 2013; Son et al., 2015; Das et al., 2018)

—, indicates that the genus was not reported in this body site.
The genera in bold are those identified in both gut and oral sites.

Nevertheless, NGS has some limitations including issues with alignment of short read sequences, detection of artifacts and microbial contaminants present in samples, in addition to the presence of human nucleic acids in clinical samples, thus limiting the analytical sensitivity of microbial detection (Davis et al., 2018). One solution to this limitation was presented as the use of targeted sequencing of the 16S rRNA gene. This gene is now considered as a reference in microbial ecology studies. However, the use of 16S rRNA-based molecular methods do not allow for a high resolution of microbiota identification, because there are biases introduced into molecular community analysis by many factors, such as sample handling, DNA extraction, PCR and

partial sequence of the 16S rRNA gene (ranging between the V1 and V4 regions) (Case et al., 2007). To reduce contamination with sequence artifacts or low accuracy of read alignment, some studies remove sequence reads attributed to low-abundance operational taxonomic units (OTUs) obtained by amplicon sequencing of the 16S rRNA gene. However, it is necessary to perform the analyses with caution, because sequence data associated with these low-abundant taxa may be biologically significant. Therefore, it may not be recommended to exclude these data even if the distinction between expected and unexpected sequences is not always straightforward (Lazarevic et al., 2016).

While microbiome studies generally describe the taxonomy, diversity and abundance of the highly abundant microbes, low-abundant species have been overlooked. Most studies included in this scoping review select a cut-off value at <1% for an organism to be considered low abundant, although some studies have reported OTUs representing 0.003% of the relative abundance (**Table 3**). The choice of such cut-off value were attributed to low read count and or other considerations such as technical artefacts, contaminations, and the presence of transient species. However, by excluding these OTUs from the analysis, the full richness and diversity of the microbiota is underestimated. Camelo-Castillo et al. (2019) stated that only the OTUs representing over 0.1% of the total sequences of each sample were considered for their analysis, as low-frequency reads, including singletons, are more likely to represent sequencing errors, contaminants, or transient organisms without a biological role at the niche under study. Although artifacts and errors are expected, important signals from low-abundant members of microbial community, including keystone organisms, may be lost due to the current technical limitations provided by this strategy. As affirmed before, low-abundant species can be responsible for major functions on the microbial community such as processing certain secondary metabolites. An example comprises organisms from the Archaea domain, that can be detected with 16S rRNA deep sequencing but in very low abundance. Those microrganisms, particularly the methanogens, play a unique role by using hydrogen to produce methane, modulating the environment and were previously described as keystone pathogens associated with periodontal diseases (Camelo-Castillo et al., 2019).

To overcome this limitation, an interesting approach was applied by Li et al (2019), that defined a core microbiome based on high ubiquity taxa in conjunction with a characteristic of high abundance such that the significance of both measurements can be made with a sufficient degree of confidence across and within samples. Using this approach, they were able to classify OTUs with low abundance (<1%) that were highly prevalent across the samples. The authors proposed that larger sample size and sequencing depth are necessary, so that the detection of low abundant taxa may be considered non-spurious across the donors (Li et al., 2019). We believe that defining the ubiquity of the low-abundant microrganisms is a good strategy that should be better explored. A clearer cut-off point to confirm the presence and importance of such species should urgently be defined (minimum values of the sample size, as well as the ubiquity).

Another approach was recommended by Liu et al. (2013), and based on single-read-based, instead of assembly-based classification which has a higher resolution for the characterization of the composition and structure of microbiota, especially for species in low abundance. Their composition and phylogeny-based algorithm uses the strategy of composition comparison, and is capable of classifying millions of very short reads relatively quickly (Liu et al., 2013). Zhang et al (2019) also described two DNA extraction methods (using prolonged lysis and homogenizing methods) which presented marked differences specifically to the low abundance genera (Zhang et al., 2019), and might represent an important improvement in the field.

Metagenomic studies produce high-throughput sequence data that attempt to classify the taxonomy and function of all microbial communities and are greatly affected by the presence of sequencing errors that may influence the estimation of taxonomic diversity (Keegan et al., 2012). There are noise and errors in the sequencing data that can be influenced by the type of platform used. In the studies included in this review, the most commonly used platform was Illumina. With this platform, when errors occur, they are predominantly substitution-type and the error percentage for most Illumina sequence reads is approximately 0.5% (1 error in 200 bases) (Mardis, 2013). The Ion Torrent PGM and 454 GS Junior platforms produced a higher error rate associated with homopolymers around 1.5 and 0.38 errors per 100 bases, respectively (Loman et al., 2012). All platforms are considered suitable for metagenomic sequencing, but no instrument can generate completely accurate data sets, each technology has advantages and disadvantages (Luo et al., 2012). The length of reads generated, sequencing depth and error rates may be taken into account when choosing the most appropriate platform to use. For example, longer reads as those provided by MiSeq (Illumina), Ion Torrent, PacBio and Oxford Nanopore Technologies, are important to consider when carrying out 16S rRNA metagenomics, or genome sequencing (Winand et al., 2020).

CONCLUSION

There is currently no consensus in the literature on the classification of low-abundant organisms. Some studies have described such organisms being detected at less than 1% relative abundance, however, most studies use the same cutoff point (i.e. <1%) to exclude them, due to the risk of contamination or artifacts. This practice may compromise the identification of the true diversity of human microbiota. Domains other than *Bacteria* are neglected due to the cut-off, excluding OTUs with relative abundance lower than 0.1% or 1%. Representatives of Archaea, Fungi or Viruses are little explored. There is growing interest in developing new bioinformatics tools, such as single-read-based, instead of assembly-based, classification to obtain a

higher resolution of the taxonomic analysis. Also, the ubiquity classification associated with the abundance could be a good strategy to identify the low-abundant microbiota. To achieve this, higher sequencing depths should be used in future microbiome investigations, as well as more holistic approaches including shotgun metagenomics should be employed to have a better view of the richness and diversity at play in health, disease and dysbiotic stages.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

The authors are grateful to the Scientific Initiation Program from the Brazilian National Council for Scientific and Technological Development (CNPq), and to the UK's Academy of Medical Sciences, Newton International Fellowship (Grant no. NIF/R5/242) for their support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2021.689197/full#supplementary-material

REFERENCES

Ai, D. M., Huang, R. C., Wen, J., Li, C., Zhu, J. P., and Xia, L. C. (2017). Integrated Metagenomic Data Analysis Demonstrates That a Loss of Diversity in Oral Microbiota Is Associated With Periodontitis. *BMC Genomics* 18 (1), 1–15. doi: 10.1186/s12864-016-3254-5

Albert, A. Y., Chaban, B., Wagner, E. C., Schellenberg, J. J., Links, M. G., Van Schalkwyk, J., et al. (2015). A Study of the Vaginal Microbiome in Healthy Canadian Women Utilizing Cpn60-Based Molecular Profiling Reveals Distinct Gardnerella Subgroup Community State Types. *PloS One* 10 (8), e0135620. doi: 10.1371/journal.pone.0135620

Al-hebshi, N. N., Abdulhaq, A., Albarrag, A., Basode, V. K., and Chen, T. (2016). Species-Level Core Oral Bacteriome Identified by 16S Rrna Pyrosequencing in a Healthy Young Arab Population. *J. Oral. Microbiol.* 8 (1), 31444. doi: 10.3402/jom.v8.31444

Balan, P., Chong, Y. S., Umashankar, S., Swarup, S., Loke, W. M., Lopez, V., et al. (2018). Keystone Species in Pregnancy Gingivitis: A Snapshot of Oral Microbiome During Pregnancy and Postpartum Period. *Front. Microbiol.* 9, 2360. doi: 10.3389/fmicb.2018.02360

Banerjee, S., Schlaeppi, K., and van der Heijden, M. G. (2018). Keystone Taxa as Drivers of Microbiome Structure and Functioning. *Nat. Rev. Microbiol.* 16 (9), 567–576. doi: 10.1038/s41579-018-0024-1

Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M.-C. C., Charles, T., et al. (2020). Microbiome Definition Re-Visited: Old Concepts and New Challenges. *Microbiome* 8 (1), 1–22. doi: 10.1186/s40168-020-00875-0

Brawner, K., Kumar, R., Serrano, C., Ptacek, T., Lefkowitz, E., Morrow, C., et al. (2017). Helicobacter Pylori Infection Is Associated With an Altered Gastric Microbiota in Children. *Mucosal Immunol.* 10 (5), 1169–1177. doi: 10.1038/mi.2016.131

Brooks, J. P., Edwards, D. J., Harwich, M. D., Rivera, M. C., Fettweis, J. M., Serrano, M. G., et al. (2015). The Truth About Metagenomics: Quantifying and Counteracting Bias in 16S rRNA Studies. *BMC Microbiol.* 15 (1), 66. doi: 10.1186/s12866-015-0351-6

Bry, L., Brigl, M., and Brenner, M. B. (2006). Cd4+-T-Cell Effector Functions and Costimulatory Requirements Essential for Surviving Mucosal Infection With Citrobacter Rodentium. *Infect. Immun.* 74 (1), 673–681. doi: 10.1128/IAI.74.1.673-681.2006

Buermans, H., and Den Dunnen, J. (2014). Next Generation Sequencing Technology: Advances and Applications. *Biochim. Biophys. Acta (BBA)-Mol. Basis Dis.* 1842 (10), 1932–1941. doi: 10.1016/j.bbadis.2014.06.015

Burmistrz, M., Dudek, B., Staniec, D., Rodriguez Martinez, J. I., Bochtler, M., Potempa, J., et al. (2015). Functional Analysis of Porphyromonas Gingivalis W83 Crispr-Cas Systems. *J. Bacteriol.* 197 (16), 2631–2641. doi: 10.1128/jb.00261-15

Burne, R., Zeng, L., Ahn, S., Palmer, S., Liu, Y., Lefebure, T., et al. (2012). Progress Dissecting the Oral Microbiome in Caries and Health. *Adv. Dental Res.* 24 (2), 77–80. doi: 10.1177/0022034512449462

Camelo-Castillo, A., Henares, D., Brotons, P., Galiana, A., Rodríguez, J. C., Mira, A., et al. (2019). Nasopharyngeal Microbiota in Children With Invasive Pneumococcal Disease: Identification of Bacteria With Potential Disease-Promoting and Protective Effects. *Front. Microbiol.* 10, 11. doi: 10.3389/fmicb.2019.00011

Camelo-Castillo, A., Novoa, L., Balsa-Castro, C., Blanco, J., Mira, A., and Tomás, I. (2015). Relationship Between Periodontitis-Associated Subgingival Microbiota and Clinical Inflammation by 16S Pyrosequencing. *J. Clin. Periodontol.* 42 (12), 1074–1082. doi: 10.1111/jcpe.12470

Caporaso, J. G., Lauber, C. L., Costello, E. K., Berg-Lyons, D., Gonzalez, A., Stombaugh, J., et al. (2011). Moving Pictures of the Human Microbiome. *Genome Biol.* 12 (5), 1–8. doi: 10.1186/gb-2011-12-5-r50

Casadevall, A., and Pirofski, X.-a. (2003). The Damage-Response Framework of Microbial Pathogenesis. *Nat. Rev. Microbiol.* 1 (1), 17–24. doi: 10.1038/nrmicro732

Case, R. J., Boucher, Y., Dahllöf, I., Holmström, C., Doolittle, W. F., and Kjelleberg, S. (2007). Use of 16S rRNA and Rpob Genes as Molecular Markers for Microbial Ecology Studies. *Appl. Environ. Microbiol.* 73 (1), 278–288. doi: 10.1128/AEM.01177-06

Centanni, M., Lawley, B., Butts, C. A., Roy, N. C., Lee, J., Kelly, W. J., et al. (2018). Bifidobacterium Pseudolongum in the Ceca of Rats Fed Hi-Maize Starch Has Characteristics of a Keystone Species in Bifidobacterial Blooms. *Appl. Environ. Microbiol.* 84 (15), 11. doi: 10.1128/aem.00547-18

Claussen, J. C., Skiecevičienė, J., Wang, J., Rausch, P., Karlsen, T. H., Lieb, W., et al. (2017). Boolean Analysis Reveals Systematic Interactions Among Low-Abundance Species in the Human Gut Microbiome. *PloS Comput. Biol.* 13 (6), e1005361. doi: 10.1371/journal.pcbi.1005361

Dame-Teixeira, N., de Cena, J. A., Côrtes, D. A., Belmok, A., dos Anjos Borges, L. G., Marconatto, L., et al. (2020). Presence of Archaea in Dental Caries Biofilms. *Arch. Oral. Biol.* 110, 104606. doi: 10.1016/j.archoralbio.2019.104606

Das, B., Ghosh, T. S., Kedia, S., Rampal, R., Saxena, S., Bag, S., et al. (2018). Analysis of the Gut Microbiome of Rural and Urban Healthy Indians Living in Sea Level and High Altitude Areas. *Sci. Rep.* 8 (1), 1–15. doi: 10.1038/s41598-018-28550-3

Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A., and Callahan, B. J. (2018). Simple Statistical Identification and Removal of Contaminant Sequences in Marker-Gene and Metagenomics Data. *Microbiome* 6 (1), 226. doi: 10.1186/s40168-018-0605-2

de Goffau, M. C., Luopajärvi, K., Knip, M., Ilonen, J., Ruohtula, T., Härkönen, T., et al. (2013). Fecal Microbiota Composition Differs Between Children With β-Cell Autoimmunity and Those Without. *Diabetes* 62 (4), 1238–1244. doi: 10.2337/db12-0526

Dobbler, P. T., Procianoy, R. S., Mai, V., Silveira, R. C., Corso, A. L., Rojas, B. S., et al. (2017). Low Microbial Diversity and Abnormal Microbial Succession Is Associated With Necrotizing Enterocolitis in Preterm Infants. *Front. Microbiol.* 8, 2243. doi: 10.3389/fmicb.2017.02243

Ekblom, R., and Galindo, J. (2011). Applications of Next Generation Sequencing in Molecular Ecology of Non-Model Organisms. *Heredity* 107 (1), 1–15. doi: 10.1038/hdy.2010.152

Feng, Y., Chen, C.-L., Chen, T.-H., Liang, Y.-H., Chen, H.-L., Lin, C.-Y., et al. (2015). Application of Next-Generation Sequencing to Study Ascitic

Microbiome in Cirrhotic Patients With or Without Spontaneous Bacterial Peritonitis. *J. Microbiol. Immunol. Infect.* 48 (5), 504–509. doi: 10.1016/j.jmii.2014.07.005

Fisher, C. K., and Mehta, P. (2014). Identifying Keystone Species in the Human Gut Microbiome From Metagenomic Timeseries Using Sparse Linear Regression. *PloS One* 9 (7), 10. doi: 10.1371/journal.pone.0102451

Garrett, W. S., Gallini, C. A., Yatsunenko, T., Michaud, M., DuBois, A., Delaney, M. L., et al. (2010). Enterobacteriaceae Act in Concert With the Gut Microbiota to Induce Spontaneous and Maternally Transmitted Colitis. *Cell Host Microbe* 8 (3), 292–300. doi: 10.1016/j.chom.2010.08.004

Ghannoum, M. A., Jurevic, R. J., Mukherjee, P. K., Cui, F., Sikaroodi, M., Naqvi, A., et al. (2010). Characterization of the Oral Fungal Microbiome (Mycobiome) in Healthy Individuals. *PloS Pathog.* 6 (1), e1000713. doi: 10.1371/journal.ppat.1000713

Hajishengallis, G., Darveau, R. P., and Curtis, M. A. (2012). The Keystone-Pathogen Hypothesis. *Nat. Rev. Microbiol.* 10 (10), 717–725. doi: 10.1038/nrmicro2873

Hajishengallis, G., and Lamont, R. J. (2016). Dancing With the Stars: How Choreographed Bacterial Interactions Dictate Nososymbiocity and Give Rise to Keystone Pathogens, Accessory Pathogens, and Pathobionts. *Trends Microbiol.* 24 (6), 477–489. doi: 10.1016/j.tim.2016.02.010

Hamady, M., and Knight, R. (2009). Microbial Community Profiling for Human Microbiome Projects: Tools, Techniques, and Challenges. *Genome Res.* 19 (7), 1141–1152. doi: 10.1101/gr.085464.108

Hauser, L. J., Feazel, L. M., Ir, D., Fang, R., Wagner, B. D., Robertson, C. E, et al. (2015). Sinus Culture Poorly Predicts Resident Microbiota. *Int. Forum Allergy Rhinol.* Wiley Online Library. 5, 3–9. doi: 10.1002/alr.21428

Heisel, T., Podgorski, H., Staley, C. M., Knights, D., Sadowsky, M. J., and Gale, C. A. (2015). Complementary Amplicon-Based Genomic Approaches for the Study of Fungal Communities in Humans. *PloS One* 10 (2), e0116705. doi: 10.1371/journal.pone.0116705

Holt, S. C., and Ebersole, J. L. (2005). Porphyromonas Gingivalis, Treponema Denticola, and Tannerella Forsythia: The 'Red Complex', a Prototype Polybacterial Pathogenic Consortium in Periodontitis. *Periodontol. 2000* 38 (1), 72–122. doi: 10.1111/j.1600-0757.2005.00113.x

Iebba, V., Guerrieri, F., Di Gregorio, V., Levrero, M., Gagliardi, A., Santangelo, F., et al. (2018). Combining Amplicon Sequencing and Metabolomics in Cirrhotic Patients Highlights Distinctive Microbiota Features Involved in Bacterial Translocation, Systemic Inflammation and Hepatic Encephalopathy. *Sci. Rep.* 8 (1), 1–14. doi: 10.1038/s41598-018-26509-y

Kang, J. Y., Kim, H.-N., Chang, Y., Yun, Y., Ryu, S., Shin, H., et al. (2017). Gut Microbiota and Physiologic Bowel 18 F-FDG Uptake. *EJNMMI Res.* 7 (1), 72. doi: 10.1186/s13550-017-0318-8

Keegan, K. P., Trimble, W. L., Wilkening, J., Wilke, A., Harrison, T., D'Souza, M., et al. (2012). A Platform-Independent Method for Detecting Errors in Metagenomic Sequencing Data: Drisee. *PloS Comput. Biol.* 8 (6), e1002541. doi: 10.1371/journal.pcbi.1002541

Kostic, A. D., Chun, E., Robertson, L., Glickman, J. N., Gallini, C. A., Michaud, M., et al. (2013). Fusobacterium Nucleatum Potentiates Intestinal Tumorigenesis and Modulates the Tumor-Immune Microenvironment. *Cell Host Microbe* 14 (2), 207–215. doi: 10.1016/j.chom.2013.07.007

Kowalska-Duplaga, K., Gosiewski, T., Kapusta, P., Sroka-Oleksiak, A., Wędrychowicz, A., Pieczarkowski, S., et al. (2019). Differences in the Intestinal Microbiome of Healthy Children and Patients With Newly Diagnosed Crohn's Disease. *Sci. Rep.* 9 (1), 1–11. doi: 10.1038/s41598-019-55290-9

Lau, J. T., Whelan, F. J., Herath, I., Lee, C. H., Collins, S. M., Bercik, P., et al. (2016). Capturing the Diversity of the Human Gut Microbiota Through Culture-Enriched Molecular Profiling. *Genome Med.* 8 (1), 72. doi: 10.1186/s13073-016-0327-7

Lazarevic, V., Gaïa, N., Girard, M., and Schrenzel, J. (2016). Decontamination of 16S Rrna Gene Amplicon Sequence Datasets Based on Bacterial Load Assessment by Qpcr. *BMC Microbiol.* 16 (1), 73. doi: 10.1186/s12866-016-0689-4

Lepp, P. W., Brinig, M. M., Ouverney, C. C., Palm, K., Armitage, G. C., and Relman, D. A. (2004). Methanogenic Archaea and Human Periodontal Disease. *Proc. Natl. Acad. Sci. U.S.A.* 101 (16), 6176–6181. doi: 10.1073/pnas.0308766101

Li, K., Bihan, M., and Methé, B. A. (2013). Analyses of the Stability and Core Taxonomic Memberships of the Human Microbiome. *PloS One* 8 (5), e63139. doi: 10.1371/journal.pone.0063139

Ling, Z., Kong, J., Liu, F., Zhu, H., Chen, X., Wang, Y., et al. (2010). Molecular Analysis of the Diversity of Vaginal Microbiota Associated With Bacterial Vaginosis. *BMC Genomics* 11 (1), 488. doi: 10.1186/1471-2164-11-488

Liu, Z., Li, J., Liu, H., Tang, Y., Zhan, Q., Lai, W., et al. (2019). The Intestinal Microbiota Associated With Cardiac Valve Calcification Differs From That of Coronary Artery Disease. *Atherosclerosis* 284, 121–128. doi: 10.1016/j.atherosclerosis.2018.11.038

Liu, J., Wang, H., Yang, H., Zhang, Y., Wang, J., Zhao, F., et al. (2013). Composition-Based Classification of Short Metagenomic Sequences Elucidates the Landscapes of Taxonomic and Functional Enrichment of Microorganisms. *Nucleic Acids Res.* 41 (1), e3–e3. doi: 10.1093/nar/gks828

Li, Y., Wang, K., Zhang, B., Tu, Q., Yao, Y., Cui, B., et al. (2019). Salivary Mycobiome Dysbiosis and its Potential Impact on Bacteriome Shifts and Host Immunity in Oral Lichen Planus. *Int. J. Oral. Sci.* 11 (2), 1–10. doi: 10.1038/s41368-019-0045-2

Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J., et al. (2012). Performance Comparison of Benchtop High-Throughput Sequencing Platforms. *Nat. Biotechnol.* 30 (5), 434–439. doi: 10.1038/nbt.2198

Luo, C., Tsementzi, D., Kyrpides, N., Read, T., and Konstantinidis, K. T. (2012). Direct Comparisons of Illumina vs. Roche 454 Sequencing Technologies on the Same Microbial Community Dna Sample. *PloS One* 7 (2), e30087. doi: 10.1371/journal.pone.0030087

Mardis, E. R. (2013). Next-Generation Sequencing Platforms. *Annu. Rev. Anal. Chem.* 6, 287–303. doi: 10.1146/annurev-anchem-062012-092628

Muzzey, D., Evans, E. A., and Lieber, C. (2015). Understanding the Basics of NGS: From Mechanism to Variant Calling. *Curr. Genet. Med. Rep.* 3 (4), 158–165. doi: 10.1007/s40142-015-0076-8

Nakayama, J., Yamamoto, A., Palermo-Conde, L. A., Higashi, K., Sonomoto, K., Tan, J., et al. (2017). Impact of Westernized Diet on Gut Microbiota in Children on Leyte Island. *Front. Microbiol.* 8, 197. doi: 10.3389/fmicb.2017.00197

Nash, A. K., Auchtung, T. A., Wong, M. C., Smith, D. P., Gesell, J. R., Ross, M. C., et al. (2017). The Gut Mycobiome of the Human Microbiome Project Healthy Cohort. *Microbiome* 5 (1), 153. doi: 10.1186/s40168-017-0373-4

Ozkan, J., Willcox, M., Wemheuer, B., Wilcsek, G., Coroneo, M., and Thomas, T. (2019). Biogeography of the Human Ocular Microbiota. *Ocul. Surf.* 17 (1), 111–118. doi: 10.1016/j.jtos.2018.11.005

Palmer, R. J.Jr (2014). Composition and Development of Oral Bacterial Communities. *Periodontol. 2000* 64 (1), 20–39. doi: 10.1111/j.1600-0757.2012.00453.x

Perez-Chaparro, P. J., Goncalves, C., Figueiredo, L. C., Faveri, M., Lobao, E., Tamashiro, N., et al. (2014). Newly Identified Pathogens Associated With Periodontitis: A Systematic Review. *J. Dent. Res.* 93 (9), 846–858. doi: 10.1177/0022034514542468

Peterson, S. N., Snesrud, E., Liu, J., Ong, A. C., Kilian, M., Schork, N. J., et al. (2013). The Dental Plaque Microbiome in Health and Disease. *PloS One* 8 (3), e58487. doi: 10.1371/journal.pone.0058487

Power, M. E., Tilman, D., Estes, J. A., Menge, B. A., Bond, W. J., Mills, L. S., et al. (1996). Challenges in the Quest for Keystones: Identifying Keystone Species is Difficult—But Essential to Understanding How Loss of Species Will Affect Ecosystems. *BioScience* 46 (8), 609–620. doi: 10.2307/1312990

Rocas, I. N., Alves, F. R., Rachid, C. T., Lima, K. C., Assuncao, I. V., Gomes, P. N., et al. (2016). Microbiome of Deep Dentinal Caries Lesions in Teeth With Symptomatic Irreversible Pulpitis. *PloS One* 11 (5), e0154653. doi: 10.1371/journal.pone.0154653

Rubinstein, M. R., Wang, X., Liu, W., Hao, Y., Cai, G., and Han, Y. W. (2013). Fusobacterium Nucleatum Promotes Colorectal Carcinogenesis by Modulating E-Cadherin/β-Catenin Signaling Via Its FadA Adhesin. *Cell Host Microbe* 14 (2), 195–206. doi: 10.1016/j.chom.2013.07.012

Sakwinska, O., Moine, D., Delley, M., Combremont, S., Rezzonico, E., Descombes, P., et al. (2016). Microbiota in Breast Milk of Chinese Lactating Mothers. *PloS One* 11 (8), e0160856. doi: 10.1371/journal.pone.0160856

Sears, C. L., and Pardoll, D. M. (2011). Perspective: Alpha-Bugs, Their Microbial Partners, and the Link to Colon Cancer. *J. Infect. Dis.* 203 (3), 306–311. doi: 10.1093/jinfdis/jiq061

Shendure, J., and Ji, H. (2008). Next-Generation DNA Sequencing. *Nat. Biotechnol.* 26 (10), 1135. doi: 10.1038/nbt1486

Simón-Soro, A., Guillen-Navarro, M., and Mira, A. (2014). Metatranscriptomics Reveals Overall Active Bacterial Composition in Caries Lesions. *J. Oral. Microbiol.* 6 (1), 25443. doi: 10.3402/jom.v6.25443

Singh, P., and Manning, S. D. (2016). Impact of Age and Sex on the Composition and Abundance of the Intestinal Microbiota in Individuals With and Without Enteric Infections. *Ann. Epidemiol.* 26 (5), 380–385. doi: 10.1016/j.annepidem.2016.03.007

Son, J. S., Zheng, L. J., Rowehl, L. M., Tian, X., Zhang, Y., Zhu, W., et al. (2015). Comparison of Fecal Microbiota in Children With Autism Spectrum Disorders and Neurotypical Siblings in the Simons Simplex Collection. *PloS One* 10 (10), e0137725. doi: 10.1371/journal.pone.0137725

Srivastava, A. (2011). *Evolution & Detection of non-Coding RNA, and Transcriptome Analyses of Two non-Model Systems* (University of Georgia).

Stobernack, T. (2019). *Porphyromonas Gingivalis–an Oral Keystone Pathogen Challenging the Human Immune System* (University of Groningen).

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., et al. (2018). Prisma Extension for Scoping Reviews (Prisma-Scr): Checklist and Explanation. *Ann. Intern. Med.* 169 (7), 467–473. doi: 10.7326/M18-0850

Trosvik, P., and de Muinck, E. J. (2015). Ecology of Bacteria in the Human Gastrointestinal Tract-Identification of Keystone and Foundation Taxa. *Microbiome* 3, 1–12. doi: 10.1186/s40168-015-0107-4

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., and Gordon, J. I. (2007). The Human Microbiome Project. *Nature* 449 (7164), 804–810. doi: 10.1038/nature06244

Wang, B., Yao, M., Lv, L., Ling, Z., and Li, L. (2017). The Human Microbiota in Health and Disease. *Engineering* 3 (1), 71–82. doi: 10.1016/J.ENG.2017.01.008

Wang, Y., Zhang, J., Chen, X., Jiang, W., Wang, S., Xu, L., et al. (2017). Profiling of Oral Microbiota in Early Childhood Caries Using Single-Molecule Real-Time Sequencing. *Front. Microbiol.* 8, 2244. doi: 10.3389/fmicb.2017.02244

Wan, Y.-D., Zhu, R.-X., Wu, Z.-Q., Lyu, S.-Y., Zhao, L.-X., Du, Z.-J., et al. (2018). Gut Microbiota Disruption in Septic Shock Patients: A Pilot Study. *Med. Sci. Monit.: Int. Med. J. Exp. Clin. Res.* 24, 8639. doi: 10.12659/MSM.911768

White, J. R., Nagarajan, N., and Pop, M. (2009). Statistical Methods for Detecting Differentially Abundant Features in Clinical Metagenomic Samples. *PloS Comput. Biol.* 5 (4), e1000352. doi: 10.1371/journal.pcbi.1000352

Winand, R., Bogaerts, B., Hoffman, S., Lefevre, L., Delvoye, M., Van Braekel, J., et al. (2020). Targeting the 16s Rrna Gene for Bacterial Identification in Complex Mixed Samples: Comparative Evaluation of Second (Illumina) and Third (Oxford Nanopore Technologies) Generation Sequencing Technologies. *Int. J. Mol. Sci.* 21 (1), 298. doi: 10.3390/ijms21010298

Wu, L., Zeng, T., Zinellu, A., Rubino, S., Kelvin, D. J., and Carru, C. (2019). A Cross-Sectional Study of Compositional and Functional Profiles of Gut Microbiota in Sardinian Centenarians. *MSystems* 4 (4), e00325–e00319. doi: 10.1128/mSystems.00325-19

Xiao, J., Fiscella, K. A., and Gill, S. R. (2020). Oral Microbiome: Possible Harbinger for Children's Health. *Int. J. Oral. Sci.* 12 (1), 1–13. doi: 10.1038/s41368-020-0082-x

Zakrzewski, M., Simms, L. A., Brown, A., Appleyard, M., Irwin, J., Waddell, N., et al. (2019). Il23r-Protective Coding Variant Promotes Beneficial Bacteria and Diversity in the Ileal Microbiome in Healthy Individuals Without Inflammatory Bowel Disease. *J. Crohn's Colitis* 13 (4), 451–461. doi: 10.1093/ecco-jcc/jjy188

Zeng, X., Gao, X., Peng, Y., Wu, Q., Zhu, J., Tan, C., et al. (2019). Higher Risk of Stroke Is Correlated With Increased Opportunistic Pathogen Load and Reduced Levels of Butyrate-Producing Bacteria in the Gut. *Front. Cell. Infect. Microbiol.* 9, 4. doi: 10.3389/fcimb.2019.00004

Zerón, A. (2014). Genoma, Microbioma Y Epigenoma Humano. Una Visión Contemporánea De La Tríada Ecológica. *Rev. ADM* 71 (4), 161–172.

Zhang, C., Cleveland, K., Schnoll-Sussman, F., McClure, B., Bigg, M., Thakkar, P., et al. (2015). Identification of Low Abundance Microbiome in Clinical Samples Using Whole Genome Sequencing. *Genome Biol.* 16 (1), 265. doi: 10.1186/s13059-015-0821-z

Zhang, C., Thakkar, P. V., Powell, S. E., Sharma, P., Vennelaganti, S., Betel, D., et al. (2019). A Comparison of Homogenization vs. Enzymatic Lysis for Microbiome Profiling in Clinical Endoscopic Biopsy Tissue Samples. *Front. Microbiol.* 9, 3246. doi: 10.3389/fmicb.2018.03246

Zhang, Y., Wang, X., Li, H., Ni, C., Du, Z., and Yan, F. (2018). Human Oral Microbiota and its Modulation for Oral Health. *BioMed. Pharmacother.* 99, 883–893. doi: 10.1016/j.biopha.2018.01.146

Zhu, H., Liu, Y., Li, S., Jin, Y., Zhao, L., Zhao, F., et al. (2018). Altered Gut Microbiota After Traumatic Splenectomy Is Associated With Endotoxemia. *Emerg. Microbes Infect.* 7 (1), 1–10. doi: 10.1038/s41426-018-0202-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cellular and Infection Microbiology

ORIGINAL RESEARCH
published: 17 June 2021
doi: 10.3389/fcimb.2021.663756

Structure and Function of Oral Microbial Community in Periodontitis Based on Integrated Data

Zhengwen Cai[1,2], *Shulan Lin*[1,2,3], *Shoushan Hu*[1,2] *and Lei Zhao*[1,2,3]*

[1] State Key Laboratory of Oral Diseases, West China College of Stomatology, Sichuan University, Chengdu, China, [2] National Clinical Research Center for Oral Diseases, West China College of Stomatology, Sichuan University, Chengdu, China, [3] Department of Periodontics, West China Hospital of Stomatology, Sichuan University, Chengdu, China

Objective: Microorganisms play a key role in the initiation and progression of periodontal disease. Research studies have focused on seeking specific microorganisms for diagnosing and monitoring the outcome of periodontitis treatment. Large samples may help to discover novel potential biomarkers and capture the common characteristics among different periodontitis patients. This study examines how to screen and merge high-quality periodontitis-related sequence datasets from several similar projects to analyze and mine the potential information comprehensively.

Methods: In all, 943 subgingival samples from nine publications were included based on predetermined screening criteria. A uniform pipeline (QIIME2) was applied to clean the raw sequence datasets and merge them together. Microbial structure, biomarkers, and correlation network were explored between periodontitis and healthy individuals. The microbiota patterns at different periodontal pocket depths were described. Additionally, potential microbial functions and metabolic pathways were predicted using PICRUSt to assess the differences between health and periodontitis.

Results: The subgingival microbial communities and functions in subjects with periodontitis were significantly different from those in healthy subjects. *Treponema, TG5*, *Desulfobulbus*, *Catonella*, *Bacteroides*, *Aggregatibacter*, *Peptostreptococcus*, and *Eikenella* were periodontitis biomarkers, while *Veillonella*, *Corynebacterium*, *Neisseria*, *Rothia*, *Paludibacter*, *Capnocytophaga*, and *Kingella* were signature of healthy periodontium. With the variation of pocket depth from shallow to deep pocket, the proportion of Spirochaetes, Bacteroidetes, TM7, and Fusobacteria increased, whereas that of Proteobacteria and Actinobacteria decreased. Synergistic relationships were observed among different pathobionts and negative relationships were noted between periodontal pathobionts and healthy microbiota.

Conclusion: This study shows significant differences in the oral microbial community and potential metabolic pathways between the periodontitis and healthy groups.

OPEN ACCESS

Edited by:
Thuy Do,
University of Leeds, United Kingdom

Reviewed by:
Marcelo Freire,
J. Craig Venter Institute (La Jolla), United States
Daniel Hagenfeld,
University Hospital of Münster, Germany

***Correspondence:**
Lei Zhao
jollyzldoc@163.com

Specialty section:
This article was submitted to Microbiome in Health and Disease, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: *03 February 2021*
Accepted: *31 May 2021*
Published: *17 June 2021*

Citation:
Cai Z, Lin S, Hu S and Zhao L (2021) Structure and Function of Oral Microbial Community in Periodontitis Based on Integrated Data. Front. Cell. Infect. Microbiol. 11:663756. doi: 10.3389/fcimb.2021.663756

Our integrated analysis provides potential biomarkers and directions for in-depth research. Moreover, a new method for integrating similar sequence data is shown here that can be applied to other microbial-related areas.

Keywords: 16S, periodontitis, bacteria, microbiome, metabolite, biomarker, high-throughput nucleotide sequencing

INTRODUCTION

Periodontitis is an inflammatory condition affecting periodontal tissue and is the result of uncontrolled gingivitis (Chapple et al., 2018). The pathogenesis of periodontitis remains unclear, and the diagnosis of periodontitis heavily depends on the clinical manifestation and periodontal detection indicators (periodontal probing depth, clinical attachment loss, bleeding on probing, and alveolar bone loss) (Kinane et al., 2017; Papapanou et al., 2018). At present, it is widely acknowledged that dental plaque is the key initiating factor. Destruction of the periodontium is attributed to microbial dysbiosis as well as the excessive immune response of the hosts (Kinane et al., 2017). The transition from healthy gingiva to periodontitis occurs because of the accumulation of pathogenic microorganisms and is affected by multiple risk factors such as modifiable habits and immutable genetic predisposition, eventually leading to oral ecological disturbance (Genco and Borgnakke, 2013; Papapanou et al., 2018). However, no specific pathobionts have yet been identified. This may be because we have overlooked specific pathobionts that are low in number, or that an imbalanced microbiota is the underlying cause, rather than specific pathobionts. For instance, the red complex in subgingival plaque (*Porphyromonas gingival*, *Treponema denticola*, and *Tannerella forsythia*) play an important role in periodontal dysbiosis (Socransky et al., 1998).

Previous studies have analyzed the correlation between microorganisms and periodontitis. Nevertheless, *in vitro* bacterial cultures and polymerase chain reaction (PCR)-based analysis alone limit the ability to observe the complex profiles of subgingival microbiota. Next-generation sequencing technology has been widely used in microecology, which has the ability to present the whole appearance of microbiota and enable researchers to explore multiple dimensions of microbial communities (Gu et al., 2019). Combined with conserved 16S ribosomal RNA targeted assays, the external interference by host genes can be eliminated, and it can accurately identify and quantify various microorganisms (Cummings et al., 2016). The evolutionary and taxonomic relationships among microorganisms can also be evaluated, in addition to predicting their characteristic properties and metabolic pathways (Langille et al., 2013).

Several studies have been conducted on human oral microbiota in subjects with healthy periodontal tissue, gingivitis, and different states of periodontitis. Samples are collected from different sites including the supragingival area (Galimanas et al., 2014), subgingival plaque (Tsai et al., 2018), saliva (Chen et al., 2018), and gingival crevicular fluid (Pei et al., 2020). The current trend is to explore target samples to identify sensitive pathogenic biomarkers by microbiology, metabolomics, or multiomics (Su et al., 2020). However, studies in a small sample can be disturbed by diversified confounders that may lead to biased conclusions. Therefore, there should be a way to use existing data to realize the hidden information mining, aimed at identifying the common microbial characteristics of periodontitis. To this end, we screened high-quality sequence datasets, merged, and processed data under a unified standard. Finally, we analyzed oral microbial communities to identify the common characteristic microorganisms as well as the functions and potential metabolic pathways in different periodontitis patients.

MATERIALS AND METHODS

Microbiome Data Source Collection and Eligibility Criteria

All included microbial datasets were retrieved according to the predetermined design. Relevant studies and data were collected by December 30, 2020, through electronic databases: (1) The National Library of Medicine (MEDLINE by PubMed) was searched using the following keywords: ((((periodontitis) OR (chronic periodontitis)) OR (aggressive periodontitis)) OR (periodontal disease)) AND (16S) AND (subgingival). (2) The Genomes of National Center for Biotechnology Information was searched using the following search strategy: 16S [All Fields] AND periodontitis [All Fields] AND subgingival [All Fields]. The search results yielded 919 publications. All abstracts were browsed and selected by two authors (Z.W. Cai and S.L. Lin) to remove duplicate and non-clinical original articles. Only subgingival microbial sequencing-related studies were retained. The full text of 38 studies was obtained. We searched for the details including study design information, inclusion criteria of participants, and the results of clinical periodontal indicators, to assess whether the inclusion criteria and the results of clinical periodontal indices were consistent with our predetermined criteria, which were based on the latest classification of periodontal disease (Chapple et al., 2018; Papapanou et al., 2018). The inclusion criteria for the periodontitis group were as follows: (1) periodontal probing depth (PPD) >4 mm, (2) clinical attachment loss (CAL) >3 mm, and (3) bleeding on probing (BOP) at >10% of sites. The inclusion criteria for healthy individuals were as follows: (1) without PPD or PPD <4 mm, (2) mean CAL <2 mm, and (3) mean BOP <10%. The exclusion criteria for all subjects were (1) pregnancy or systemic disease, (2) periodontal therapy sought within the past 3 months, and (3) use of antibiotics in the past 3 months. Twenty-two studies were eliminated because of unavailable sequence datasets or

because they did not meet our predetermined criteria (**Supplementary Material 1**). Sixteen articles were included and their datasets downloaded from NCBI or ENA.

Raw Sequence Processing and Re-Filter

The 16 raw sequence datasets were merged and processed with a uniform standard *via* QIIME2 pipeline version 2020.8 using default parameters (Bolyen et al., 2019). DADA2 (Wolf and Evans, 2018) was used to denoise the data and assess the sequence quality score (QS). The parameters for trimming and truncation settings were 10 and 150, respectively. Samples with overall quality <25 were eliminated. Then, the amplicon sequence variants (ASVs, obtained from DADA2) sequence data were clustered into operational taxonomic units (OTUs) at 97% similarity in closed-reference of the Greengenes database gg-13-8 version (DeSantis et al., 2006), and the Feature Table was created. Through this step, we matched the different short reads with the representative sequence. The short reads without a matching representative sequence in the library were eliminated, and the representative sequence taxa were classified with a trained Naïve Bayesian classifier for their annotations (Wang et al., 2007). Next, we re-filtered the Feature Table to eliminate samples with microbial features <4 or the frequency of microorganism <1000. Finally, seven more publications were eliminated because they were not suited for further analysis (**Supplementary Material 1**); nine datasets were included in the final analysis (**Table 1**). The risk of bias in the included studies was assessed according to the Downs-Black checklist (**Supplementary Material 2**) (Downs and Black, 1998). Two authors (Z.W. Cai and S.S. Hu) independently assessed the risk of bias, and any disagreement was resolved through consultation with the third author (L. Zhao). The quality level of eight studies was fair, and one study was of good quality. The average score of the nine articles was 16.25, and the equality level was fair. **Figure 1** shows the flow diagram of the process of literature selection.

Datasets for Analysis

The datasets from nine publications (Griffen et al., 2012; Galimanas et al., 2014; Bizzarro et al., 2016; Califf et al., 2017; Chen et al., 2018; Pérez-Chaparro et al., 2018; Wei et al., 2019; Liu et al., 2020; Shi et al., 2020) were divided into two groups—periodontal disease (PD) and healthy control (HC)—according to our predetermined criteria. The mean age of the HC group was 34.4 ± 7.1 years, and the mean age of the PD group was 43.7 ± 8.2 years. Some samples collected from different sites (such as supragingival dental plaque, tongue, and saliva) were excluded. Only subgingival samples were included for the analysis, which could better demonstrate the relationship between microorganisms and periodontitis. Other irrelevant confounding factors were also eliminated as much as possible: data of patients with periodontitis who were post-treated with sodium hypochlorite in Califf's study were deleted. The patients treated with antibiotics in Bizzarro's study or post-treated with no-surgical periodontal treatment in Chen's study were removed. The healthy group in Griffen's study was excluded because some participants' BOP was >10%. Moreover, three studies (Bizzarro et al., 2016; Califf et al., 2017; Pérez-Chaparro et al., 2018) with definite parameters of periodontal pocket depth were selected to explore the relationship between clinical periodontal pocket depth and the variation of microbiota. The other six studies were not included because there were no specific pocket depth parameters supported in the original metadata. The subgingival samples were divided into the following four groups based on the periodontal pocket depth: 0–3 mm (healthy control); 3–4 mm; 5–6 mm; and 7–9 mm.

Statistics Analysis

The microbial structure was evaluated by alpha diversity (Shannon diversity index, Observed features vector, Simpson diversity index, and Chao1 index) and beta diversity (Jaccard distance, Bray-Curtis distance, unweighted UniFrac distance, and weighted UniFrac distance matrix) based on a rarifying sample depth of 1000 in QIIME2 (Bolyen et al., 2019). Beta diversity was tested by multivariate homogeneity of dispersions (PERMDISP) (Anderson, 2006) and permutational multivariate analysis of variance (PERMANOVA) (Anderson, 2001) to test the homogenous dispersion and variance between groups. Microbial composition analysis was carried out on the Microbiome-Analyst platform (Chong et al., 2020). Default parameters were used to preprocess the data, which included a count filter and a variance filter. The count filter removed the samples in small number, and the variance filter deleted the constant features in each group.

TABLE 1 | Summary of the studies included in pooled analysis.

Author	Accession	Sample-source	Region	Description (Number of participants)	
				HC	PD
Califf et al.	PRJEB19122	Sub, Supra	V4V5	–	34
Galimanas et al.	PRJEB6047	Sub, Supra	V3	11	13
Bizzarro et al.	PRJNA289294	Sub	V5-V7	–	37
Griffen et al.	SRP009299	Sub	V1V2/V4	–	29
Wei et al.	PRJNA509532	Sub, Buccal mucosa	V4V5	9	23
Shi et al.	SRP228020	Sub, GCF	V4	10	24
Liu et al.	SRP102224	Sub	V3V4	–	12
Pérez et al.	PRJNA324274	Sub	V3	7	9
Chen et al.	SRP075100	Sub, Saliva	V4	21	48

PD, periodontal disease; HC, healthy control; Sub, subgingival plaque; Supra, supragingival plaque; GCF, gingival crevicular fluid.
Some samples in those sequence datasets were removed due to not meeting the inclusion criteria.

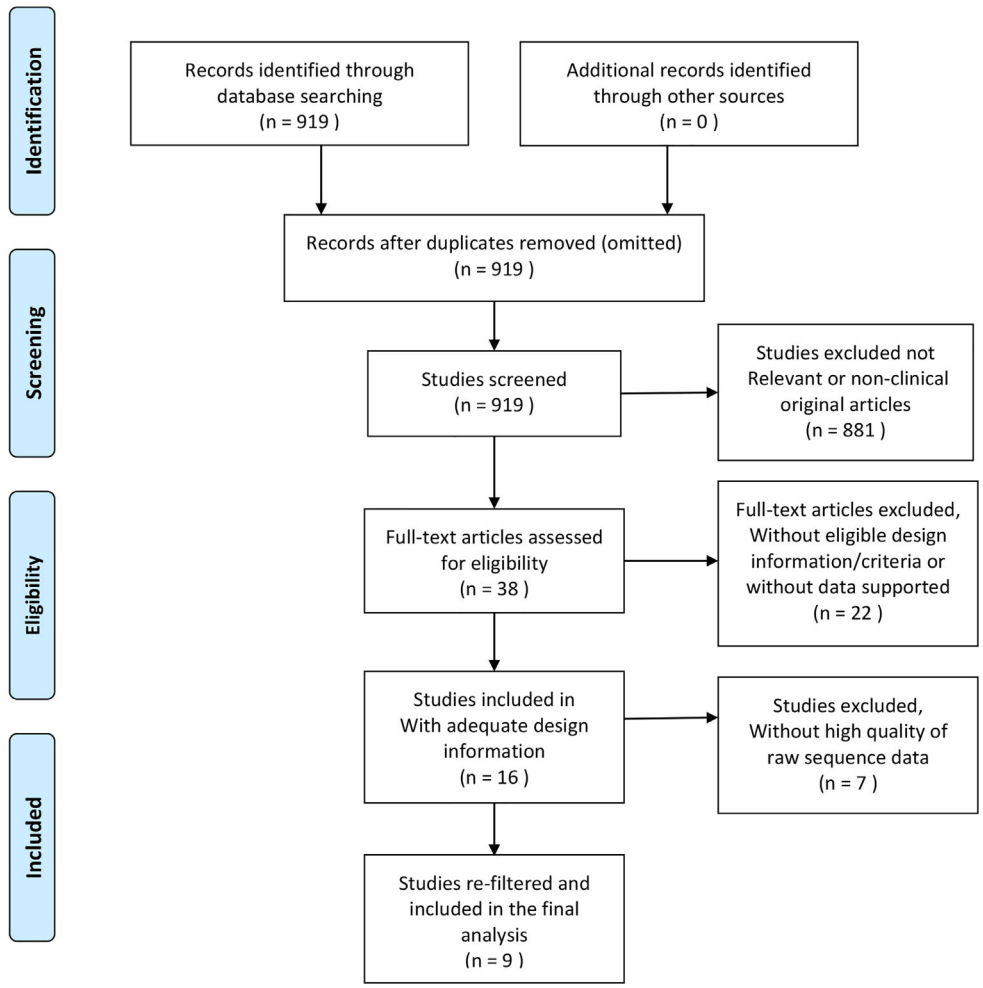


FIGURE 1 | Modified flow diagram of collecting and screening articles, processing, and re-filtering data.

Centered log ratio (CLR) transformation was performed for normalization prior to data analysis. The Mann–Whitney U test was selected as the default statistical test. Linear discriminant analysis effect size (LEfSe) was adopted for the microbial comparison analysis (significance level, $p<0.05$ and linear discriminant analysis [LDA] score >2 were chosen to characterize the phenotype) (Segata et al., 2011). Correlation network analysis used Spearman's rank correlation with the threshold set to 0.3. Microbial community function was predicted *via* PICRUSt (Langille et al., 2013) to explore the potential interactions among host, environment, and microbial community. The principle was to match the whole genome of the corresponding homologous ancestor through 16S sequencing, and then map it to metabolites as well as pathways to achieve functional prediction. The enrichment pathway analysis of the PD and HC groups was on level 2 and 3 functional dimensions based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (www.kegg.jp). RNA-seq methods (Algorithm: edgeR, adjusted p-value cut-off: <0.05) were adopted to analyze the significant difference in microbial functions and microbiota at different pocket depths (Robinson et al., 2010).

RESULTS

Oral Microbial Structure and Composition

The beta diversity (Jaccard distance matrix, $p<0.001$, PERMANOVA) demonstrated the clusters of subgingival microbial structure between the PD and HC groups (**Figure 2A**). It also showed inhomogeneous dispersion between both groups (Jaccard distance matrix, $p<0.001$, PERMDISP), which means that both location effect and dispersion effect existed. Other beta diversity analyses are presented in **Supplementary Figure 1A**. The alpha diversity Chao1 and observed index showed significant difference ($p<0.01$), whereas the Simpson and Shannon index showed no difference between HC and PD (**Supplementary Figure 1B**).

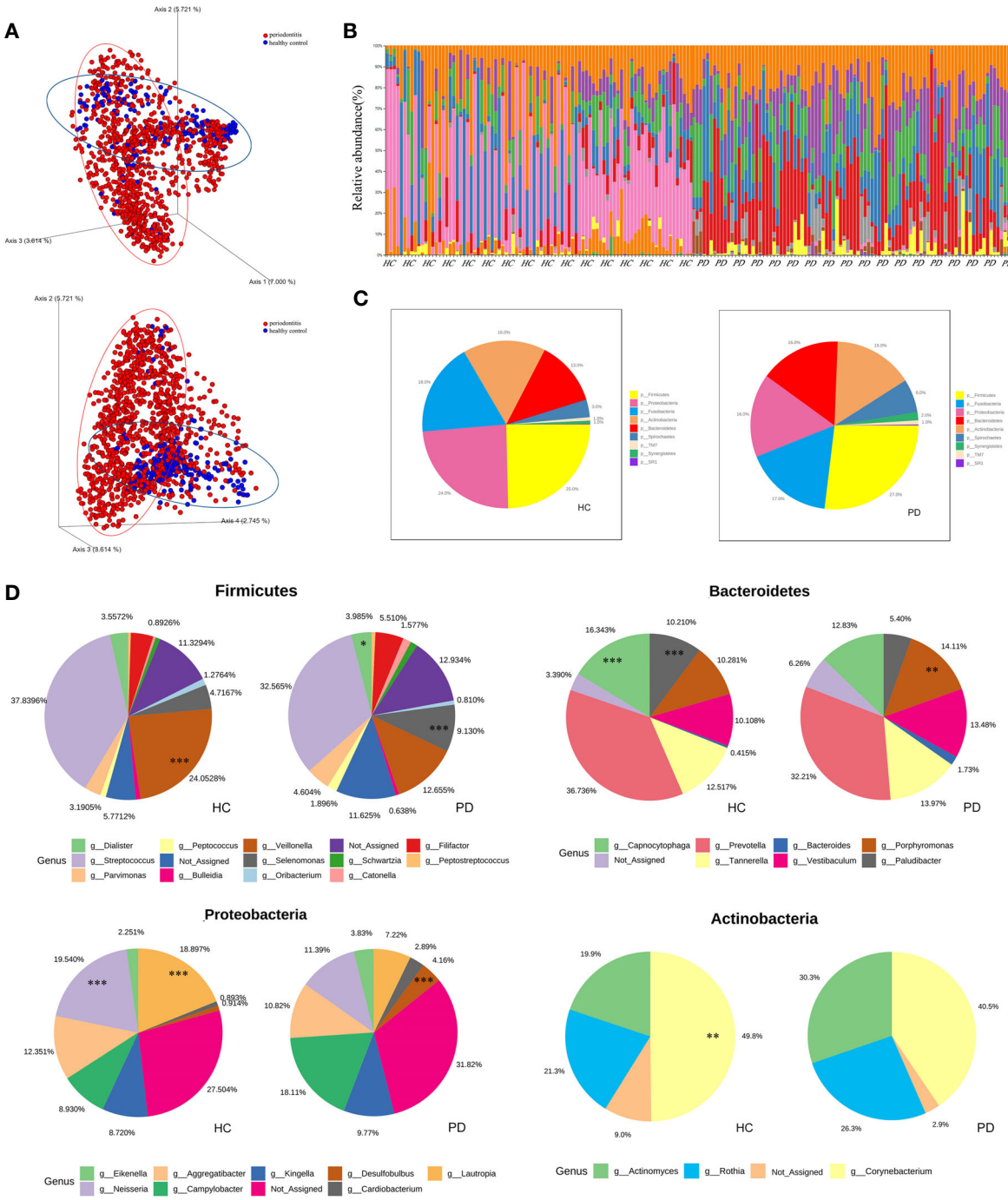


FIGURE 2 | **(A)** The 3D-PCoA plot based on Jaccard distance matrix (p<0.001, PERMANOVA) illustrates the beta diversity of oral microbiota. PD (periodontitis, red spots), HC (healthy control, blue spots). **(B)** The taxonomic bar plots on the phylum level. Each column represents a sample (HC left and PD right), and each small fragment in different colors represents different phyla. **(C)** The pie charts demonstrate the difference of microbial composition between HC (left) and PD (right) on the phylum level. Different colors correspond to the phyla on the list. **(D)** The subcategories composition of four phyla. Each pair of pie charts shows the comparison of microbial abundance between HC and PD on the genus level (*p < 0.05; **p < 0.01; ***p < 0.001, tested by RNA-seq methods, algorithm: edgeR).

A taxonomic bar plot was used to compare the profile of HC and PD groups at the phylum level (**Figure 2B**). It was observed that the microbial community in PD was more complex than that in HC, and the proportions of several phyla changed, which was regarded as a state of dysbiosis. The pie chart shows differences in microbial composition between HC and PD on the phylum level (**Figure 2C**). The dominant phyla of PD were Firmicutes (27%), Fusobacteria (17%), Proteobacteria (16%), Bacteroidetes (16%),

Actinobacteria (15%), and Spirochaetes (6%). The dominant phyla of HC were Firmicutes (25%), Proteobacteria (24%), Fusobacteria (18%), Actinobacteria (16%), Bacteroidetes (13%), and Spirochaetes (3%). To illustrate the subtle differences in the composition of HC and PD, we compared the composition of subcategories of some phyla between HC and PD separately: Firmicutes, Proteobacteria, Bacteroidetes, and Actinobacteria (**Figure 2D**). In Firmicutes, the genus *Veillonella* ($p<0.001$) increased in HC, whereas *Selenomonas* and *Dialister* ($p<0.001$) were seen more abundantly in PD. In Proteobacteria, the genus *Neisseria* ($p<0.001$) and *Lautropia* ($p<0.001$) were found abundantly in HC, while *Desulfobulbus* ($p<0.001$) was rich in PD. In Bacteroidetes, the proportion of *Porphyromonas* ($p<0.01$) increased, whereas *Capnocytophaga* ($p<0.001$) and *Paludibacter* ($p<0.001$) decreased in PD. In phylum Actinobacteria, *Corynebacterium* ($p<0.01$) increased in HC. LEfSe was used to identify significant differences in taxa. (1) On the class level: the biomarkers of PD were Bacteroidia, Spirochaetes, Synergistia, and Deltaproteobacteria, while Actinobacteria, Betaproteobacteria, and Flavobacteria were the biomarkers of HC. (2) On the genus level: *Treponema, TG5*, *Desulfobulbus*, *Catonella*, *Bacteroides*, *Aggregatibacter*, *Peptostreptococcus*, and *Eikenella* were biomarkers for periodontitis, while *Veillonella*, *Corynebacterium*, *Neisseria*, *Rothia*, *Paludibacter*, *Capnocytophaga*, and *Kingella* were biomarkers for the healthy group (**Figure 3A**).

Microbial Correlation Network

Correlation network analysis among the microbial community is shown in **Table 2**. The characteristic genera of PD, i.e., *Treponema*, *Tannerella, TG5*, *Desulfobulbus*, *Porphyromonas*, *Treponema*, and *Filifactor* were positively correlated with each other (**Figure 3B**). There were also positive correlations between normal oral bacteria and healthy-related microbes (*Kingella*, *Capnocytophaga*, *Rothia*, *Veillonella*, *Streptococcus*, and *Corynebacterium*). Negative correlations were between oral normal microbes and periodontal pathobionts. For example, *TG5*, *Treponema*, *Tannerella*, and *Desulfobulbus* were negatively correlated with *Kingella* as well as *Veillonella*. More relationships among the microbial community are presented in **Supplementary Material 3**.

Microbial Composition Changing With PPD

The taxa plot displayed the microbial abundance of different pocket depths on the phylum level. With the variation of PPD from shallow to deep pockets, the proportion of Spirochaetes, Bacteroidetes, TM7, and Fusobacteria increased, whereas Proteobacteria and Actinobacteria decreased (**Figure 4A**). The microbial composition of HCs (PPD: 0–3 mm), shallow layer (PPD: 3–4 mm) group, and deep layer (PPD: 7–9 mm) group are shown on the genus level in **Figure 4B**. Healthy controls consisted of *Neisseria*, *Streptococcus*, and some other bacteria. The shallow layer group consisted of

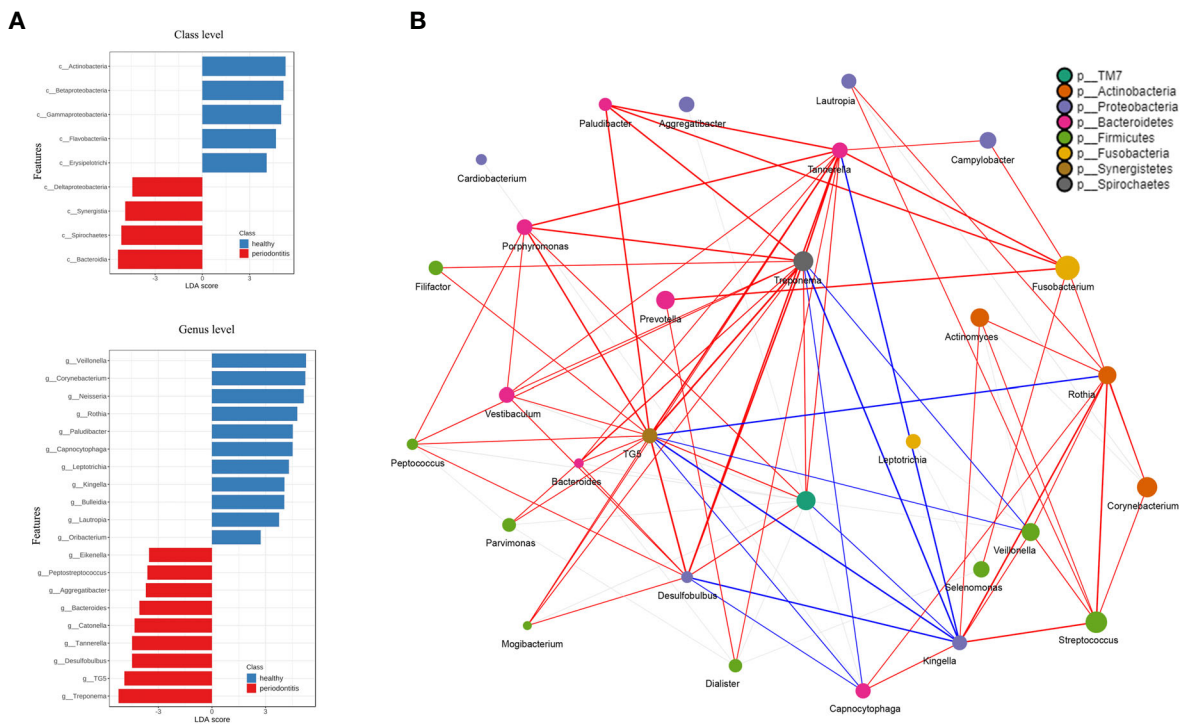


FIGURE 3 | **(A)** The ordinate is the taxa with significant differences between the groups, and the abscissa is a bar graph to visually display the LDA analysis logarithmic score value of each taxa. The longer the length, the more significant the difference in the taxon. Red bars indicate periodontitis and blue bars indicate healthy control. **(B)** The network analysis shows the correlation between microorganisms on the genus level. Each genus is colored according to its phylum. The edges show greater correlations and the node size reflects the abundance. Red and blue lines represent positive and negative correlations, respectively.

TABLE 2 | Correlation network analysis of the microbial community.

Taxon1	Taxon2	Correlation	P.value	Statistic
Treponema	*TG5*	0.7191	<0.01	32561635.83
Desulfobulbus	*TG5*	0.6455	<0.01	41094316.01
Tannerella	*Treponema*	0.5774	<0.01	48986356.52
Desulfobulbus	*Treponema*	0.5446	<0.01	52793791.74
Tannerella	*TG5*	0.5364	<0.01	53736459.74
Capnocytophaga	*Kingella*	0.5071	<0.01	57139373.92
Kingella	*Rothia*	0.4634	<0.01	62197661.63
Corynebacterium	*Rothia*	0.4582	<0.01	62806659.08
Streptococcus	*Veillonella*	0.4462	<0.01	64198665.84
Rothia	*Streptococcus*	0.4448	<0.01	64356577.13
Dialister	*Prevotella*	0.4427	<0.01	64602407.19
Desulfobulbus	*Tannerella*	0.4417	<0.01	64720393.22
Corynebacterium	*Lautropia*	0.4357	<0.01	65409187.11
Filifactor	*Treponema*	0.4325	<0.01	65785509.61
Peptococcus	*Treponema*	0.429	<0.01	66187131.79
Actinomyces	*Rothia*	0.4137	<0.01	67957682.15
Porphyromonas	*Tannerella*	0.4076	<0.01	68666527.88
Paludibacter	*Treponema*	0.4049	<0.01	68985946.96
Campylobacter	*Fusobacterium*	0.4016	<0.01	69367430.26
Desulfobulbus	*Vestibaculum*	0.3982	<0.01	69759006.29
Mogibacterium	*TG5*	0.3982	<0.01	69763542.92
Porphyromonas	*TG5*	0.393	<0.01	70363518.18
Paludibacter	*Tannerella*	0.3885	<0.01	70882356.24
Mogibacterium	*Treponema*	0.3855	<0.01	71235324.49
Fusobacterium	*Selenomonas*	0.3812	<0.01	71724014.35
Bacteroides	*TG5*	0.3793	<0.01	71948445.92
Porphyromonas	*Treponema*	0.377	<0.01	72212475.12
Desulfobulbus	*Mogibacterium*	0.3711	<0.01	72897777.71
Rothia	*TG5*	-0.3042	<0.01	151183647.29
Capnocytophaga	*Desulfobulbus*	-0.3091	<0.01	151744548.94
Desulfobulbus	*Kingella*	-0.325	<0.01	153592374.07
TG5	*Veillonella*	-0.3299	<0.01	154162465.49
Kingella	*Tannerella*	-0.3476	<0.01	156211678.52
Treponema	*Veillonella*	-0.3685	<0.01	158629770.93
Capnocytophaga	*Treponema*	-0.375	<0.01	159392440.50
Capnocytophaga	*TG5*	-0.3865	<0.01	160722624.16
Kingella	*TG5*	-0.4239	<0.01	165050919.50
Kingella	*Treponema*	-0.4479	<0.01	167842107.45

Correlation network analysis on genus level used Spearman rank correlation with the threshold set at 0.3. Only part of the correlation is presented.

Fusobacterium, Corynebacterium, *Actinomyces*, *Streptococcus*, and some other bacteria. In the deep layer (PPD: 7–9 mm) group, *Fusobacterium*, *Porphyromonas*, and *Treponema* were the dominant genera. The comparison of microorganisms in the deep layer and the shallow layer are presented in **Table 3**. In the deep layer, *Desulfobulbus*, *TG5*, *SHD_231, Tannerella*, *Porphyromonas*, and some other pathobionts increased significantly, whereas *Pseudomonas*, *Haemophilus*, *Actinomyces*, *Capnocytophaga*, and some oral normal bacteria decreased significantly. A heatmap was used to show the correlations among different taxa and PPD (**Figure 4C**). The correlation coefficient between the pathobionts (such as *Mogibacterium*, *Tannerella*, *Filifactor*, *TG5*, *Treponema*, *Desulfobulbus*, and *Peptostreptococcus*) and the deep periodontal pockets is higher than the correlation coefficient between these pathobionts and the shallow pocket depth. These pathobionts illustrated an increasing trend with the deepening of pocket depth. In contrast, some microorganisms such as *Corynebacterium*, *Rothia*, *Kingella*, *Neisseria*, and *Haemophilus* had a higher correlation coefficient in the groups with healthy (PPD: 0–3 mm) and shallow (PPD: 3–4 mm) pocket depths. These normal microbes displayed a decreasing trend with the pocket depth.

Microbial Community Functions Analysis

The KEGG functional orthologs (KOs) were gathered to comprehensively analyze the involved enrichment pathways. The function and pathways of microbes between the PD and HC groups were compared on different functional dimensions. At level 2, cell motility, cellular processing and signaling, nucleotide metabolism, metabolism of cofactors and vitamins, and nervous system function were significantly different ($p<0.05$). At level 3, bacterial motility proteins and flagellar assembly increased in the PD group. There were some differences in amino acid metabolism (e.g., tyrosine, histidine, D-arginine, D-ornithine, and glycine) between the HC and PD groups. Synthesis and degradation of ketone bodies, nitrogen metabolism, and sulfur metabolism were also significantly different ($p<0.05$). Part of the significant functions and pathways are shown in **Table 4**. More details about the differential analysis on levels 2 and 3 are provided in **Supplementary Material 4** and **5**. The random forest model

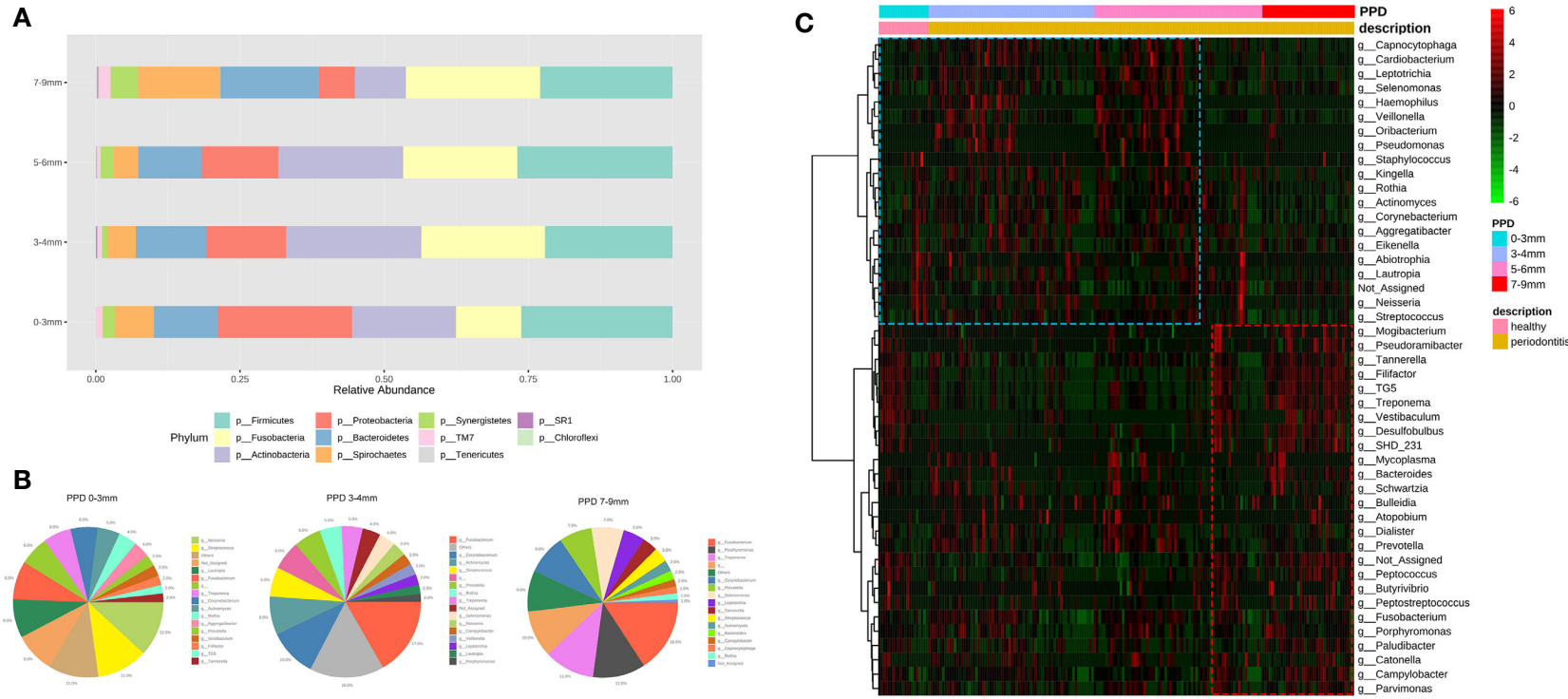


FIGURE 4 | **(A)** The taxonomic bar plots on the genus level. Each bar represents a group of different periodontal probing depth (0–3 mm, 3–4 mm, 5–6 mm, and 7–9 mm). **(B)** Three pie charts display the genera (abundance >1%) in the sites of PPD 0–3 mm, 3–4 mm, and 7–9 mm. The taxonomy "others" is a cluster of genera whose abundance are less than 1%. **(C)** The heatmap shows the correlations between taxa and PPD. Each column represents a sample, each row represents a taxon, and each lattice represents a correlation coefficient between a taxon and PPD group (red lattice, positive correlation; green lattice, negative correlation). With the deepening of pocket depth, the pathobionts increase (red box) and some normal microbes decrease (blue box).

TABLE 3 | Comparison of microorganisms in the deep layer (PPD: 7–9 mm) and the shallow layer (PPD: 3–4 mm).

Taxon	log2FC	LogCPM	P values	FDR
Desulfobulbus	3.7079	13.03	<0.001	<0.001
TG5	3.2558	14.706	<0.001	<0.001
Pseudomonas	-7.4192	16.079	<0.001	<0.001
Haemophilus	-6.0582	13.966	<0.001	<0.001
Eubacterium	2.983	10.262	<0.001	<0.001
SHD_231	3.1276	10.859	<0.001	<0.001
Filifactor	3.3358	14.779	<0.001	<0.001
Tannerella	2.5596	15.028	<0.001	<0.001
Treponema	2.4324	16.494	<0.001	<0.001
Lautropia	-4.2751	15.045	<0.001	<0.001
Actinomyces	-2.9909	16.941	<0.001	<0.001
Capnocytophaga	-3.0637	13.944	<0.001	<0.001
Streptococcus	-2.3346	16.191	<0.001	<0.001
Mogibacterium	2.1055	10.219	<0.001	<0.001
Abiotrophia	-3.1646	10.654	<0.001	<0.001
Corynebacterium	-3.0281	17.348	<0.001	<0.001
Peptococcus	2.045	11.502	<0.001	<0.001
Rothia	-3.3356	16.289	<0.001	<0.001
Vestibaculum	2.9434	14.144	<0.001	<0.001
Neisseria	-3.4407	14.256	<0.001	<0.001
Leptotrichia	-2.344	14.919	<0.001	<0.001
Kingella	-2.703	13.456	<0.001	<0.001
Peptostreptococcus	2.0265	10.73	<0.001	<0.001
Veillonella	-2.2552	14.784	<0.001	<0.001
Oribacterium	-2.1402	10.499	<0.001	<0.001
Aggregatibacter	-2.5911	13.703	<0.001	<0.001
Porphyromonas	1.5748	15.452	<0.001	<0.001
Mycoplasma	1.4871	10.702	<0.01	<0.01
Bacteroides	1.4567	11.735	<0.01	<0.01
Cardiobacterium	-1.702	12.594	<0.01	<0.01
Parvimonas	1.2651	13.251	<0.01	<0.01
Selenomonas	-0.94956	15.082	<0.05	<0.05
Fusobacterium	0.54974	17.479	0.086998	0.11185
Butyrivibrio	0.71479	9.2838	0.098256	0.12282
Dialister	-0.76662	13.014	0.11462	0.1394
Paludibacter	0.59143	13.375	0.16633	0.19697
Bulleidia	-0.58357	10.01	0.21877	0.25243
Campylobacter	0.29517	14.966	0.38402	0.43202
Eikenella	-0.26099	12.195	0.56464	0.61973
Schwartzia	0.20297	12.061	0.61726	0.64448
Staphylococcus	-0.10987	9.9382	0.80758	0.80758

Log2FC (log2 fold change) represents the ratio of two groups (PPD 7–9 vs. PPD 3–4) based log2. LogCPM (log counts per million) represents the expression level of variables. FDR is the false discovery rate as correction of P value.

TABLE 4 | Significant functions and pathways on L2 and L3 compared with periodontitis and healthy groups.

Variables (L2)	log2FC	LogCPM	P values	FDR
Cell Motility	0.3819	14.092	<0.001	<0.001
Environmental Adaptation	0.10176	10.496	<0.001	<0.01
Signal Transduction	0.087304	13.795	<0.001	<0.01
Metabolism of Terpenoids and Polyketides	-0.02687	14.146	<0.001	<0.01
Metabolism of Cofactors and Vitamins	-0.02879	15.525	<0.001	<0.01
Folding, Sorting and Degradation	-0.02063	14.726	<0.001	<0.01
Cellular Processes and Signaling	-0.0408	15.158	<0.01	<0.05
Nervous System	-0.07875	9.5907	<0.01	<0.05
Variables (L2)	log2FC	LogCPM	P values	FDR
Bacterial motility proteins	0.40809	13.024	<0.001	<0.001
Bacterial chemotaxis	0.48527	11.742	<0.001	<0.001
Methane metabolism	0.088374	13.439	<0.001	<0.001
Flagellar assembly	0.58686	11.586	<0.001	<0.001
Ether lipid metabolism	0.56359	5.9541	<0.001	<0.001
Other ion-coupled transporters	-0.060619	13.597	<0.001	<0.001
Xylene degradation	0.28929	8.4607	<0.001	<0.01
Pentose and glucuronate interconversions	0.11254	11.654	<0.001	<0.01
Carotenoid biosynthesis	-0.43936	6.6335	<0.001	<0.01
Ubiquitin system	-0.40326	7.6626	<0.001	<0.01
Nitrogen metabolism	-0.040959	12.795	<0.001	<0.01
Insulin signaling pathway	0.082719	9.7838	<0.001	<0.01
D-Arginine and D-ornithine metabolism	-0.29459	5.8248	<0.001	<0.01
Base excision repair	-0.046719	12.258	<0.001	<0.01
Sulfur metabolism	-0.12471	11.39	<0.01	<0.05
Polycyclic aromatic hydrocarbon degradation	-0.077289	10.401	<0.01	<0.05
Tyrosine metabolism	-0.056456	11.869	<0.01	<0.05
Synthesis and degradation of ketone bodies	0.20517	9.0478	<0.01	<0.05
Histidine metabolism	0.053466	12.395	<0.01	<0.05
Linoleic acid metabolism	0.16046	8.5719	<0.01	<0.05
Chloroalkane and chloroalkene degradation	0.090688	10.423	<0.01	<0.05
Carbon fixation pathways in prokaryotes	0.034009	13.338	<0.01	<0.05
Primary immunodeficiency	-0.070153	9.19	<0.01	<0.05
Glycine, serine, and threonine metabolism	0.0204	13.083	<0.01	<0.05
Type I diabetes mellitus	-0.05003	9.2339	<0.01	<0.05

Log2FC (log2 fold change) represents the ratio of two groups (PD vs. HC) based log2. LogCPM (log counts per million) represents the expression level of variables. FDR is the false discovery rate as correction of P value.

distinguished PD from HC almost without error (class error<0.01), implying special functions and metabolic pathways of periodontitis-related microbiota (**Supplementary Figure 2**).

DISCUSSION

Overall Review

Our study elucidated the subgingival microbial structure of periodontitis patients *via* integrated datasets. Extensive literature searches and rigorous screening criteria were performed. Datasets were merged and processed using uniform standards for eventual analysis. Subgingival microbial community, periodontitis biomarkers, potential functions of microbiota, and their collaborative network were also evaluated. Furthermore, we described the variation of microbial composition in different PPDs. Our results showed that some pathobionts were consistent with those reported in previous studies (Kirst et al., 2015; Liu et al., 2020) and supported the finding that periodontal dysbiosis was not due to specific microorganisms, rather due to the increasing level of pathobionts. The two reasons for this are likely that (1) microbial community dysbiosis leads to periodontal disease, and (2) periodontitis is caused by some specific pathogenic bacteria that have not yet been identified.

Microorganisms Associated With Periodontitis

Compared with previous studies and the included subgroup studies, our analysis yielded some consistent results and novel potential periodontitis-related microbes. *Porphyromonas*, *Treponema*, and *Tannerella* were found closely related to periodontitis. Moreover, with the deepening of PPD, *Porphyromonas* and *Treponema* occupied the main components of subgingival microbes, while the healthy periodontium-related genera *Neisseria* and *Lautropia* decreased. The abundance of *Spirochaetes*, *Synergistes, Desulfobulbus*, and *Bacteroides* also increased in PD. These results were consistent with those reported by (Galimanas et al., 2014; Califf et al., 2017; Pérez-Chaparro et al., 2018). Healthy gingiva-associated genera *Rothia*, *Capnocytophaga*, *Veillonella*, *Corynebacterium*, and *Neisseria* were found in our results, which were also partly reported by (Galimanas et al., 2014; Bizzarro et al., 2016; Chen et al., 2018). Notably, Proteobacteria appeared to be a point of contention with different reports in several articles. In Shi's and Griffen's studies, Proteobacteria was higher in healthy controls than in periodontitis patients. By contrast, Galimanas reported that Proteobacteria was associated with periodontitis, although he later noted that Proteobacteria was associated with the healthy population in subgingival microbiota. Our findings showed that Proteobacteria was more closely related to HC and the proportion of Proteobacteria decreased with the deepening of the periodontal pocket. Additionally, we identified some potential genera associated with periodontitis, such as *TG5* and *Catonella*, whose relationship with periodontitis has been rarely reported; thus, more trials are required to validate their pathogenic mechanisms in periodontitis.

LEfSe analysis showed that *Corynebacterium* and *Rothia* were biomarkers for healthy periodontium. *Rothia* is among the normal genera that colonize the oral cavity. Although it was detected in some opportunistic infectious diseases (Ramanan et al., 2014), our findings support its classification as a typical oral bacterium. This is consistent with Meuric's research (Meuric et al., 2017) that the ratio of *Porphyromonas*, *Treponema*, and *Tannerella* to *Rothia* and *Corynebacterium* is an excellent predictor of periodontitis, which regards *Corynebacterium* and *Rothia* as non-pathogenic genera. *Veillonella*, *Kingella*, and *Neisseria* were thought to be healthy biomarkers. *Veillonella* can consume the lactic acid produced by *Streptococcus mutans* to prevent caries (Sanz et al., 2017). In a clinical trial of periodontal therapy, *Kingella* and *Veillonella* were found to be more associated with therapeutic success (Colombo et al., 2012). For *Neisseria*, its abundance declined from HCs and the shallow layer to the deep layer in periodontal pocket, which showed it was a biomarker for healthy periodontium.

Our results showed that *Desulfobulbus*, *Treponema*, and *Tannerella* were periodontitis biomarkers. The correlation between *Desulfobulbus* and periodontitis was discovered recently, represented by *Desulfobulbus oralis*, which has not been valued previously owing to limitations in culture and isolation (Cross et al., 2018). It was found that *D. oralis* can directly induce the inflammatory response in oral epithelial cells to promote the occurrence of periodontitis. *Treponema denticola*, *Porphyromonas gingivalis*, and *Tannerella forsythia* belong to the red complex, which are considered to be the most periodontitis-related microbial aggregation (Socransky et al., 1998). In our results, *Porphyromonas* and *Treponema* both displayed significant dominance in the PPD 7–9 mm group with a 11% abundance ratio. Inversely, they were <2% in the PPD 0–3 mm group. Our network correlation analysis also reflected the synergy among these microbes. *P. gingivalis* is critically related to periodontitis. This black anaerobic bacterium relies on its fimbriae, lipopolysaccharides, proteases, and other virulence factors to colonize on teeth and periodontal tissues, and it can co-aggregate a variety of other potential pathogenic microorganisms (Mysak et al., 2014). Additionally, it can interfere with host immune functions such as cytokine secretion, degrade recruitment, and weaken leukocyte defenses in periodontal tissues (Kobayashi-Sakamoto et al., 2003). Several years ago, the subgingival concentration of *Treponema* growth was considered significantly related to PPD and attachment loss (Armitage et al., 1982). *Tannerella* is another pathogenic genus implicated in periodontitis, which is associated with subgingival bleeding and regarded as a risk marker of PD (Suda et al., 2004). The abundance of Fusobacteria increased with increasing PPD. On the genus level, the proportion of *Fusobacterium* changed from 8% in the 0–3 mm PPD group to 16% in the 7–9 mm PPD group. The role of Fusobacteria in deep periodontal pockets cannot be ignored. Its pathogenic ability is to co-aggregate and help periodontal pathobionts to colonize, which acts as a bridge for dental plaque biofilm formation (Rickard et al., 2003). Moreover, *F. nucleatum* can invade epithelial cells to escape host immunity and trigger inflammatory responses, and FadA was recognized as the crucial virulence factor (Han et al., 2000).

TG5 and *Catonella* were identified as potential periodontitis-related pathobionts in our study. Few studies have reported its existence in patients with periodontitis. Only few research studies have confirmed the virulence and pathogenic mechanisms of *TG5* in periodontitis. *Catonella* is an oral pathobiont associated with oral infections and oral cancer (Zhao et al., 2017). However, its role in the development of periodontal disease remains to be investigated.

In network analysis, we can clearly observe positive correlations among pathobionts, and positive correlations among normal and healthy oral microorganisms. Our results showed that the microbes associated with PD or HC can be classified into two communities. The microorganisms cooperate with others in the same community, but are negatively correlated with the bacteria in the other community. This association was consistent with the results reported by Liu (Liu et al., 2020) and supported the antagonistic relationships between pathogenic and non-pathogenic bacteria. It was worth emphasizing that *Mogibacterium*, *Parvimonas*, and *Filifactor* were all positively correlated with some known pathogenic microorganisms in the correlation analysis, although they were not found in the LEfSe analysis. This may indicate a new direction in the discovery of pathobionts, although their abundance is not high.

Remarkable Microbial Metabolic Pathways

The differences in metabolic pathways and functions caused by alteration of microbiota were obvious. The increasing levels of bacterial motility proteins and flagellar assembly may imply that the invasion ability of pathobionts plays an important role in

periodontitis. The metabolism of tyrosine was significantly different in our results ($p<0.01$). Liebsch et al. (Liebsch et al., 2019) showed that dental plaque and pocket depth were positively correlated with the metabolites derived from phenylalanine and tyrosine catabolism. For example, phenylacetate, a bacterial metabolite, was significantly associated with periodontal disease and may be a candidate marker for periodontal disease screening (Liebsch et al., 2019). The metabolism level of ketone bodies increased in PD ($p<0.05$). This was consistent with another study that reported decreased pyruvate and pyruvic acid levels in the saliva of patients with chronic periodontitis (Romano et al., 2018). Sulfur metabolism was different between the PD and HC groups, which may be associated with the production of volatile sulfur compounds (VSCs). These VSCs are known to be produced by anaerobic microbes and are toxic to periodontal tissue (Hampelska et al., 2020). It is also the major reason for halitosis in patients with periodontitis. A longitudinal study evaluated the correlation between periodontitis progress and VSCs, and the results showed a positive association between the two (Makino et al., 2012).

Limitations and Prospect

In this study, we determined the biomarkers of periodontitis based on the abundance of microorganisms; however, our results are inadequate and more extensive research, such as on virulence factors, is needed to confirm their pathogenesis in periodontitis. At the literature inclusion stage, we did not retrieve all the studies related to subgingival microbiota of periodontitis patients. However, it should be pointed out that the data acquisition and analysis in our study are different from a systematic review and meta-analysis. The biggest hurdle was to access complete and high-quality datasets. Because many datasets are unavailable, it is challenging to obtain all datasets to analyze. The short reads were matched with the Greengenes database library (gg-13-8 version) in Closed Reference way to annotate. However, it screened some unidentified microorganisms in this method. A more complete microbial gene database and powerful computers are needed to improve this analysis. Heterogeneity across studies is a confounding factor. For instance, 16SrRNA sequencing analysis can be biased by PCR. This is inevitable at present, and we can only reduce this heterogeneity through unified criteria and data processing methods. We screened literature with predetermined criteria and only included high-quality datasets for merging, aiming to decrease the influence of multiple variables between different researches as much as possible. Some variables remain inevitably among the similar studies, but the variables from different studies can be minimized, which has been described by (Kirst et al., 2015). Some other studies confirmed that the idea of merging datasets under strict screening criteria and unified sequence data processing can be feasible (Sze and Schloss, 2016; Meuric et al., 2017). Overall, this method could magnify the pathogenic features of periodontitis and minimize variables from different studies, such as individual differences, experimental differences, and technical differences, helping to identify the common pathobionts among different periodontitis patients. Perhaps, a more unified protocol for high-throughput sequencing studies can be designed, which will be conducive to the realization of data aggregation. Additionally, the results of functional profiles predicted from 16S amplicons were not as accurate as those of whole-genome sequencing, which should be further validated (Jing et al., 2021).

CONCLUSION

We applied strict and unified standards to process sequence datasets, and analyzed the microbial community structure and functions in periodontitis. The results showed significant differences in the structure of microorganisms and potential functions and metabolic pathways between the PD and HC groups. Furthermore, we revealed that the composition of the subgingival microbiota changed at different PPD sites. Our results identified some potential periodontitis biomarkers and explored the functions of subgingival microbiota in periodontitis. Besides, we described a feasible method to pool microbial sequence data, which can be used in other related areas. With the updates to microbial database and the improvement of sequencing technology, the advantages of this method may be greater, which can be used to identify more unknown and unannotated pathobionts in the future.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: National Center for Biotechnology (NCBI) and The European Nucleotide Archive (ENA): PRJEB19122, PRJEB6047, PRJNA289294, SRP009299, PRJNA509532, SRP228020, SRP102224, PRJNA324274, SRP075100.

AUTHOR CONTRIBUTIONS

ZC and LZ contributed to conception and design. ZC and SL contributed to data acquisition, screening, processing, and manuscript drafting. ZC, SL, and SH contributed to analysis and interpretation of data. ZC, SL, SH, and LZ contributed to manuscript revisions. All authors contributed to the article and approved the submitted version.

FUNDING

This study is supported by the National Natural Science Foundation of China (Grant No. 81970944 and No.81991502) and a research grant from West China Hospital of Stomatology (LCYJ2019-4).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2021.663756/full#supplementary-material

REFERENCES

Anderson, M. J. (2001). A New Method for non-Parametric Multivariate Analysis of Variance. *Austral Ecol.* 26 (1), 32–46. doi: 10.1111/j.1442-9993.2001.01070.pp.x

Anderson, M. J. (2006). Distance-Based Tests for Homogeneity of Multivariate Dispersions. *Biometrics* 62 (1), 245–253. doi: 10.1111/j.1541-0420.2005.00440.x

Armitage, G. C., Dickinson, W. R., Jenderseck, R. S., Levine, S. M., and Chambers, D. W. (1982). Relationship Between the Percentage of Subgingival Spirochetes and the Severity of Periodontal Disease. *J. Periodontol.* 53 (9), 550–556. doi: 10.1902/jop.1982.53.9.550

Bizzarro, S., Laine, M. L., Buijs, M. J., Brandt, B. W., Crielaard, W., Loos, B. G., et al. (2016). Microbial Profiles at Baseline and Not the Use of Antibiotics Determine the Clinical Outcome of the Treatment of Chronic Periodontitis. *Sci. Rep.* 6, 20205. doi: 10.1038/srep20205

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, Interactive, Scalable and Extensible Microbiome Data Science Using QIIME 2. *Nat. Biotechnol.* 37 (8), 852–857. doi: 10.1038/s41587-019-0209-9

Califf, K. J., Schwarzberg-Lipson, K., Garg, N., Gibbons, S. M., Caporaso, J. G., Slots, J., et al. (2017). Multi-Omics Analysis of Periodontal Pocket Microbial Communities Pre- and Posttreatment. *mSystems* 2 (3). doi: 10.1128/mSystems.00016-17

Chapple, I. L. C., Mealey, B. L., Van Dyke, T. E., Bartold, P. M., Dommisch, H., Eickholz, P., et al. (2018). Periodontal Health and Gingival Diseases and Conditions on an Intact and a Reduced Periodontium: Consensus Report of Workgroup 1 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions. *J. Clin. Periodontol.* 45 Suppl 20, S68–s77. doi: 10.1111/jcpe.12940

Chen, C., Hemme, C., Beleno, J., Shi, Z. J., Ning, D., Qin, Y., et al. (2018). Oral Microbiota of Periodontal Health and Disease and Their Changes After Nonsurgical Periodontal Therapy. *Isme J.* 12 (5), 1210–1224. doi: 10.1038/s41396-017-0037-1

Chong, J., Liu, P., Zhou, G., and Xia, J. (2020). Using MicrobiomeAnalyst for Comprehensive Statistical, Functional, and Meta-Analysis of Microbiome Data. *Nat. Protoc.* 15 (3), 799–821. doi: 10.1038/s41596-019-0264-1

Colombo, A. P., Bennet, S., Cotton, S. L., Goodson, J. M., Kent, R., Haffajee, A. D., et al. (2012). Impact of Periodontal Therapy on the Subgingival Microbiota of Severe Periodontitis: Comparison Between Good Responders and Individuals With Refractory Periodontitis Using the Human Oral Microbe Identification Microarray. *J. Periodontol.* 83 (10), 1279–1287. doi: 10.1902/jop.2012.110566

Cross, K. L., Chirania, P., Xiong, W., Beall, C. J., Elkins, J. G., Giannone, R. J., et al. (2018). Insights Into the Evolution of Host Association Through the Isolation and Characterization of a Novel Human Periodontal Pathobiont, Desulfobulbus Oralis. *mBio* 9 (2). doi: 10.1128/mBio.02061-17

Cummings, L. A., Kurosawa, K., Hoogestraat, D. R., SenGupta, D. J., Candra, F., Doyle, M., et al. (2016). Clinical Next Generation Sequencing Outperforms Standard Microbiological Culture for Characterizing Polymicrobial Samples. *Clin. Chem.* 62 (11), 1465–1473. doi: 10.1373/clinchem.2016.258806

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., et al. (2006). Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible With ARB. *Appl. Environ. Microbiol.* 72 (7), 5069–5072. doi: 10.1128/aem.03006-05

Downs, S. H., and Black, N. (1998). The Feasibility of Creating a Checklist for the Assessment of the Methodological Quality Both of Randomised and non-Randomised Studies of Health Care Interventions. *J. Epidemiol. Community Health* 52 (6), 377–384. doi: 10.1136/jech.52.6.377

Galimanas, V., Hall, M. W., Singh, N., Lynch, M. D., Goldberg, M., Tenenbaum, H., et al. (2014). Bacterial Community Composition of Chronic Periodontitis and Novel Oral Sampling Sites for Detecting Disease Indicators. *Microbiome* 2, 32. doi: 10.1186/2049-2618-2-32

Genco, R. J., and Borgnakke, W. S. (2013). Risk Factors for Periodontal Disease. *Periodontol. 2000* 62 (1), 59–94. doi: 10.1111/j.1600-0757.2012.00457.x

Griffen, A. L., Beall, C. J., Campbell, J. H., Firestone, N. D., Kumar, P. S., Yang, Z. K., et al. (2012). Distinct and Complex Bacterial Profiles in Human Periodontitis and Health Revealed by 16S Pyrosequencing. *Isme J.* 6 (6), 1176–1185. doi: 10.1038/ismej.2011.191

Gu, W., Miller, S., and Chiu, C. Y. (2019). Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection. *Annu. Rev. Pathol.* 14, 319–338. doi: 10.1146/annurev-pathmechdis-012418-012751

Hampelska, K., Jaworska, M. M., Babalska, Z., and Karpiński, T. M. (2020). The Role of Oral Microbiota in Intra-Oral Halitosis. *J. Clin. Med.* 9 (8). doi: 10.3390/jcm9082484

Han, Y. W., Shi, W., Huang, G. T., Kinder Haake, S., Park, N. H., Kuramitsu, H., et al. (2000). Interactions Between Periodontal Bacteria and Human Oral Epithelial Cells: Fusobacterium Nucleatum Adheres to and Invades Epithelial Cells. *Infect. Immun.* 68 (6), 3140–3146. doi: 10.1128/iai.68.6.3140-3146.2000

Jing, G., Zhang, Y., Cui, W., Liu, L., Xu, J., and Su, X. (2021). Meta-Apo Improves Accuracy of 16S-Amplicon-Based Prediction of Microbiome Function. *BMC Genomics* 22 (1), 9. doi: 10.1186/s12864-020-07307-1

Kinane, D. F., Stathopoulou, P. G., and Papapanou, P. N. (2017). Periodontal Diseases. *Nat. Rev. Dis. Primers* 3, 17038. doi: 10.1038/nrdp.2017.38

Kirst, M. E., Li, E. C., Alfant, B., Chi, Y. Y., Walker, C., Magnusson, I., et al. (2015). Dysbiosis and Alterations in Predicted Functions of the Subgingival Microbiome in Chronic Periodontitis. *Appl. Environ. Microbiol.* 81 (2), 783–793. doi: 10.1128/aem.02712-14

Kobayashi-Sakamoto, M., Isogai, E., and Hirose, K. (2003). Porphyromonas Gingivalis Modulates the Production of Interleukin 8 and Monocyte Chemotactic Protein 1 in Human Vascular Endothelial Cells. *Curr. Microbiol.* 46 (2), 109–114. doi: 10.1007/s00284-002-3782-x

Langille, M. G., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., et al. (2013). Predictive Functional Profiling of Microbial Communities Using 16S rRNA Marker Gene Sequences. *Nat. Biotechnol.* 31 (9), 814–821. doi: 10.1038/nbt.2676

Liebsch, C., Pitchika, V., Pink, C., Samietz, S., Kastenmüller, G., Artati, A., et al. (2019). The Saliva Metabolome in Association to Oral Health Status. *J. Dent. Res.* 98 (6), 642–651. doi: 10.1177/0022034519842853

Liu, G., Chen, F., Cai, Y., Chen, Z., Luan, Q., and Yu, X. (2020). Measuring the Subgingival Microbiota in Periodontitis Patients: Comparison of the Surface Layer and the Underlying Layers. *Microbiol. Immunol.* 64 (2), 99–112. doi: 10.1111/1348-0421.12759

Makino, Y., Yamaga, T., Yoshihara, A., Nohno, K., and Miyazaki, H. (2012). Association Between Volatile Sulfur Compounds and Periodontal Disease Progression in Elderly non-Smokers. *J. Periodontol.* 83 (5), 635–643. doi: 10.1902/jop.2011.110275

Meuric, V., Le Gall-David, S., Boyer, E., Acuña-Amador, L., Martin, B., Fong, S. B., et al. (2017). Signature of Microbial Dysbiosis in Periodontitis. *Appl. Environ. Microbiol.* 83 (14). doi: 10.1128/aem.00462-17

Mysak, J., Podzimek, S., Sommerova, P., Lyuya-Mi, Y., Bartova, J., Janatova, T., et al. (2014). Porphyromonas Gingivalis: Major Periodontopathic Pathogen Overview. *J. Immunol. Res.* 2014, 476068. doi: 10.1155/2014/476068

Papapanou, P. N., Sanz, M., Buduneli, N., Dietrich, T., Feres, M., Fine, D. H., et al. (2018). Periodontitis: Consensus Report of Workgroup 2 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions. *J. Periodontol.* 89 Suppl 1, S173–s182. doi: 10.1002/jper.17-0721

Pei, J., Li, F., Xie, Y., Liu, J., Yu, T., and Feng, X. (2020). Microbial and Metabolomic Analysis of Gingival Crevicular Fluid in General Chronic Periodontitis Patients: Lessons for a Predictive, Preventive, and Personalized Medical Approach. *Epma J.* 11 (2), 197–215. doi: 10.1007/s13167-020-00202-5

Pérez-Chaparro, P. J., McCulloch, J. A., Mamizuka, E. M., Moraes, A., Faveri, M., Figueiredo, L. C., et al. (2018). Do Different Probing Depths Exhibit Striking Differences in Microbial Profiles? *J. Clin. Periodontol.* 45 (1), 26–37. doi: 10.1111/jcpe.12811

Ramanan, P., Barreto, J. N., Osmon, D. R., and Tosh, P. K. (2014). Rothia Bacteremia: A 10-Year Experience at Mayo Clinic, Rochester, Minnesota. *J. Clin. Microbiol.* 52 (9), 3184–3189. doi: 10.1128/jcm.01270-14

Rickard, A. H., Gilbert, P., High, N. J., Kolenbrander, P. E., and Handley, P. S. (2003). Bacterial Coaggregation: An Integral Process in the Development of Multi-Species Biofilms. *Trends Microbiol.* 11 (2), 94–100. doi: 10.1016/s0966-842x(02)00034-3

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* 26 (1), 139–140. doi: 10.1093/bioinformatics/btp616

Romano, F., Meoni, G., Manavella, V., Baima, G., Tenori, L., Cacciatore, S., et al. (2018). Analysis of Salivary Phenotypes of Generalized Aggressive and Chronic Periodontitis Through Nuclear Magnetic Resonance-Based Metabolomics. *J. Periodontol.* 89 (12), 1452–1460. doi: 10.1002/jper.18-0097

Sanz, M., Beighton, D., Curtis, M. A., Cury, J. A., Dige, I., Dommisch, H., et al. (2017). Role of Microbial Biofilms in the Maintenance of Oral Health and in the Development of Dental Caries and Periodontal Diseases. Consensus Report of Group 1 of the Joint EFP/ORCA Workshop on the Boundaries Between

Caries and Periodontal Disease. *J. Clin. Periodontol.* 44 Suppl 18, S5–s11. doi: 10.1111/jcpe.12682

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., et al. (2011). Metagenomic Biomarker Discovery and Explanation. *Genome Biol.* 12 (6), R60. doi: 10.1186/gb-2011-12-6-r60

Shi, M., Wei, Y., Nie, Y., Wang, C., Sun, F., Jiang, W., et al. (2020). Alterations and Correlations in Microbial Community and Metabolome Characteristics in Generalized Aggressive Periodontitis. *Front. Microbiol.* 11, 573196. doi: 10.3389/fmicb.2020.573196

Socransky, S. S., Haffajee, A. D., Cugini, M. A., Smith, C., and Kent, ,.R. (1998). Microbial Complexes in Subgingival Plaque. *J. Clin. Periodontol.* 25 (2), 134–144. doi: 10.1111/j.1600-051x.1998.tb02419.x

Suda, R., Kobayashi, M., Nanba, R., Iwamaru, M., Hayashi, Y., Lai, C. H., et al. (2004). Possible Periodontal Pathogens Associated With Clinical Symptoms of Periodontal Disease in Japanese High School Students. *J. Periodontol.* 75 (8), 1084–1089. doi: 10.1902/jop.2004.75.8.1084

Su, X., Jing, G., Zhang, Y., and Wu, S. (2020). Method Development for Cross-Study Microbiome Data Mining: Challenges and Opportunities. *Comput. Struct. Biotechnol. J.* 18, 2075–2080. doi: 10.1016/j.csbj.2020.07.020

Sze, M. A., and Schloss, P. D. (2016). Looking for a Signal in the Noise: Revisiting Obesity and the Microbiome. *mBio* 7 (4). doi: 10.1128/mBio.01018-16

Tsai, C. Y., Tang, C. Y., Tan, T. S., Chen, K. H., Liao, K. H., and Liou, M. L. (2018). Subgingival Microbiota in Individuals With Severe Chronic Periodontitis. *J. Microbiol. Immunol. Infect.* 51 (2), 226–234. doi: 10.1016/j.jmii.2016.04.007

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences Into the New Bacterial Taxonomy. *Appl. Environ. Microbiol.* 73 (16), 5261–5267. doi: 10.1128/aem.00062-07

Wei, Y., Shi, M., Zhen, M., Wang, C., Hu, W., Nie, Y., et al. (2019). Comparison of Subgingival and Buccal Mucosa Microbiome in Chronic and Aggressive Periodontitis: A Pilot Study. *Front. Cell Infect. Microbiol.* 9, 53. doi: 10.3389/fcimb.2019.00053

Wolf, S. M., and Evans, B. J. (2018). Return of Results and Data to Study Participants. *Science* 362 (6411), 159–160. doi: 10.1126/science.aav0005

Zhao, H., Chu, M., Huang, Z., Yang, X., Ran, S., Hu, B., et al. (2017). Variations in Oral Microbiota Associated With Oral Cancer. *Sci. Rep.* 7 (1), 11773. doi: 10.1038/s41598-017-11779-9

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cellular and Infection Microbiology

ORIGINAL RESEARCH
published: 16 August 2021
doi: 10.3389/fcimb.2021.691092

Oral Microbiota Composition and Function Changes During Chronic Erythematous Candidiasis

Xin Lyu[1†‡], Hui Zheng[2†‡], Xu Wang[1,2‡], Heyu Zhang[2‡], Lu Gao[2], Zhe Xun[2], Qian Zhang[2], Xuesong He[3], Hong Hua[1], Zhimin Yan[1] and Feng Chen[2*]*

[1] Department of Oral Medicine, Peking University School and Hospital of Stomatology, National Center of Stomatology, National Clinical Research Center for Oral Diseases, National Engineering Laboratory for Digital and Material Technology of Stomatology, Beijing Key Laboratory of Digital Stomatology, Research Center of Engineering and Technology for Computerized Dentistry Ministry of Health, NMPA Key Laboratory for Dental Materials, Beijing, China, [2] Central Laboratory, Peking University School and Hospital of Stomatology, National Center of Stomatology, National Clinical Research Center for Oral Diseases, National Engineering Laboratory for Digital and Material Technology of Stomatology, Beijing Key Laboratory of Digital Stomatology, Research Center of Engineering and Technology for Computerized Dentistry Ministry of Health, NMPA Key Laboratory for Dental Materials, Beijing, China, [3] Department of Microbiology, The Forsyth Institute, Cambridge, MA, United States

Oral microbiota is constantly changing with the host state, whereas the oral microbiome of chronic erythematous candidiasis remains poorly understood. The aim of this study was to compare oral microbial signatures and functional profiling between chronic erythematous candidiasis and healthy subjects. Using shotgun metagenomic sequencing, we analyzed the microbiome in 12 chronic erythematous candidiasis, 12 healthy subjects, and 2 chronic erythematous candidiasis cured by antifungal therapy. We found that the salivary microbiota of chronic erythematous candidiasis was significantly different from that of healthy subjects. Among them, *Rothia mucilaginosa* and *Streptococcus mitis* were the most abundant disease-enriched species (Mann-Whitney U-test, $P < 0.05$). In addition, co-occurrence network analysis showed that *C. albicans* formed densely connected modules with oral bacterial species and was mainly positive connected to *Streptococcus* species. Furthermore, we investigated the functional potentials of the microbiome and identified a set of microbial marker genes associated with chronic erythematous candidiasis. Some of these genes enriching in chronic erythematous candidiasis are involved in eukaryotic ribosome, putative glutamine transport system, and cytochrome bc1 complex respiratory unit. Altogether, this study revealed the changes of oral microbial composition, the co-occurrence between *C. albicans* and oral bacteria, as well as the changes of microbial marker genes during chronic erythematous candidiasis, which provides evidence of oral microbiome as a target for the treatment and prevention of chronic erythematous candidiasis.

Keywords: oral candidiasis, *Candida albicans*, oral microbiota, metagenomics, functional potentials

OPEN ACCESS

Edited by:
Dongmei Deng,
VU University Amsterdam, Netherlands

Reviewed by:
Claudio Cermelli,
University of Modena and Reggio Emilia, Italy
Fernanda Brighenti,
São Paulo State University, Brazil

***Correspondence:**
Feng Chen
chenfeng2011@hsc.pku.edu.cn
Zhimin Yan
yzhimin96@163.com

†Present address:
Xin Lyu,
Department of Oral Medicine, Beijing Stomatological Hospital, Capital Medical University, Beijing, China
Hui Zheng,
Department of Orthodontics, School and Hospital of Stomatology, Fujian Medical University, Fuzhou, China

‡These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Microbiome in Health and Disease, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: *05 April 2021*
Accepted: *12 July 2021*
Published: *16 August 2021*

Citation:
Lyu X, Zheng H, Wang X, Zhang H, Gao L, Xun Z, Zhang Q, He X, Hua H, Yan Z and Chen F (2021) Oral Microbiota Composition and Function Changes During Chronic Erythematous Candidiasis. Front. Cell. Infect. Microbiol. 11:691092. doi: 10.3389/fcimb.2021.691092

INTRODUCTION

Oral microbiota is a reflection of the host state and plays an important role in the development of various diseases. Previous studies have shown us the profiling of human oral microbiota in many oral diseases, such as periodontitis, dental caries, and oral squamous cell carcinoma with the use of 16S rRNA sequence analysis or shotgun whole-genome metagenomic methods (Li et al., 2014; Baker et al., 2021; Sarkar et al., 2021). However, there remains some diseases which are well worth exploring from the perspective of oral microbiota.

Chronic erythematous candidiasis is the most common type of Oral candidiasis (Oral candidosis, OC) (Hu et al., 2020), which is the most common opportunistic fungal disease occurring in oral cavity. It is estimated that 5% of newborns, 10% of elderly patients, 30-94% of individuals with malignant tumors, and nearly 90% of HIV-infected patients can develop into OC (Davies et al., 2008; Poncet et al., 2009). *Candida albicans* (*C. albicans*), a symbiotic microorganism carried by about 80% of the general population, is widely believed to be the main causative agent of OC, and accounts for up to 95% of cases (Vila et al., 2020). A variety of local and systemic factors can lead to the overgrowth of *C. albicans* on oral mucosa, and make it from commensal to pathogenic (Millsop and Fazel, 2016).

Increasing evidence indicates that *C. albicans* exhibits diverse interactions with oral bacterial species, ranging from antagonistic to synergistic (Montelongo-Jauregui and Lopez-Ribot, 2018; Abrantes and Africa, 2020). On the one hand, *C. albicans* was shown to co-aggregate with varieties of oral bacterial flora, such as *Streptococci*, *Fusobacterium nucleatum* and *Porphyromonas gingivalis* (Wu et al., 2015; Montelongo-Jauregui and Lopez-Ribot, 2018; Xiao et al., 2018; Vila et al., 2020). On the other hand, some oral bacteria can affect the colonization and activity of *C. albicans*. For example, *Streptococcus oralis* and *Porphyromonas gingivalis* can enhance the expression level of genes encoding cell surface adhesin in *C. albicans* and the biofilm formation ability of *C. albicans* (Cavalcanti et al., 2016; Bartnicka et al., 2019).

Although various studies have explored the relationship between *C. albicans* and oral bacteria, few studies clarified the whole oral microbiome during *C. albicans* infection, especially in chronic erythematous candidiasis patients without the presence of other factors that can affect the composition of microbiota. In addition, mapping the complex nature of oral microbiota in chronic erythematous candidiasis patients is of great significance in expanding our understanding of oral microbiome and chronic erythematous candidiasis itself. Therefore, this study aimed to compare the salivary microbiota and its gene function between chronic erythematous candidiasis and healthy subjects, in order to profile the microbial communities in chronic erythematous candidiasis.

MATERIALS AND METHODS

Subject Recruitment

Study participants aged 45-65 years were recruited from the Department of Oral Medicine, Peking University School and Hospital of Stomatology, China. Participants were excluded from the research if they (1) had removable dentures, (2) suffered hyposalivation (unstimulated salivary flow rate <1ml/10min) (Lopez-Pintor et al., 2016), diabetes, cancer, anemia, HIV positive, or other severe local or systemic infections, (3) used steroid drugs, immunosuppressants, antibiotics or mouthwash in the last 3 months, (4) had head or neck radiotherapy within the 3 previous months, (5) had smoking history. Participants with clinical manifestations of chronic erythematous candidiasis (Coronado-Castellote and Jimenez-Soriano, 2013) and a positive result on a mycological examination (smear and culture) were included in the disease group (DIS, n = 12). Participants with no clinical manifestations of oral candidiasis, no other oral mucosal disease and a negative result on a mycological examination (smear and culture) were included in the healthy control group (HC, n = 12). Besides, 2 participants suffered from chronic erythematous candidiasis before, but now had been cured by antifungal therapy were also included into our research. The social demographics including age and gender were collected. A comprehensive oral examination including salivary pH, probing depth (PD), DMFT (decayed, missing and filled teeth) was performed by an experienced dentist.

The study protocol was reviewed and approved by the Ethics Committee of the Peking University Health Science Center (PKUSSIRB-2013034). All participants received both written and oral information before consenting to participate in our study.

Saliva Sample Collection

For DIS and HC group, their salivary samples were collected after taking the questionnaire and completing the initial screening. For the cured group, their salivary samples were collected when the patients were cured and discontinued the antifungal drugs for 1 week. Unstimulated mixed saliva was collected by spitting into a 50-ml sterile collection tube, with at least 2 ml of volume. All participants were asked to skip breakfast and not to do tooth brushing 3 h before saliva collection. Saliva was collected between 8 a.m. and 10 a.m. by a single dentist in a quiet room. Samples were centrifuged at 10,000×g for 20 min at 4°C. Sediments were stored at -80°C until DNA extraction.

DNA Extraction, Library Preparation, and Whole Genome Shotgun Sequencing

Genomic DNA extraction was performed using the FastDNA SPIN kit for Soil (MP Biomedicals, USA). The DNA concentration was determined with the Nanodrop 8000 (Thermo Scientific, USA), and DNA quality was estimated by agarose gel electrophoresis. Genomic DNA was sheared by the Biorupter ® Pico sonication device (Diagenode, Belgium). DNA fragments of approximately 200 bp were selected with agarose gel electrophoresis. DNA libraries were constructed using the NEBNext Ultra DNA Library Prep Kit for Illumina (Illumina Inc, USA) according to the instruction manual. The insert size of the DNA libraries constructed from salivary samples varied from 155 to 266 bp (mean 210 ± 35.5 bp). All library sequencing was performed with 2×125-bp paired-end on the Illumina HiSeq 2000 platform (Illumina Inc, USA).

Bioinformatics Analysis of Sequence Data

Sequences with more than 3 ambiguous bases were removed. We screened reads for the minimum percentage of high-quality bases (Q30, >=50%) and trimmed low-quality bases (<Q30) on the terminal end. Paired reads with at least 1 read mapped to the human reference genome (GRCh37/hg19) were removed by SOAP2 software (-m 100 -x 1000).

The filtered clean reads were mapped to a database of predefined single-copy phylogenetic marker genes, with default options embedded in the MOCAT pipeline (Sunagawa et al., 2013). To estimate the *C. albicans* load, paired-end reads were mapped to the *C. albicans* reference genome using BWA with default settings. We simply defined the relative abundance of *C. albicans* as the relative abundance of sequences mapped to the reference genome.

Functional profiling of the microbial community was performed based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database of gene families and modules. The procedures were as follows: firstly, assembled the quality- and human- filtered WGS sequences into contigs using Soap2; then, detected the Open Reading Frames (ORFs) using Markergene software; finally, mapped the ORFs against protein-coding sequences from the KEGG Orthology using bowtie2.

Statistical Analysis

The data were presented as mean ± standard error unless otherwise indicated. Univariate statistical analyses were performed using t-test and Mann-Whitney U-test. A P value < 0.05 was considered statistically significant.

RESULTS

Participants' Characteristics and General Sequence Information

We carried out shotgun metagenomic sequencing of 26 salivary samples, including 12 from healthy subjects (HC group), 12 from chronic erythematous candidiasis patients (DIS group), and 2 from chronic erythematous candidiasis patients cured by antifungal therapy. No significant differences were found in age, salivary pH, PD, DMFT between HC and DIS (**Figure 1** and **Table S1**). It was worth mentioning that salivary pH value showed a decreasing trend in DIS compared with HC, although the difference was not statistically significant (independent t-test, $P = 0.074$).

In total, 154.4 gigabases (Gb) of paired-end sequence data were generated with an average of 27.6 million reads (5.5 Gb) per sample. A total of 97.2 ± 0.9% of these reads remained after filtering low-quality reads. Human DNA, which accounted for 48 ± 23.2% (range from 2.7% to 79.5%) of the high-quality reads, was filtered for further processing (**Figure S1**).

Phylogenetic Analysis of Microbial Community Composition

The alpha diversity of total microbiota in both DIS and HC was calculated, showing no significant inter-group differentiation (data not shown). In the DIS group, $4.9\times10^{-3} \pm 12.1\times10^{-3}$% (range from 9.53×10^{-5}% to 4.3×10^{-2}%) of the metagenomic reads were mapped to the *C. albicans* genome from the NCBI. In HC and in treated samples, sequences homologous to *C. albicans* were barely detected. The relative abundance of *C. albicans* was significantly different between DIS and HC (or cured participants) (**Figure 2A**; independent t-test, $P < 0.001$).

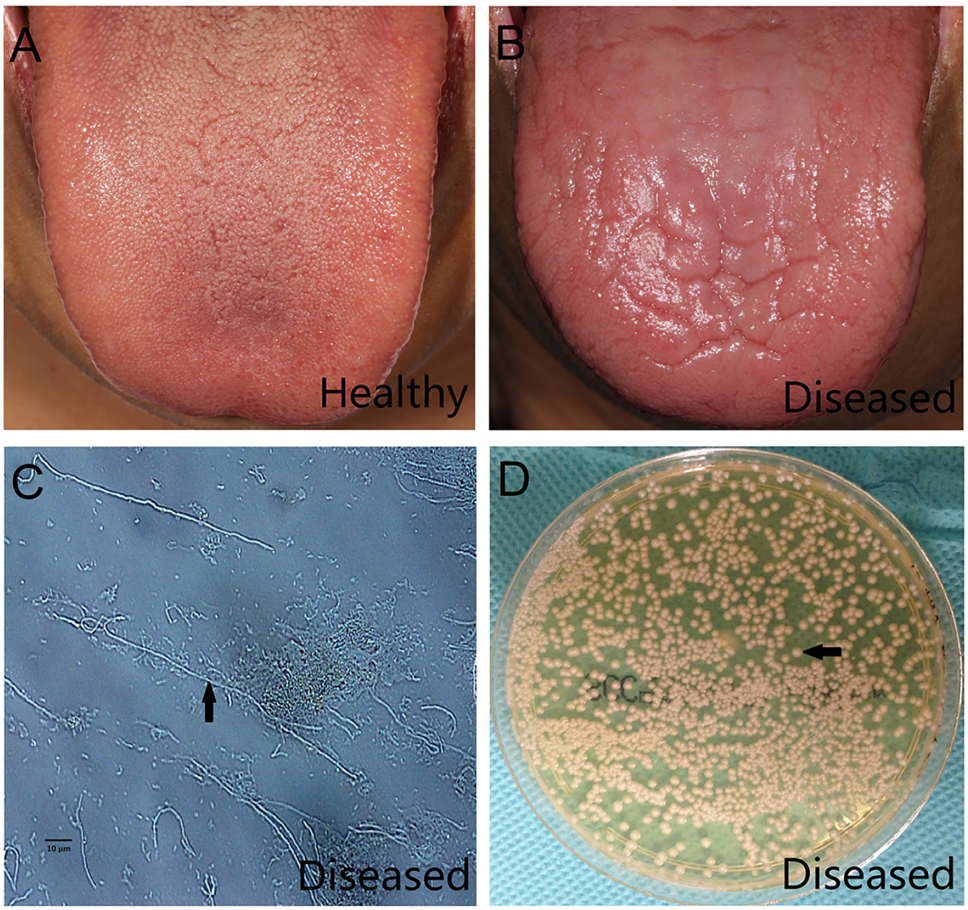
For taxonomic information from the short metagenomic sequences, metagenomic operational taxonomic units (mOTUs) were established based on single-copy phylogenetic marker genes. A total of 443 bacterial species in all samples were identified, which belonged to 9 different phyla (**Figures 2B, C**). The top 20 most abundant species accounted for 65.8% of the total taxa. The salivary community was dominated by *Neisseria* spp., which accounted for 35.9% of the total microbiota. Thirteen species of *Neisseria* were detected, including *N. flavescens* (8.6 ± 4.6%), *N. mucosa* (6.6 ± 3.1%), and *N. meningitides* (5.2 ± 2.4%). Of the 20 most abundant species, 8 belonged to *Neisseria* spp., 4 belonged to *Streptococcus*, and the remaining species belonged to *Prevotella, Rothia, Haemophilus, Actinomyces, Veillonella, Lautropia* and *Porphyromonas*.

Differences in Salivary Microbiome Between Chronic Erythematous Candidiasis and Controls

Taxa at different taxonomic levels were presented as DIS *versus* HC to describe oral microbial community changes (**Figure 3A**). The microbial communities of HC and DIS were roughly separated in a principal coordinate's analysis based on differentially present taxa using the Bray-Curtis distance (**Figure 3B**), indicating the distinction between DIS and HC. At the class level, Actinobacteria and Bacilli were more enriched in DIS. At the genus level, *Rothia*, *Gemella* and *Streptococcus* were more enriched in DIS, while *Aggregatibacter*, *Campylobacter* and *Simonsiella* were more enriched in HC (**Figure 3A**). At the species level, 15 species were disease-enriched (*Rothia mucilaginosa* and *Streptococcus mitis* were the most abundant species), and 10 species were control-enriched (*Prevotella pallens* and *Campylobacter concisus* were the most abundant species) (Mann-Whitney U-test, $P < 0.05$) (**Figures 3C, D**). And it's worth noting that 9/15 disease-enriched species belonged to *Streptococci*.

C. albicans Showed Highly Correlated With *Streptococcus*

To analyze the relationship between *C. albicans* and oral bacteria, we constructed a salivary co-occurrence network based on taxonomic datasets of DIS, HC and cured DIS (**Figure S2**). The network consisted of 411 species (92% of the total) and 2625 correlations (edges); 97.2% of the species were positively correlated. We found that *C. albicans* formed densely connected modules with oral bacteria. Among them, 5 of the 8 species that directly correlated with *C. albicans* as well as 11 of the 25 species that indirectly correlated with *C. albicans* were members of *Streptococcus* respectively (Spearman correlation, $P < 0.05$; **Figure 4A**).



E Demographic, laboratorial and clinical information of the participants.

	Control group (n = 12)	Disease group (n = 12)	*P*-value*
Age (year)	60.17 ± 8.07	63.75 ± 7.88	0.283
Female (%)	12 (100%)	12 (100%)	
Colony count (per ml)	Not detectable	> 200	
Salivary pH	7.13 ± 0.47	6.77 ± 0.48	0.074
PD (mm)	2.97 ± 0.16	2.99 ± 0.13	0.681
DMFT	1.00 (0.00 ; 2.75)	0.00 (0.00 ; 2.50)	0.437
Missing tooth	0.00 (0.00 ; 1.00)	0.00 (0.00 ; 0.00)	0.402

FIGURE 1 | Clinical, laboratory and demographic characteristics of the study participants. **(A)** A clinical picture of the normal dorsum of tongue from a healthy individual. **(B)** A clinical picture of the dorsum of a tongue with chronic erythematous candidiasis from the disease group. Atrophy of the filiform papilla and erythema of the tongue could be observed. **(C)** Many hyphae (arrow) were detected by a smear test using optical microscopy (original magnification ×400). **(D)** A positive result of a salivary culture obtained from Sabouraud's agar. *Candida albicans* colonies (arrow) could be detected. **(E)** Demographic, laboratory and clinical information of the participants. *Independent t-test for age, salivary pH and PD; Mann-Whitney test for DMFT and missing tooth. The values are the means ± standard deviations, the medians (Q25; Q75) and the numbers of participants (percentage). PD, Probing depth; DMFT, Decayed, missing and filled tooth.

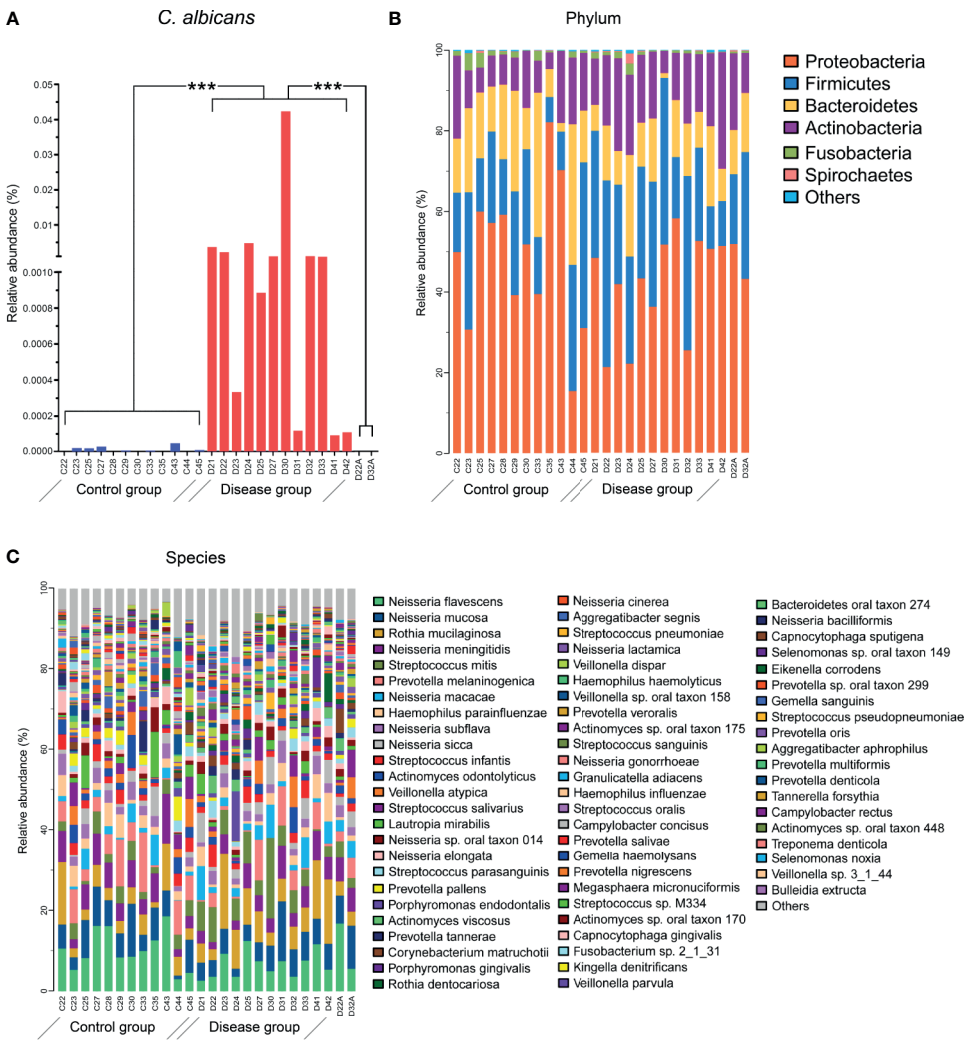


FIGURE 2 | Microbial compositions of salivary samples. **(A)** The relative abundance of *C. albicans* was significantly different between the HC and DIS groups. ***P <0.001. **(B)** The relative abundance levels of bacteria at the phylum levels. **(C)** The relative abundance levels of bacteria at the species levels. Sequence annotation and relative abundance estimation were performed with MOCAT.

Then, correlations between *C. albicans* and *Streptococcus* spp. were evaluated by a newly constructed co-occurrence network (Spearman correlation, $P < 0.05$; **Figure 4B**). Of the 34 *Streptococcus* species detected in saliva to our extent of sequencing, 33 were included in the network and were connected with *C. albicans*. *Streptococcus* spp. and *C. albicans* formed a network with densely connected nodes and formed 338 edges. All correlations between *Streptococcus* spp. and *C. albicans* were positive, indicating a potential mutually promotional relationship.

Differences in Functional Potentials Between Chronic Erythematous Candidiasis and Controls

To identify the functional role of salivary microbiota and its changes during *C. albicans* infection, we analyzed functional genes using the KEGG database. A total of 6495 KEGG orthologues (KOs) were identified, and the relative abundance levels of KOs were estimated (**Table S2**). The most abundant KOs identified from microbial communities in all samples included genes encoding Bacterial secretion system, ABC transport system, RNA polymerase, Genetic Information Processing, and Aminoacyl-tRNA biosynthesis.

Under the criteria (Mann-Whitney U-test, $P < 0.05$), 688 KOs (≈10% of the total we detected) showed significantly different abundance levels in DIS compared to those in HC metagenomes, with 202 enriched KOs and 486 depleted KOs in DIS (**Table S2**). As shown in **Figure 5**, KOs representing lipopolysaccharide biosynthesis system, reductive citrate cycle, bacterial ribosome were enriched in HC, while KOs representing eukaryotic ribosome, putative glutamine transport system, cytochrome bc1 complex respiratory unit were enriched in DIS. The results revealed that the genetic phenotypes of metabolism and other

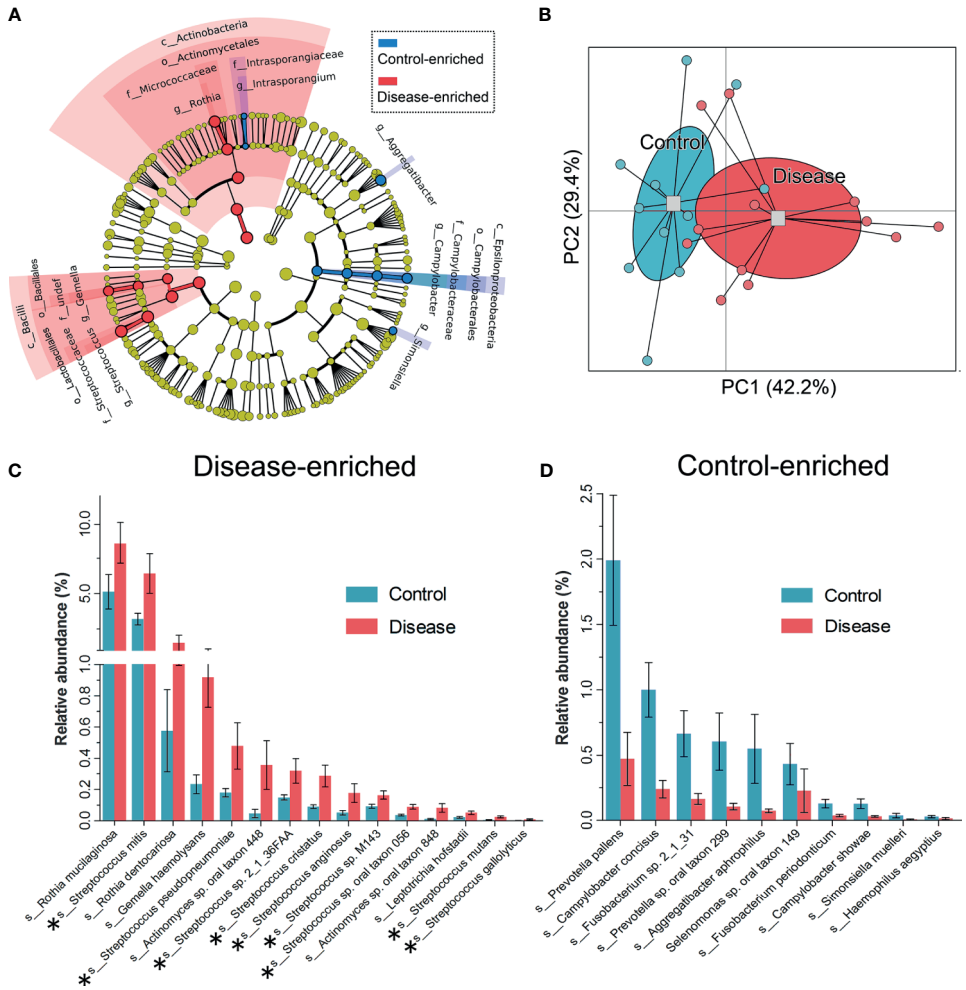


FIGURE 3 | Taxa differentiations between the HC and DIS group. **(A)** Taxa with significant differences in relative abundance at the phylum, class, order, family and genus levels are shown via taxonomic relationships. **(B)** Principal coordinates analysis based on the differentially present taxa using the Bray-Curtis distance. **(C)** Species enriched in the DIS group are shown, the bar represents the mean±sem. Members of *Streptococcus* are marked with "*". **(D)** Species enriched in the HC group are shown, the bar represents the mean±sem.

important physiological functions had changed significantly in oral microbiome of chronic erythematous candidiasis.

DISCUSSION

The Human Microbiome Project (HMP) has been carried out over ten years, and it was found that the complex and common host–microbiome interactions play an important role in the health and various diseases of human (Proctor et al., 2019). Oral microbiome is one of the most diverse microbiomes in human body. Though its composition has been explored in many oral diseases, there are still unexplored areas worth investigating (Wade, 2013). In the present study, we profiled the salivary microbiome of chronic erythematous candidiasis patients with the use of shotgun metagenomic sequencing technique and compared it with that of healthy participants. It was demonstrated that the composition and gene function of salivary microbiota had changed a lot during chronic erythematous candidiasis. What's more, a strong correlation between *C. albicans* and *Streptococcus* could be found. This project is a complement of the previous studies in oral microbiomes under the disease state. More importantly, it may provide some potential targets for the therapy and prognosis of chronic erythematous candidiasis.

Great attention has been paid to the relationship between *Candida* and oral microflora in different host states. Using 454 pyrosequencing, Kraneveld EA et al. revealed that oral *Candida* load was correlated negatively with the diversity of the salivary microbiome and positively with the relative abundance of *Streptococci* in older Dutch adults (Kraneveld et al., 2012). With 16s rRNA amplicon sequencing, Xiao J et al. found that *C. albicans* influenced the composition and diversity of oral bacterial community in the context of severe early childhood

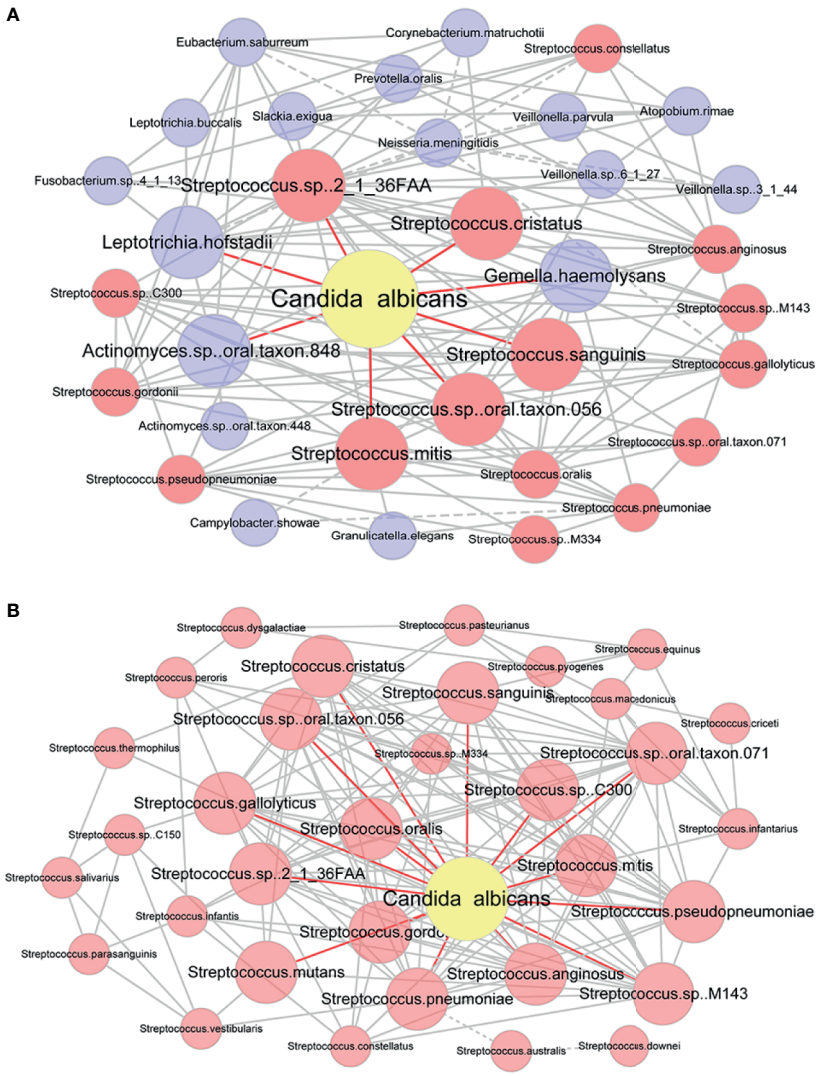


FIGURE 4 | Co-occurrence network of *C. albicans* and bacteria. **(A)** Co-occurrence network modules formed around *C. albicans*. Taxa that were directly correlated with *C. albicans* and those directly corrected with them were used to construct the network. Each pair of nodes connected by an edge was significantly and highly correlated (Spearman correlation test, $P < 0.05$, $|r| \geq 0.6$). Solid and dashed lines indicate positive and negative correlations. Taxa that were directly correlated with *C. albicans* are in bold. Nodes that represent members of *Streptococcus* are highlighted in red. **(B)** Co-occurrence network of *C. albicans* and *Streptococcus* spp. Each pair of nodes connected by an edge was significantly and highly correlated (Spearman correlation test, $P < 0.05$). Solid and dashed lines indicate positive and negative correlations. Taxa that were directly correlated with *C. albicans* are in bold.

caries (Xiao et al., 2018). By the use of 16s rRNA gene sequencing, Bertolini M et al. demonstrated that *C. albicans* infection was associated with loss of mucosal bacterial diversity in both oral and small intestinal mucosa in a mouse intravenous chemotherapy model (Bertolini et al., 2019). However, few studies explored the oral microbiome during chronic erythematous candidiasis without the presence of other factors that can affect the composition of microbiota.

Thus, in the present study, shotgun metagenomic sequencing technique was used to depict the salivary microbiome of chronic erythematous candidiasis patients. What's more, by adopting stringent inclusion criteria, many interference factors that can influence oral microbiota, such as wearing removable dentures and systemic diseases (O'Donnell et al., 2015), were eliminated. As a result, we collected a more valuable and reliable data on the interaction between *C. albicans* and oral microflora.

The composition of salivary microbiota showed significant difference between chronic erythematous candidiasis and healthy subjects. We found that *Prevotella rallens* was the most abundant species in healthy participants and significantly decreased in chronic erythematous candidiasis. It was reported that *Prevotella* spp. in all oral sites of clinically healthy individuals are part of the core microbiome (Ishaq et al., 2017). Therefore, changes in the composition of *Prevotella* genus may serve as a signal of a potential observational index for assessing or providing a prognosis for *C. albicans* infection and other unhealthy states.

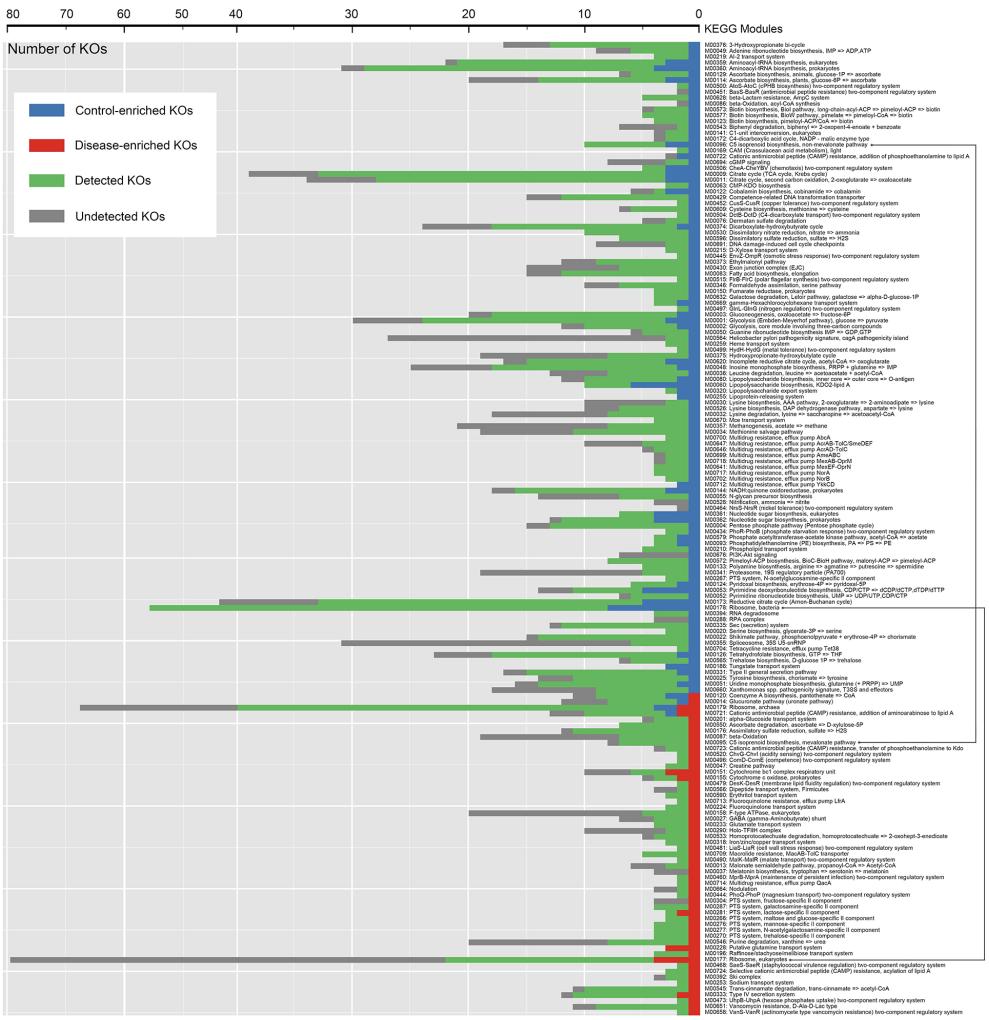


FIGURE 5 | Distribution of KEGG modules for HC-enriched and DIS-enriched KEGG orthologues (KOs). Each bar represents the number of KOs involved in each module. Green bars represent KOs detected in salivary samples to the extent of our sequencing. Blue and red bars represent KOs that were differentially detected in samples from healthy controls and patients with chronic erythematous candidiasis.

On the contrary, *Rothia mucilaginosa* and *Streptococcus mitis* were the most abundant disease-enriched species. Both bacteria are normal inhabitant of the human oral cavity and their abundance changed a lot in some abnormal status of oral cavity. It has been reported that the abundance of *Rothia mucilaginosa* was significantly increased in tongue leukoplakia lesions (Amer et al., 2017) and smokeless tobacco users (Halboub et al., 2020). And the abundance of oral *Streptococcus mitis* was reported to have changed a lot in periodontitis patients (Lundmark et al., 2019). In addition, the relative abundance of *Streptococcus*, *Rothia* and *Gemella* was increased in chronic erythematous candidiasis, but decreased after treatment (the 22nd and 32nd sample), which indicates that *Candida* infection has affected the native oral microbiome. Among them, the relative abundance of *Gemella* is also changed in an unhealthy oral cavity and is reported to be higher in erosive oral lichen planus (Yu et al., 2020). Hence, these microorganisms changing during chronic erythematous candidiasis may play potential significant roles in the disease process.

Therefore, we further analyzed the co-occurrence network between *C. albicans* and oral bacteria, and demonstrated the positive correlations between *Streptococcus* spp. and *C. albicans*. Although previous studies investigated the interactions between oral *Streptococcus* spp. and *C. albicans*, they had been focusing on certain species of *Streptococcus* (Ellepola et al., 2019; Abrantes and Africa, 2020). As what has been summarized in a review (Montelongo-Jauregui and Lopez-Ribot, 2018), 7 *Streptococcus* spp. associated with *C. albicans* were listed, all of them could be found in our study showing the correlations between *C. albicans* and 33 *Streptococcus* species. In addition, the results of this study can well represent the *in vivo* and the clinical environment, which is a supplement and extension to the previous studies in the influence of *C. albicans* infection on bacterial dysbiosis in mice model (Bertolini et al., 2019).

Some studies have also explored the underlying mechanisms of the correlations between *Streptococcus* spp. and *C. albicans*. *C. albicans* can positively influence the growth and biofilm formation of *Streptococcus* spp., through different but connected factors and mechanisms, such as fatty acids, carboxylic acids, farnesol and glucans (Metwalli et al., 2013; Ellepola et al., 2019). In addition, *C. albicans* can also induce the expression of virulence genes (e.g., *gtfB, fabM*) in *Streptococcus mutans* (Falsetta et al., 2014). On the other hand, *Streptococci* are considered to play an important role in establishing *C. albicans* colonization. *Streptococcus gordonii*, *Streptococcus oralis*, *Streptococcus sanguis* and *Streptococcus mutans* had been found to be structurally associated and co-aggregate with *C. albicans*, facilitated its yeast-to-hypha transition, and promoted its morphogenesis progress, adherence, colonization and biofilm formation through different mechanisms, such GlcNAc, lactate, H_2O_2, peptidoglycan, and Al-2 (Diaz et al., 2012; Metwalli et al., 2013; Dutton et al., 2016).

In our study, 9 of the 15 species enriched in DIS belonged to *Streptococcus* spp., and the co-occurrence network analysis revealed positive correlations between *Streptococcus* spp. and *C. albicans*. In order to support this result, we carried out an *in vivo* mice experiment. The mice were infected on their tongue with *C. albicans* only or co-infected with *C. albicans* and *Streptococcus mutans* for 5 days. We found that co-infection with *C. albicans* and *Streptococcus mutans* triggered significantly greater weight loss than infection with *C. albicans* only, and *Streptococcus mutans* enhanced the colonization of *C. albicans* on mice tongues (**Figure S3**). The above results indicated the interactions between *Streptococcus* spp. and *C. albicans* may serve as an important role in the occurrence and development of chronic erythematous candidiasis. However, more detailed animal or *in vitro* experiments about the relationships between *C. albicans* and key bacteria changed in chronic erythematous candidiasis should be carried out in the future.

In addition, the gene function of salivary microbiome was different between chronic erythematous candidiasis and healthy subjects. It was obvious that the putative glutamine transport system and cytochrome bc1 complex respiratory unit were enriched in DIS. Glutamine transport system is an important part of the nitrogen assimilation metabolism of microorganism. Nitrogen and nicotinate/nicotinamide metabolic pathways had been confirmed to be involved in *C. albicans* morphogenesis, such as filament formation which plays an important role in penetrating endothelial tissue (Han et al., 2019). Cytochrome bc1 (Complex III) is an important part of the electron respiratory transport chain in organisms. The disruption of electron transport chain function increased intracellular levels of reactive oxygen species in yeast. And, the inhibition of cytochrome bc1 could significantly increase the sensitivity of *C. albicans* to photodynamic therapy (Chabrier-Rosello et al., 2010). Basing on the changes in energy metabolism, respiratory transport chain and other biological functions of oral microbiome, we might find some new drug targets for treating chronic erythematous candidiasis.

In conclusion, our study revealed the oral microbial changes occurring in chronic erythematous candidiasis. Close relationships were seen between *C. albicans* and oral bacteria, especially *Streptococcus*. In addition, functional potentials of oral microbiome also changed during chronic erythematous candidiasis, which may provide us with some new insights into the role of oral microecology in pathological and clinical manifestations of chronic erythematous candidiasis.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the China National GeneBank (CNGB) repository, accession number CNP0001917.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Peking University Health Science Center. The patients/participants provided their written informed consent to participate in this study. The animal study was reviewed and approved by Ethics Committee of the Peking University Health Science Center. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

XL: Writing original draft, sample collection, experimentation, data analysis, and data visualization. HuZ: Data analysis, Data visualization. XW: Data analysis, Data visualization. HeZ: Data analysis, Data visualization. LG: Data visualization. ZX: Experimentation. QZ: Experimentation. XH: Data analysis. HH: Sample collection. ZY: Conceptualization, supervision, data interpretation, data analysis, and sample collection. FC: Supervision, data interpretation, and experimentation. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financially supported by the National Natural Science Foundation of China (Nos. 81000441, 81570985 and 81991501).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2021.691092/full#supplementary-material

REFERENCES

Abrantes, P., and Africa, C. (2020). Measuring *Streptococcus Mutans*, *Streptococcus Sanguinis* and *Candida Albicans* Biofilm Formation Using a Real-Time Impedance-Based System. *J. Microbiol. Methods* 169, 105815. doi: 10.1016/j.mimet.2019.105815

Amer, A., Galvin, S., Healy, C. M., and Moran, G. P. (2017). The Microbiome of Potentially Malignant Oral Leukoplakia Exhibits Enrichment for *Fusobacterium*, *Leptotrichia*, Campylobacter, and *Rothia* Species. *Front. Microbiol.* 8, 2391. doi: 10.3389/fmicb.2017.02391

Baker, J. L., Morton, J. T., Dinis, M., Alvarez, R., Tran, N. C., Knight, R., et al. (2021). Deep Metagenomics Examines the Oral Microbiome During Dental Caries, Revealing Novel Taxa and Co-Occurrences With Host Molecules. *Genome Res.* 31 (1), 64–74. doi: 10.1101/gr.265645.120

Bartnicka, D., Karkowska-Kuleta, J., Zawrotniak, M., Satala, D., Michalik, K., Zielinska, G., et al. (2019). Adhesive Protein-Mediated Cross-Talk Between *Candida Albicans* and *Porphyromonas Gingivalis* in Dual Species Biofilm Protects the Anaerobic Bacterium in Unfavorable Oxic Environment. *Sci. Rep.* 9 (1), 4376. doi: 10.1038/s41598-019-40771-8

Bertolini, M., Ranjan, A., Thompson, A., Diaz, P. I., Sobue, T., Maas, K., et al. (2019). *Candida Albicans* Induces Mucosal Bacterial Dysbiosis That Promotes Invasive Infection. *PloS Pathog.* 15 (4), e1007717. doi: 10.1371/journal.ppat.1007717

Cavalcanti, I. M., Nobbs, A. H., Ricomini-Filho, A. P., Jenkinson, H. F., and Del, B. C. A. (2016). Interkingdom Cooperation Between *Candida Albicans*, *Streptococcus Oralis* and *Actinomyces Oris* Modulates Early Biofilm Development on Denture Material. *Pathog. Dis.* 74 (3), ftw002. doi: 10.1093/femspd/ftw002

Chabrier-Rosello, Y., Giesselman, B. R., De Jesus-Andino, F. J., Foster, T. H., Mitra, S., and Haidaris, C. G. (2010). Inhibition of Electron Transport Chain Assembly and Function Promotes Photodynamic Killing of *Candida*. *J. Photochem. Photobiol. B.* 99 (3), 117–125. doi: 10.1016/j.jphotobiol.2010.03.005

Coronado-Castellote, L., and Jimenez-Soriano, Y. (2013). Clinical and Microbiological Diagnosis of Oral Candidiasis. *J. Clin. Exp. Dent.* 5 (5), e279–e286. doi: 10.4317/jced.51242

Davies, A. N., Brailsford, S. R., Beighton, D., Shorthose, K., and Stevens, V. C. (2008). Oral Candidosis in Community-Based Patients With Advanced Cancer. *J. Pain Symptom Manage.* 35 (5), 508–514. doi: 10.1016/j.jpainsymman.2007.07.005

Diaz, P. I., Xie, Z., Sobue, T., Thompson, A., Biyikoglu, B., Ricker, A., et al. (2012). Synergistic Interaction Between *Candida Albicans* and Commensal Oral *Streptococci* in a Novel *In Vitro* Mucosal Model. *Infect. Immun.* 80 (2), 620–632. doi: 10.1128/IAI.05896-11

Dutton, L. C., Paszkiewicz, K. H., Silverman, R. J., Splatt, P. R., Shaw, S., Nobbs, A. H., et al. (2016). Transcriptional Landscape of Trans-Kingdom Communication Between *Candida Albicans* and *Streptococcus Gordonii*. *Mol. Oral Microbiol.* 31 (2), 136–161. doi: 10.1111/omi.12111

Ellepola, K., Truong, T., Liu, Y., Lin, Q., Lim, T. K., Lee, Y. M., et al. (2019). Multi-Omics Analyses Reveal Synergistic Carbohydrate Metabolism in *Streptococcus Mutans-Candida Albicans* Mixed-Species Biofilms. *Infect. Immun.* 87 (10), e00339–19. doi: 10.1128/IAI.00339-19

Falsetta, M. L., Klein, M. I., Colonne, P. M., Scott-Anne, K., Gregoire, S., Pai, C. H., et al. (2014). Symbiotic Relationship Between *Streptococcus Mutans* and *Candida Albicans* Synergizes Virulence of Plaque Biofilms *In Vivo*. *Infect. Immun.* 82 (5), 1968–1981. doi: 10.1128/IAI.00087-14

Halboub, E., Al-Ak'Hali, M. S., Alamir, A. H., Homeida, H. E., Baraniya, D., Chen, T., et al. (2020). Tongue Microbiome of Smokeless Tobacco Users. *BMC Microbiol.* 20 (1), 201. doi: 10.1186/s12866-020-01883-8

Han, T. L., Cannon, R. D., Gallo, S. M., and Villas-Boas, S. G. (2019). A Metabolomic Study of the Effect of *Candida Albicans* Glutamate Dehydrogenase Deletion on Growth and Morphogenesis. *NPJ Biofilms Microbiomes* 5 (1), 13. doi: 10.1038/s41522-019-0086-5

Hu, L., Zhou, P., Zhao, W., Hua, H., and Yan, Z. (2020). Fluorescence Staining vs. Routine KOH Smear for Rapid Diagnosis of Oral Candidiasis-A Diagnostic Test. *Oral. Dis.* 26 (5), 941–947. doi: 10.1111/odi.13293

Ishaq, H. M., Mohammad, I. S., Guo, H., Shahzad, M., Hou, Y. J., Ma, C., et al. (2017). Molecular Estimation of Alteration in Intestinal Microbial Composition in Hashimoto's Thyroiditis Patients. *BioMed. Pharmacother.* 95, 865–874. doi: 10.1016/j.biopha.2017.08.101

Kraneveld, E. A., Buijs, M. J., Bonder, M. J., Visser, M., Keijser, B. J., Crielaard, W., et al. (2012). The Relation Between Oral *Candida* Load and Bacterial Microbiome Profiles in Dutch Older Adults. *PloS One* 7 (8), e42770. doi: 10.1371/journal.pone.0042770

Li, Y., He, J., He, Z., Zhou, Y., Yuan, M., Xu, X., et al. (2014). Phylogenetic and Functional Gene Structure Shifts of the Oral Microbiomes in Periodontitis Patients. *ISME J.* 8 (9), 1879–1891. doi: 10.1038/ismej.2014.28

Lopez-Pintor, R. M., Casanas, E., Gonzalez-Serrano, J., Serrano, J., Ramirez, L., de Arriba, L., et al. (2016). Xerostomia, Hyposalivation, and Salivary Flow in Diabetes Patients. *J. Diabetes Res.* 2016, 4372852. doi: 10.1155/2016/4372852

Lundmark, A., Hu, Y., Huss, M., Johannsen, G., Andersson, A. F., and Yucel-Lindberg, T. (2019). Identification of Salivary Microbiota and its Association With Host Inflammatory Mediators in Periodontitis. *Front. Cell Infect. Microbiol.* 9, 216. doi: 10.3389/fcimb.2019.00216

Metwalli, K. H., Khan, S. A., Krom, B. P., and Jabra-Rizk, M. A. (2013). *Streptococcus Mutans*, *Candida Albicans*, and the Human Mouth: A Sticky Situation. *PloS Pathog.* 9 (10), e1003616. doi: 10.1371/journal.ppat.1003616

Millsop, J. W., and Fazel, N. (2016). Oral Candidiasis. *Clin. Dermatol.* 34 (4), 487–494. doi: 10.1016/j.clindermatol.2016.02.022

Montelongo-Jauregui, D., and Lopez-Ribot, J. L. (2018). *Candida* Interactions With the Oral Bacterial Microbiota. *J. Fungi (Basel)* 4 (4), 122. doi: 10.3390/jof4040122

O'Donnell, L. E., Robertson, D., Nile, C. J., Cross, L. J., Riggio, M., Sherriff, A., et al. (2015). The Oral Microbiome of Denture Wearers Is Influenced by Levels of Natural Dentition. *PloS One* 10 (9), e0137717. doi: 10.1371/journal.pone.0137717

Poncet, S., Milohanic, E., Maze, A., Abdallah, J. N., Ake, F., Larribe, M., et al. (2009). Correlations Between Carbon Metabolism and Virulence in Bacteria. *Contrib. Microbiol.* 16, 88–102. doi: 10.1159/000219374

Proctor, L. M., Creasy, H. H., Fettweis, J. M., Lloyd-Price, J., Mahurkar, A., and Zhou, W. (2019). The Integrative Human Microbiome Project. *Nature* 569, 641–648. doi: 10.1038/s41586-019-1238-8

Sarkar, P., Malik, S., Laha, S., Das, S., Bunk, S., Ray, J. G., et al. (2021). Dysbiosis of Oral Microbiota During Oral Squamous Cell Carcinoma Development. *Front. Oncol.* 11, 614448. doi: 10.3389/fonc.2021.614448

Sunagawa, S., Mende, D. R., Zeller, G., Izquierdo-Carrasco, F., Berger, S. A., Kultima, J. R., et al. (2013). Metagenomic Species Profiling Using Universal Phylogenetic Marker Genes. *Nat. Methods* 10 (12), 1196–1199. doi: 10.1038/nmeth.2693

Vila, T., Sultan, A. S., Montelongo-Jauregui, D., and Jabra-Rizk, M. A. (2020). Oral Candidiasis: A Disease of Opportunity. *J. Fungi (Basel)* 6 (1), 15. doi: 10.3390/jof6010015

Wade, W. G. (2013). The Oral Microbiome in Health and Disease. *Pharmacol. Res.* 69 (1), 137–143. doi: 10.1016/j.phrs.2012.11.006

Wu, T., Cen, L., Kaplan, C., Zhou, X., Lux, R., Shi, W., et al. (2015). Cellular Components Mediating Coadherence of *Candida Albicans* and *Fusobacterium Nucleatum*. *J. Dent. Res.* 94 (10), 1432–1438. doi: 10.1177/0022034515593706

Xiao, J., Grier, A., Faustoferri, R. C., Alzoubi, S., Gill, A. L., Feng, C., et al. (2018). Association Between Oral Candida and Bacteriome in Children With Severe ECC. *J. Dent. Res.* 97 (13), 1468–1476. doi: 10.1177/0022034518790941

Yu, F. Y., Wang, Q. Q., Li, M., Cheng, Y. H., Cheng, Y. L., Zhou, Y., et al. (2020). Dysbiosis of Saliva Microbiome in Patients With Oral Lichen Planus. *BMC Microbiol.* 20 (1), 75. doi: 10.1186/s12866-020-01733-7

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Cellular and Infection Microbiology

ORIGINAL RESEARCH
published: 18 August 2021
doi: 10.3389/fcimb.2021.728933

The Clinical Potential of Oral Microbiota as a Screening Tool for Oral Squamous Cell Carcinomas

Xinxuan Zhou[1†], Yu Hao[1,2†], Xian Peng[1†], Bolei Li[1,2†], Qi Han[1], Biao Ren[1], Mingyun Li[1], Longjiang Li[1], Yi Li[1], Guo Cheng[3,4], Jiyao Li[1,2], Yue Ma[3], Xuedong Zhou[1,2*] and Lei Cheng[1,2*]*

[1] State Key Laboratory of Oral Diseases & West China Hospital of Stomatology & National Clinical Research Center for Oral Diseases, Sichuan University, Chengdu, China, [2] Department of Operative Dentistry and Endodontics, West China Hospital of Stomatology, Sichuan University, Chengdu, China, [3] West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, China, [4] Laboratory of Molecular Translational Medicine, Centre for Translational Medicine, Key Laboratory of Birth Defects and Related Diseases of Women and Children, Ministry of Education, West China Second University Hospital, Sichuan University, Chengdu, China

Introduction: The oral squamous cell carcinoma (OSCC) is detrimental to patients' physical and mental health. The prognosis of OSCC depends on the early diagnosis of OSCC in large populations.

Objectives: Here, the present study aimed to develop an early diagnostic model based on the relationship between OSCC and oral microbiota.

Methods: Overall, 164 samples were collected from 47 OSCC patients and 48 healthy individuals as controls, including saliva, subgingival plaque, the tumor surface, the control side (healthy mucosa), and tumor tissue. Based on 16S rDNA sequencing, data from all the five sites, and salivary samples only, two machine learning models were developed to diagnose OSCC.

Results: The average diagnostic accuracy rates of five sites and saliva were 98.17% and 95.70%, respectively. Cross-validations showed estimated external prediction accuracies of 96.67% and 93.58%, respectively. The false-negative rate was 0%. Besides, it was shown that OSCC could be diagnosed on any one of the five sites. In this model, *Actinobacteria*, *Fusobacterium*, *Moraxella*, *Bacillus*, and *Veillonella* species exhibited strong correlations with OSCC.

Conclusion: This study provided a noninvasive and inexpensive way to diagnose malignancy based on oral microbiota without radiation. Applying machine learning methods in microbiota data to diagnose OSCC constitutes an example of a microbial assistant diagnostic model for other malignancies.

Keywords: oral microbiota, OSCC, machine learning methods, diagnose, sequencing

OPEN ACCESS

Edited by:
Naile Dame-Teixeira, University of Brasilia, Brazil

Reviewed by:
Indranil Chattopadhyay, Central University of Tamil Nadu, India
Vivian Petersen Wagner, The University of Sheffield, United Kingdom

***Correspondence:**
Lei Cheng
chenglei@scu.edu.cn
Xuedong Zhou
zhouxd@scu.edu.cn
Yue Ma
Gordonrozen@qq.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Microbiome in Health and Disease, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: *22 June 2021*
Accepted: *27 July 2021*
Published: *18 August 2021*

Citation:
Zhou X, Hao Y, Peng X, Li B, Han Q, Ren B, Li M, Li L, Li Y, Cheng G, Li J, Ma Y, Zhou X and Cheng L (2021) The Clinical Potential of Oral Microbiota as a Screening Tool for Oral Squamous Cell Carcinomas. Front. Cell. Infect. Microbiol. 11:728933. doi: 10.3389/fcimb.2021.728933

INTRODUCTION

Oral cancer is a significant threat to patients' physical and mental health. According to the Global Cancer Statistics (Bray et al., 2018), an estimated 350,000 new cases and 170,000 deaths from oral cavity cancers occurred in 2018. Most global oral squamous cell carcinoma (OSCC) cases are diagnosed in Asia. In developing countries, in particular, oral cancers rank the eighth most common cancers in males. Worryingly, the incidence of the oral cavity cancers appears to be increasing in many parts of the world (Simard et al., 2014). The most common oral cancer is OSCC, with a 95% rate. The prognosis for oral cancers is notably poor, with a mean all-stage, 5-year survival rate of <50% (Kujan et al., 2005).

Therefore, it is essential to diagnose OSCC at an early stage, especially in large populations, and the prognosis of the treatment could benefit from the early detection of OSCC. In the diagnosis of OSCC and many other tumors, pathologic diagnosis is the gold standard, and radiologic examinations provide useful supplementary data. However, it is difficult to apply these traditional methods as primary diagnostic methods for OSCC in large populations due to their invasive, radioactive, and expensive nature. Therefore, an effective, convenient, and noninvasive method is necessary as a screening tool for OSCC in large populations.

In recent years, many investigations have explored the association between oral bacteria and OSCC (Ahn et al., 2012; Pushalkar et al., 2012; Schmidt et al., 2014). Therefore, oral bacteria might be a potential biomarker to develop a promising early diagnostic method for OSCC. However, we still face considerable challenges in developing a novel diagnostic model based on oral bacteria. First, efforts are underway to find out the core microbiome or species for OSCC diagnosis. Previous studies have investigated the relationship between some single species and OSCC, including *Porphyromonas gingivalis* (Chang et al., 2019; Park et al., 2019; de Mendoza et al., 2020) and *Staphylococcus aureus* (Wang et al., 2019). Investigators have also indicated the differences in the oral microbiome between OSCC patients and healthy individuals *via* bioinformatics analysis. Some other previous studies have indicated significant losses in the richness and diversity of oral microbiota in OSCC patients compared with healthy subjects. The relative frequencies of *Streptococcus*, *Dialister*, and *Veillonella* species differentiate the tumor from a healthy state (Guerrero-Preston et al., 2016). Other studies (Krogh et al., 1987) found significantly higher frequencies of *Porphyromonas*, *Actinomycetes*, *Haemophilus*, and *Enterobacter* species on the surface of OSCC tissues. Hooper et al. demonstrated that microbial diversity increased in tumor tissues by using 16S rDNA sequencing technology (Hooper et al., 2007). However, the exact core microbiome remains unclear, and thus, diagnostic models were not established to detect OSCC based on the microbiome.

Second, the oral cavity is a complicated environment, and the microbiome is different in different sites, including the tongue, teeth, mucous membranes, palate, and gums (Aas et al., 2005; Avila et al., 2009; Zarco et al., 2012). Segata et al. reported that the composition of microbial communities varies in seven oral cavity surfaces, demonstrating that the buccal mucosa, keratinized gingiva, hard palate, saliva, tongue, tonsils, throat, and subgingival and supragingival plaques were distinct more or less (Segata et al., 2012). Therefore, it is necessary and vital to determine which site should be selected to analyze the microbiome for OSCC diagnosis.

This study showed that OSCC could be diagnosed based on oral microbiota, and a diagnostic model could be developed with the help of machine learning methods. Moreover, the microbiota in the saliva, subgingival plaque, tumor surface, the control side (normal mucosa), and intratumoral tissue were useful for OSCC diagnosis. What is more, this diagnostic model can effectively avoid missed diagnoses; therefore, it is a potential early OSCC diagnostic method for large populations.

MATERIALS AND METHODS

Study Design

This study consisted of three stages. In stage I, the demographic data and microbiome were characterized using descriptive methods to provide a clear profile of both internal and external samples and the whole study data. Also, the microbiome and demographic data were analyzed using exploratory methods to test the study assumption, i.e., whether OSCC patients have microbiome patterns different from those of healthy people. In stage II, random forests were developed to determine the different patterns and further analyze the specified operational taxonomic unit (OTU) role in the differences between microbiome patterns of healthy and OSCC individuals. In stage III, *post hoc* analyses were carried out to evaluate the different aspects of the performance of the diagnostic model developed in stage II, i.e., external discrimination capacity and its reliability on the sample size of the random forest prediction model based on the oral microbiome.

Participant Information

The institutional review board of the West China Hospital Stomatology of Sichuan University approved the study (approval number: WCHSIRB-D-2013-047). All the patients provided written informed consent forms before sample collection.

The sample collection protocol conformed to the Manual of Procedure for Human Microbiome Project Core Microbiome Sampling Protocol A HMP Protocol #07-001 (McInnes and Cutting, 2010; Segata et al., 2012; Sturød et al., 2020). There were 47 OSCC patients, all from China, who met the inclusion criteria, which required the use of no alcohol, no tobacco, no antibiotics, no cortisone, no cytokines (which could provoke the immune system like interleukin), and no immunosuppressant drugs like methotrexate six months before the sampling procedure. The age of the patients ranged from 34 to 78 years. The patients with DMF >4, calculus index ≥2, and oral fungal or mucosal diseases were excluded (Kalogirou and Sklavounou-Andrikopoulou, 2017; Xun et al., 2018). The control group followed the same criteria.

All patients were sampled before treatment to ensure that the microbiome was not affected by chemotherapy, radiotherapy, and oral prophylaxis. Of the 47 OSCC patients, 47 salivary samples (the saliva group), 18 subgingival plaque samples (the

pla group), 21 surfaces of tumor samples (the tum–muc group), 16 control side of healthy mucosa samples (the con–muc group), and 16 tumor tissue samples (the tum group) were collected (**Table 1**). OTU composition is a whole community structure that reflects various conditions of the microenvironment, and it is affected by factors such as diet, nutrition, and living habits. Therefore, the OTU composition of samples from different regions might be significantly different. Therefore, if this factor is not eliminated and only local or single-source samples are selected for the construction of the model, regional differences might cover it when applied to the population in other regions, resulting in unsatisfactory prediction performance. Forty-six healthy individuals were included as a control, consisting of 21 salivary samples from the same region as patients in Sichuan Province, and 25 salivary samples from another center, Peking University Hospital of Stomatology (Xun et al., 2018), in Beijing, to avoid this error (**Table 1**).

Sample Collection

The participants were asked not to take in any food and not brush or floss for at least 12 h before the sample collection session. The protocol for sample collection in each site followed the Manual of Procedure for Human Microbiome Project: Core Microbiome Sampling Protocol A (HMP Protocol #07-001) (McInnes and Cutting, 2010; Segata et al., 2012; Sturød et al., 2020). The participants were taught to stop swallowing for 1 min and collect 5 ml of saliva in 50-ml Falcon tubes for saliva collection. For plaque collection, buccal swabs were used to take plaque samples from the participants, which were stored in 2-ml EP tubes. For the bacterial flora on the oral mucosa, swabs were used to wipe the lesion and the other side of the oral mucosa for 10 s, respectively, avoiding the tooth and internal tumor. All the samples were then transferred into phosphate-buffered saline (PBS) solution and stored at −80°C immediately. For the internal tumor, dental instruments were disinfected to cut the internal tumor into 1 × 1 × 1-cm^3 cubes on a sterile platform; the tumor samples were then steeped in sterile povidone-iodine for 3 min and vortexed several times using 500 μl of PBS. The tumor samples were divided into two parts, with one being steeped in Tris-EDTA buffer (pH = 7.4) stored at −80°C and with the other one being used for cultivation (McInnes and Cutting, 2010).

DNA Extraction and PCR Amplification

Microbial DNA was extracted from all the samples using the E.Z.N.A.® soil DNA Kit (Omega Bio-Tek, Norcross, GA, USA) according to the manufacturer's protocols (Zhu et al., 2015; Wang et al., 2016; Li et al., 2019). The V4–V5 region of the bacterial 16S ribosomal RNA gene was amplified by PCR (95°C for 2 min, followed by 25 cycles at 95°C for 30 s, 55°C for 30 s, and 72°C for 30 s, and a final extension at 72°C for 5 min) using primers 515F 5′-barcode-GTGCCAGCMGCCGCGG)-3′ and 907R 5′-CCGTCAATTCMTTTRAGTTT-3′ (Li et al., 2018; Xie et al., 2018; Zhou et al., 2018), where the barcode is an eight-base sequence unique to each sample. PCRs were performed in triplicate in a 20-μl mixture containing 4 μl of 5× FastPfu buffer, 2 μl of 2.5 mM of dNTPs, 0.8 μl of each primer (5 μM), 0.4 μl of FastPfu polymerase, and 10 ng of template DNA.

TABLE 1 | Samples in different groups.

	OSCC (n = 47)	Healthy control (n = 21)	External set (n = 25)	Total (n = 93)
con_muc	16			16
Pla	18			18
Saliva	47	21	25	93
Tum	16			16
Tum_muc	21			21
Overall	118	21	25	164

OSCC, oral squamous cell carcinoma.

Illumina MiSeq Sequencing

Amplicons were extracted from 2% agarose gels, purified using the AxyPrep DNA Gel Extraction Kit (Axygen Biosciences, Union City, CA, USA) according to the manufacturer's instructions, and quantified using QuantiFluor™-ST (Promega, USA) (Xie et al., 2018). According to the standard protocols, purified amplicons were pooled in equimolar and paired-end sequenced (2 × 300) on an Illumina MiSeq platform. The Sequencing Depth of all samples was enough for analysis. The rarefaction analysis and read count statistics of all samples are shown in the **Supplementary Material** (**Figure S2** and **Table S4**). The raw reads were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database (Accession Number: SRP119028) (Zhu et al., 2015; Wang et al., 2016; Yin et al., 2016; Xie et al., 2018).

Processing of Sequencing Data

Raw FASTQ files were demultiplexed and quality-filtered using QIIME (Version 1.9.1) with the following criteria (Caporaso et al., 2010): i) the 300-bp reads were truncated at any site with an average quality score of <20 over a 50-bp sliding window, discarding the truncated reads that were shorter than 50 bp. ii) Exact barcode matching, two nucleotide mismatches in primer matching, and reads containing ambiguous characters were removed. iii) Only sequences that overlapped longer than 10 bp were assembled according to their overlap sequence. Reads that could not be assembled were discarded.

OTUs were clustered with 97% similarity cutoff using UPARSE Version 7.1 (http://drive5.com/uparse/), and chimeric sequences were identified and removed using UCHIME. The taxonomy of each 16S rRNA gene sequence was analyzed by RDP Classifier (Cole et al., 2005) (http://rdp.cme.msu.edu/) against the silva (SSU123) 16S rRNA database using a confidence threshold of 70% (Dewhirst et al., 2010).

Statistical Analysis

In stage I, the demographic and microbiome characteristics of the subjects were presented. An exploratory analysis was carried out to explore the potential capacity of pattern differences between samples from healthy and OSCC individuals. The Shannon index, Chao index, Simpson diversity index

(Chao et al., 1992; Chao and Shen, 2003), beta diversity index, network analysis, and functional analysis were used to explore whether the microbiome profiles in samples differed between OSCC and healthy individuals.

In stage II, since exploratory analysis showed that microbiome patterns differed between OSCC and healthy individuals, random forests were developed to show that such a pattern of the total microbiome in healthy subjects was different from that in OSCC individuals. The proper discriminations of this algorithm in high-dimensional datasets have been shown in various fields. Based on the model, the OTUs with great importance in distinguishing OSCC patients from healthy individuals were also extracted to provide clues for further studies on the mechanism of interaction of microbiome and cancer incidence.

In stage III, further analyses were carried out on the prediction model based on the random forests to evaluate the external prediction capacity and the dependence on the sample size.

To evaluate external prediction capacity, although the algorithms based on CART, bagging, and bootstrap have strong resistance against the overfitting, still in practice in some cases, such prediction models cannot perform well in external datasets. Therefore, a batch of cross-validations was carried out. In each cross-validation, a fixed proportion of samples was first randomly selected as the training set to build random forests. The rest of the samples used to test the forests' prediction capacity were used to predict whether the forests could correctly discriminate the OSCC patients from healthy individuals in the external population. Given that all the samples in the test set would not be used to train the forests, each time, the forests were tested using an external validation set. This process was repeated for large numbers to ensure that each sample would be in training and test sets for at least once. The average performance over the tests would be used to evaluate the expected external discrimination capacity of OSCC patients using random forests based on the microbiome.

As in cross-validation, not all the samples would be used to train the model, and the prediction capacity would decrease due to the loss of sample size. Therefore, it is of interest how many samples can build a reliable prediction model and whether the prediction capacity can be improved by introducing more samples. Therefore, different batches of cross-validations with different sample sizes of the training set were carried out to evaluate how the prediction capacity changes in terms of the sample size.

RESULTS

Characteristics and Exploratory Analysis

The diversity indexes, i.e., Shannon, Chao, and Simpson indexes, showed that the diversities of oral microbiome increased significantly in OSCC patients compared with healthy individuals (**Figure 1A**).

A Venn graph (**Figure 1B**) was used to determine the number of common and distinguished OTUs between OSCC patients and healthy controls. Samples with similar levels of 97% OTU were used for the analysis. OSCC patients and the healthy group exhibited significant differences in the OTU level, with only 428 out of 1,747 OTUs in common; 311 of 1,747 OTUs were unique for OSCC patients.

The Bray–Curtis principal coordinate analysis (PCoA) showed that healthy individuals' microbial community was concentrated, while the microbial community of patients was relatively discrete. Besides, the microbiome in samples from both OSCC and healthy individuals from the same center (West China College of Stomatology), i.e., OSCC and healthy control group, was similar. In contrast, those from the external center (Peking University Hospital of Stomatology) exhibited a different pattern (**Figure 1C**). This result supported our suspicion that microbiome profiles might differ significantly between different populations from different regions rather than those between OSCC and healthy individuals. Therefore, if the prediction model were built only with samples from a local or internal set of samples, its generalizability would be significantly limited, and the application of such a prediction model to external populations might be inappropriate. This is also evaluated by external prediction evaluation in stage III.

The key OTU phylotypes in OSCC patients and the healthy group were analyzed, which showed different phyla in the two groups. Five locations (saliva, subgingival plaque, tumor surface, normal mucosa in the control side, and intratumoral tissue) were sampled to investigate the frequencies of oral microbial communities in OSCC patients. All the results are presented in the **Supplementary Material**. This raised the interesting question of whether different sampling sites affected the model diagnosis.

These exploratory results implied that the microbiome pattern between the healthy and OSCC subjects was significantly different. The significant differences suggested that the oral microbiome does have the potential capacity to discriminate the OSCC patients from all the individuals.

Phylogenetic Profiles of Oral Microbial Communities in Oral Squamous Cell Carcinoma Patients

We examined the similarities and differences of genera present in the healthy group and as depicted in **Figures 2A, B**. Phylotypes with a median relative abundance larger than 0.01% of total abundance were included for comparison. To identify key OTU phylotypes in OSCC patients and healthy group, abundances of OTUs were analyzed by Wilcoxon's rank-sum test with the Benjamini–Hochberg method.

The OTUs representing different phyla were not similar between the two groups. The healthy group was observed to contain *Streptococcus* (22.73%), followed by *Neisseria* (18.23%), *Prevotella* (14.56%), *Porphyromonas* (7.33%), *Haemophilus* (6.72%), and *Veillonella* (4.05%). The OSCC group was found to contain *Streptococcus* (11.09%), followed by *Neisseria*

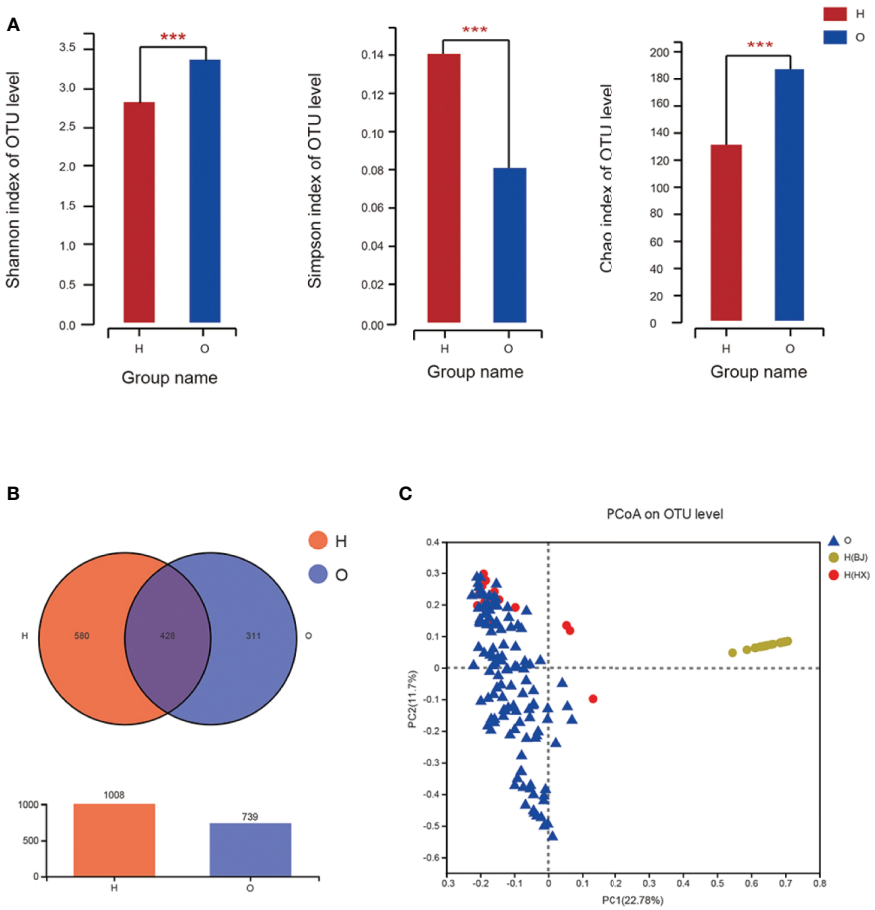


FIGURE 1 | The diversities of oral microbiome in OSCC patients and healthy individuals. **(A)** Shannon, Chao and Simpson Indexes of all OTUs with relative importance greater than 0. 01 between OSCC patients and healthy control. **(B)** Venn graph between OSCC patients and healthy control. **(C)** PCoA of bray_curtis between OSCC and healthy individuals. O represented OSCC patients, H represented healthy people, and H(HX) means people from West China College of Stomatology, H(BJ) means people from Peking University Hospital of Stomatology (* means $0.01 < P \leq 0.05$, ** means $0.001 < P \leq 0.01$, *** means $P \leq 0.001$).

(12.88%), *Prevotella* (12.03%), *Porphyromonas* (4.18%), *Haemophilus* (3.35%), and *Veillonella* (6.01%).

The stacked column plots also showed the differences between the two groups in terms of phylum (**Figure 3A**), class (**Figure 3B**), order (**Figure 3C**), family (**Figure 3D**), genus (**Figure 3E**), and species (**Figure 3F**). Overall, the abundance of the OSCC group was higher than that of the healthy group. On the phylum level, there were less Bacteroidetes and Proteobacteria and more Firmicutes and Fusobacteria in the OSCC group. On the class level, there were less Bacilli, Bacteroidia, and Betaproteobacteria and more Negativicutes in the OSCC group. On the order level, there were less Bacteroidales, Lactobacillales, and Neisseriales and more Selenomonadales in the OSCC group. On the family level, there were less Streptococcaceae, Prevotellaceae, and Neisseriaceae and more Veillonellaceae in the OSCC group. On the genus level, there were less *Streptococcus*, *Neisseria*, and *Prevotella* and more *Veillonella* and *Fusobacterium* in the OSCC group. On the species level, there were less *Haemophilus* and more *Veillonella* in the OSCC group.

In the co-occurrence network deduced from bacteria enriched in the OSCC group and healthy group, the node in the network represented the sample node or the genus node, and the line between the sample node and the species node represented that the sample contains the genus. **Figure 4** shows the genus with abundance greater than 50. Both groups contained *Veillonella*, *Alloprevotella*, *Capnocytophaga*, *Neisseria*, *Gemella*, etc. Only the healthy group contained *Rothia*, and only the OSCC group contained *Lactococcus*, *Aggregatibacter*, *Peptostreptococcus*, etc.

Analysis of similarities (ANOSIM) (**Table 2**) was significant for the overall model ($R^2 = 0.13291$, $p = 0.05$), and pairwise comparisons revealed a significant difference between control subjects who remained healthy and those with OSCC. Although the coefficient of determination is very low, there is a difference between the two groups. On the one hand, it is suggested that there can be a clear difference between the two, which can be

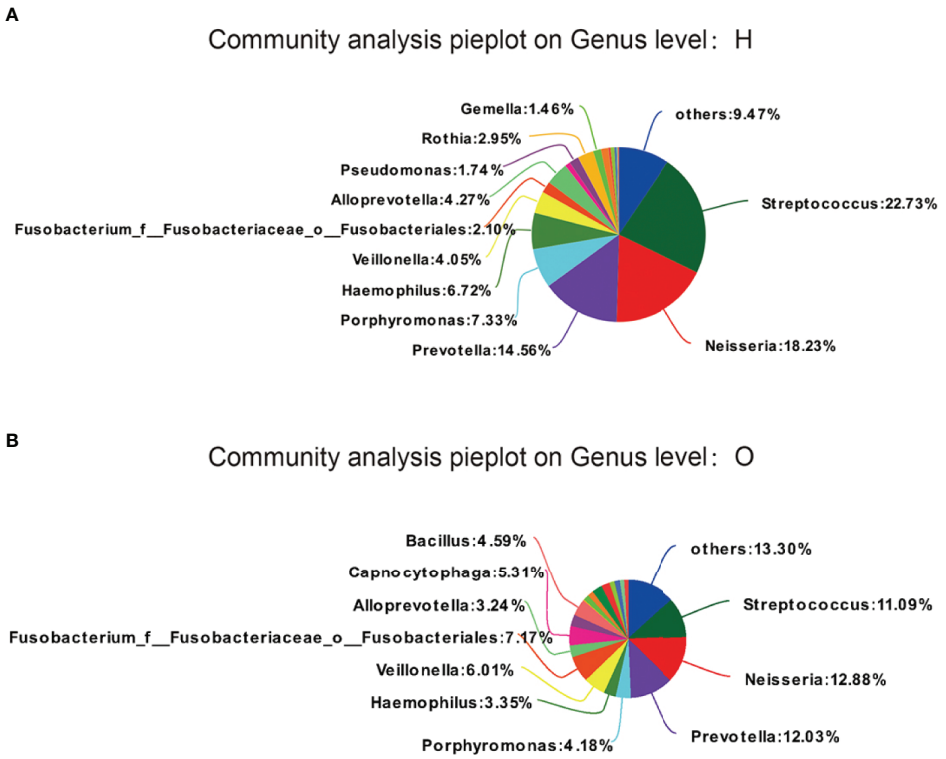


FIGURE 2 | OTUs phylotypes in healthy group **(A)** and OSCC patients **(B)**, analysed by Wilcoxon rank-sum test with Benjamini-Hochberg method.

used as an auxiliary diagnosis of OSCC; on the other hand, this mode difference may not be large and may lack discrimination in particular cases. Meanwhile, considering the number of OTUs, the samples size is relatively small, so a special method is needed to identify such slight differences. This is the machine learning diagnostic model mentioned later.

Abundance of Oral Microbial Communities in Oral Squamous Cell Carcinoma

To illustrate that the location in the oral cavity has an effect on the microbiota of the particular niche (saliva, subgingival plaque, surface of tumor, normal mucosa in the control side, and intratumoral tissue), we sampled microbiota in these five locations.

The Simpson index and Shannon index reflected the diversity of microorganisms in saliva, subgingival plaque, surface of tumor, normal mucosa, and intratumoral tissue. According to **Figure 5**, there were significant differences in the Simpson index and Shannon index of these five locations, which indicated that the diversity in different parts of the oral cavity was different. Among them, the Simpson index and Shannon index in intratumoral tissue were the highest, indicating that the diversity in intratumoral tissue was relatively high.

It can be seen from the Venn diagram (**Figure 6**) that the number of OTU that did not overlap on normal mucosa, subgingival plaque, saliva, intratumoral tissue, and surface of tumor was 34, 24, 42, 99, and 14, respectively, while the number of OTU that completely overlapped on the five sites was as high as 340, accounting for 55%~74% of the total number of each site. In other words, the composition of bacteria in saliva, subgingival plaque, surface of tumor, normal mucosa, and intratumoral tissue was very similar.

The size of nodes in **Figure 7** represents the abundance of genus, and different colors represent different genera. The colors of the lines indicate positive and negative correlations, red indicates positive correlation, and green indicates negative correlation. The thickness of the line indicates the correlation coefficient. The thicker the line, the higher the correlation between genera. The more lines, the more close the correlation.

It showed that the tumor site has the highest correlation between genera, and saliva site genus correlation is the lowest. In the tumor tissue, *Dialister*, *Johnsonella*, *Peptostreptococcus*, *Parvimonas*, and other bacteria were closely related to other bacteria. On the tumor surface, *Peptostreptococcus*, *Filifactor*, *Selenomonas*, and other bacteria were closely related to other bacteria. In the subgingival plaque, *Selenomonas*, *Peptostreptococcus*, *Prevotella*, and other bacteria were closely related to other bacteria, and the correlation was mostly positive. On the healthy mucosa, *Prevotella* was negatively associated with most microbes. In saliva, however, most of the microbes had a low microbiological correlation.

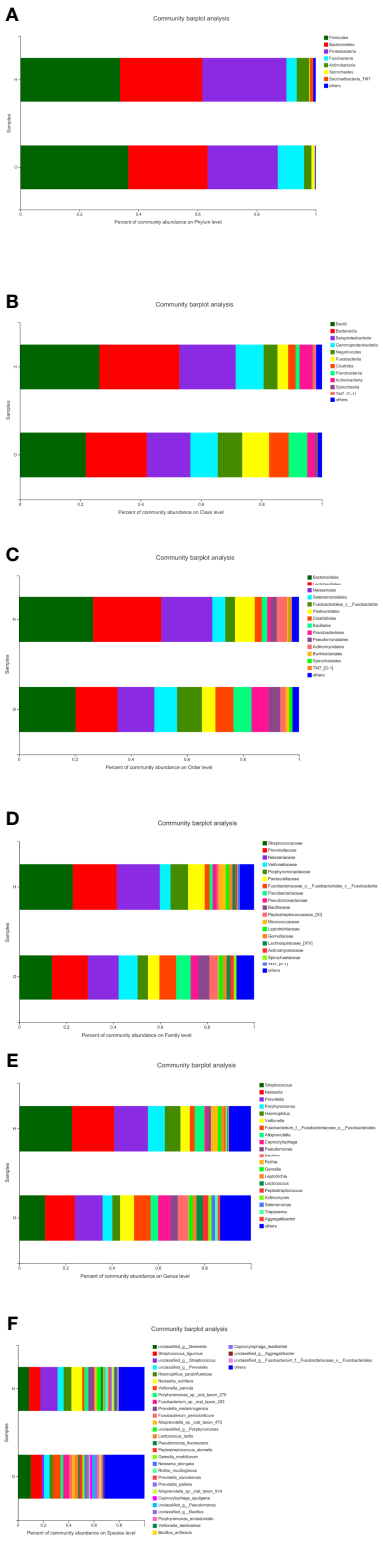


FIGURE 3 | Stacked column plots representing comparison of relative abundance of bacterial taxa between healthy and oral squamous cell carcinoma (OSCC) groups at phylum **(A)**, class **(B)**, order **(C)**, family **(D)**, genus **(E)**, and species **(F)**.

Evaluation of Oral Squamous Cell Carcinoma Prediction Random Forest Models Based on Saliva Microbiome

Given the noninvasive collection process of salivary samples and the high sensitivity in OSCC sample identification, using the microbiome in salivary samples to identify potential OSCC patients would be a more appropriate choice of screening the OSCC patients. An additional random forest model was then built only using the 93 salivary samples, in which 47 were from the OSCC patients while the remaining 46 were from healthy controls (**Table 3**). The result showed that the model's accuracy was 95.70%, and its sensitivity was 100%; i.e., four samples from healthy controls were misclassified as from OSCC patients. In addition to the model using only salivary samples, we wondered whether using samples of other sites could lead to the same performance. And the result is in the **Supplementary Material**.

Overfitting is a common concern in that a prediction model with high internal performance does not work well in other populations, especially in machine learning models. However, to evaluate such uncertainty, we carried out a batch of cross-validations. For each model, a training set containing 80% randomly selected samples was used to build a random forest model, and the rest of the samples were not used to build the model as the external test sets. Then, the average external accuracy of all the random forests provides an estimation of the model applied in external populations.

For the model built with OTUs in salivary samples, cross-validation showed an estimated external accuracy of 93.58%; i.e., 97 out of 1512 external test samples were misclassified in 84 external test sets containing 18 samples each. Still, no OSCC would be missed using the oral microbiome in salivary samples (**Table 4**).

The cross-validations of OSCC sample prediction random forests, i.e., using only salivary samples, with all the samples collected in five sites, suggested that the distinguished microbiome pattern in samples from OSCC individuals can also be used in external populations. Also, the sensitivity of external test samples still at 100% proves its high capacity in screening OSCC patients. Considering the noninvasive collection model, using random forests based on the microbiome in salivary samples would be strongly recommended to test whether the individuals are possible OSCC patients.

Another common concern is that given the well-known significant dependence on sample size, how many samples are required to build a model with favorable performance? Another batch of cross-validations, each with different sizes of training sets and test sets, was carried out to evaluate the dependence on sample size in OSCC identification using oral microbiome in salivary samples.

The sample sizes of training sets were set as 60% (56), 70% (65), 80% (75), and 90% (83) of all the 93 salivary samples. For each sample size, the random forest model was tested different times of a test sample size of 10,000 to obtain accuracy with comparable variations. The average accuracies of the tests with different sample sizes provided the association of the training sample sizes and the performance of OSCC identification models using the salivary microbiome.

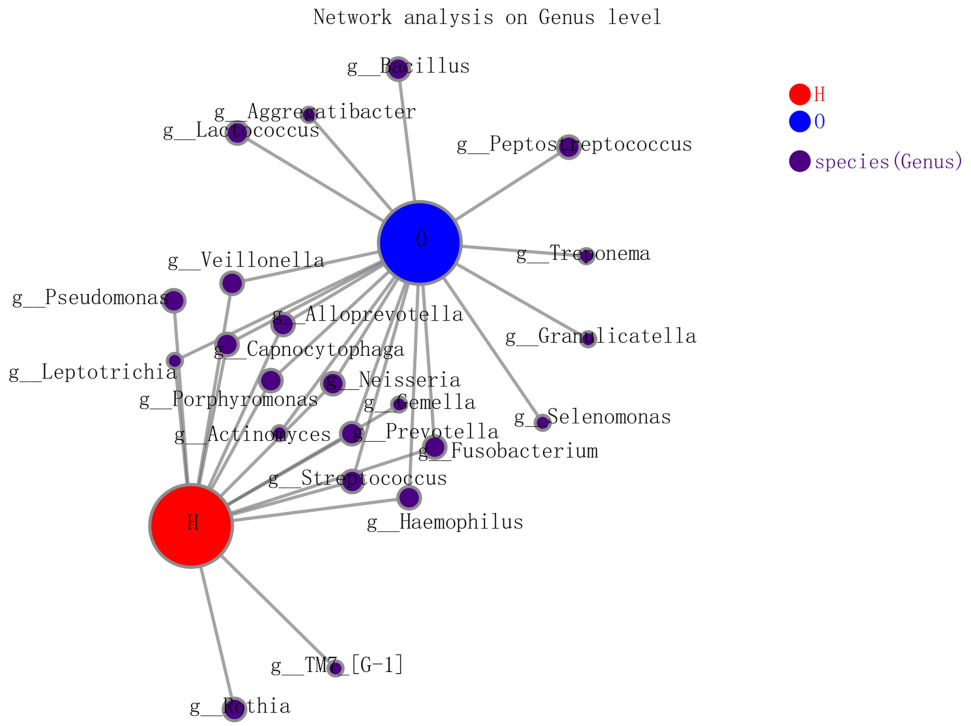


FIGURE 4 | The co-occurrence network deduced from bacteria enriched in oral squamous cell carcinoma (OSCC) group and healthy group.

TABLE 2 | ANOSIM in healthy group and OSCC patients.

	Df	SumsOfSqs	MeanSqs	F.Model	R^2	Pr(>F)
Group	1	6.864	6.8643	24.832	0.13291	0.001
Residuals	162	44.782	0.2764	0.86709		
Total	163	51.646	1			

ANOSIM, analysis of similarities; OSCC, oral squamous cell carcinoma.

Cross-validations showed that as the training sample size increased, the random forests based on salivary samples became more accurate. Seventy-five samples could supply significantly high performance, while more training samples could give rise to even greater accuracy; i.e., with all the 93 samples in training set at 95.70% accuracy, another 1.20% improvement could be obtained over 94.50% with 83 samples in the training set (**Figure 8**). Also, the larger training sample size decreased the variance of the prediction accuracy, suggesting that higher accuracy can be obtained by a model with a large training set.

Further insight into >40,000 prediction results provided a good sample of this point. In the cross-validations, all the random forests with training sample sizes >80% (75) exhibited sensitivities of 100%; therefore, no OSCC would be misclassified. However, as the training sample size decreased, false-negative predictions were found, i.e., 15 patients in 5,130 OSCC patients with a training sample size of 56 and three patients in 4,998 OSCC patients with a training sample size of 65. An interesting finding is that all the 18 false-negative predictions happened in one same sample. This sample was found from an OSCC patient having early invasive carcinoma on the oral cavity floor. At the early stages of cancer development, the characteristics of the oral microbiome were close to those of healthy individuals compared with other OSCC patients. Such samples from early cancer development stages might be misclassified as from healthy individuals due to the partly changed microbial profiles, which would be identified correctly by the models with larger training sets.

This suggested the strategy of screening model development; the prediction random forest can first be built based on small sample sizes, such as those >50, and then the accuracy can be improved as the new samples are added to the training set to renew the basic model.

Also, the difference between groups of healthy individuals from different centers suggested that a continuous and dynamic renewal of the prediction model using new samples would be of necessity for a potential change in population applied. It is recommended for each center to address the differences between microbial profiles in different populations when building its own prediction random forest.

DISCUSSION

The human microbiome, a dynamic, interconnected ecosystem reflecting the locating environments, plays a central role in the process of development, health, and disease (Bracci, 2017; Cong and Zhang, 2018; Verma et al., 2018). Although the differences between microbiome in groups having different disorder statues

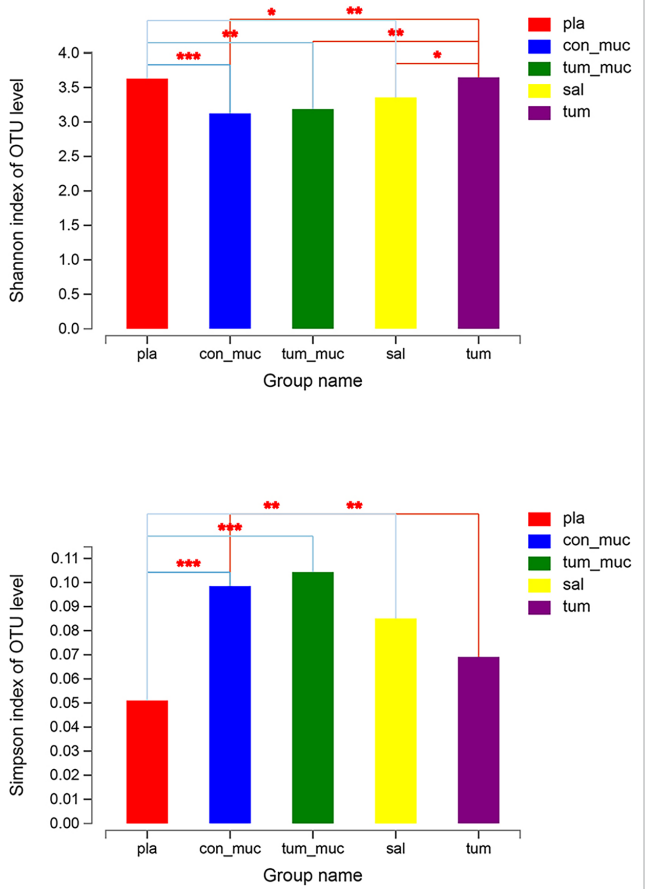


FIGURE 5 | Simpson index and Shannon index of microorganisms in subgingival plaque, normal mucosa, surface of tumor, saliva and intratumor tissue (* means $0.01 < P \leq 0.05$, ** means $0.001 < P \leq 0.01$, *** means $P \leq 0.001$).

might provide potential biomarkers, traditional analyses have low test power in identifying such differences due to the adverse effect of dimensionality. Although there is a deluge of data on the human microbiome, converting them into clinically meaningful insights remains challenging (Quince et al., 2009; Hu et al., 2013). Machine learning methods constitute proper tools for analyzing such high-dimensional datasets with a small sample size. For instance, Teng et al. (2015) developed a predictive model for early childhood caries (ECC) using oral microbiota by random forests machine learning algorithm innovatively, which became an asset for clinical work. The algorithm was also used in this study, indicating that OSCC can be diagnosed based on oral microbiota. Moreover, microbiota on any one of the five sites were useful for the diagnosis of OSCC. Thus, oral microbiota on any one of the five sites could be collected to diagnose OSCC in clinical practice.

Salivary samples would be an optimal choice for the OSCC preliminary diagnosis due to their advantages in the sample collection process. Early diagnosis plays a critical role in the treatment of OSCC, and many methods have been used in the diagnosis of OSCC. Compared with the traditional methods (CT, MRI, and PET), our novel model, based on oral microbiota, exhibited apparent advantages. First, compared with CT and PET, no radiation is involved during sample collection and examination. Second, the cost of 16S rRNA gene sequences is 20%–50% of CT/MRI for every patient and <20% of PET. Third, the examination is more convenient for patients than CT, MRI, and PET. This method only requires the collection of saliva and sequencing, without the need for professionals to purchase or learn other examination equipment, and can be easily implemented in oral medical clinics or hospitals. Fourth, the diagnosis of histopathological analysis usually takes about 3–5 days, because the preparation of tissue samples and the interpretation by the diagnostic physician are quite complex and rigorous. This method only requires sequencing and machine data processing, which will provide quick help for diagnosis. Besides, some studies have indicated that oral microbiota could provide a potential risk assessment for several other diseases, like dental caries (Stuckensen et al., 2000; Ng et al., 2005; Liao et al., 2011). Our investigations explored a novel method to detect OSCC at an early stage, expanding the application of oral microbiota in diagnosing oral diseases. Therefore, in the future, the analysis of oral microbiota might be included in annual physical examinations for large populations to detect the risk of different diseases. The selected people with a high risk of specific diseases could be referred to specialists for further confirmatory diagnosis. On the one hand, patients could benefit from the early diagnosis of the diseases; on the other hand, it could help reduce the social and public health expenses.

In the study, the accuracy of the diagnostic model was more favorable than that of the traditional methods. The accuracy of CT/MRI ranges from 66% to 86.4%. In recent years, ^{18}F-FDG PET has been recommended in the diagnosis of OSCC patients (Kitajima et al., 2015). The accuracy of ^{18}F-FDG PET ranges from 66.8% to 89.4%. In the present study, the accuracy of the novel model was 95%. Interestingly, there was no false-negative result in our diagnostic model. But there are still some false-positive individuals, and further confirmatory diagnosis could help exclude such cases.

In recent years, some studies indicated that oral microbial composition differed significantly from a healthy state to OSCC patients and non-tumoral to tumoral sites (Ahn et al., 2012; Pushalkar et al., 2012; Schmidt et al., 2014). Therefore, researchers tried to isolate some particular species and show their relationship with OSCC. In the present study, the results also provided evidence for some oral bacteria as potential research objects. As shown in **Table S2**, the top 10 features of oral microbiome in random forests were consistent with previously reported studies in which close relationships were detected between OSCC and the following bacterial species: *Porphyromonas*, *Fusobacterium*, *Prevotella*, *Leptotrichia*, *Moraxella*, *Bacillus*, and *Actinobacteria* (Sato et al., 2010; Al-Hebshi et al., 2015). Particularly, as pathogenic bacteria of periodontal disease, *P. gingivalis* and *Fusobacterium nucleatum* could promote oral carcinogenesis (Groeger et al., 2011; Gallimidi et al., 2015; Ha et al., 2015). *P. gingivalis* could

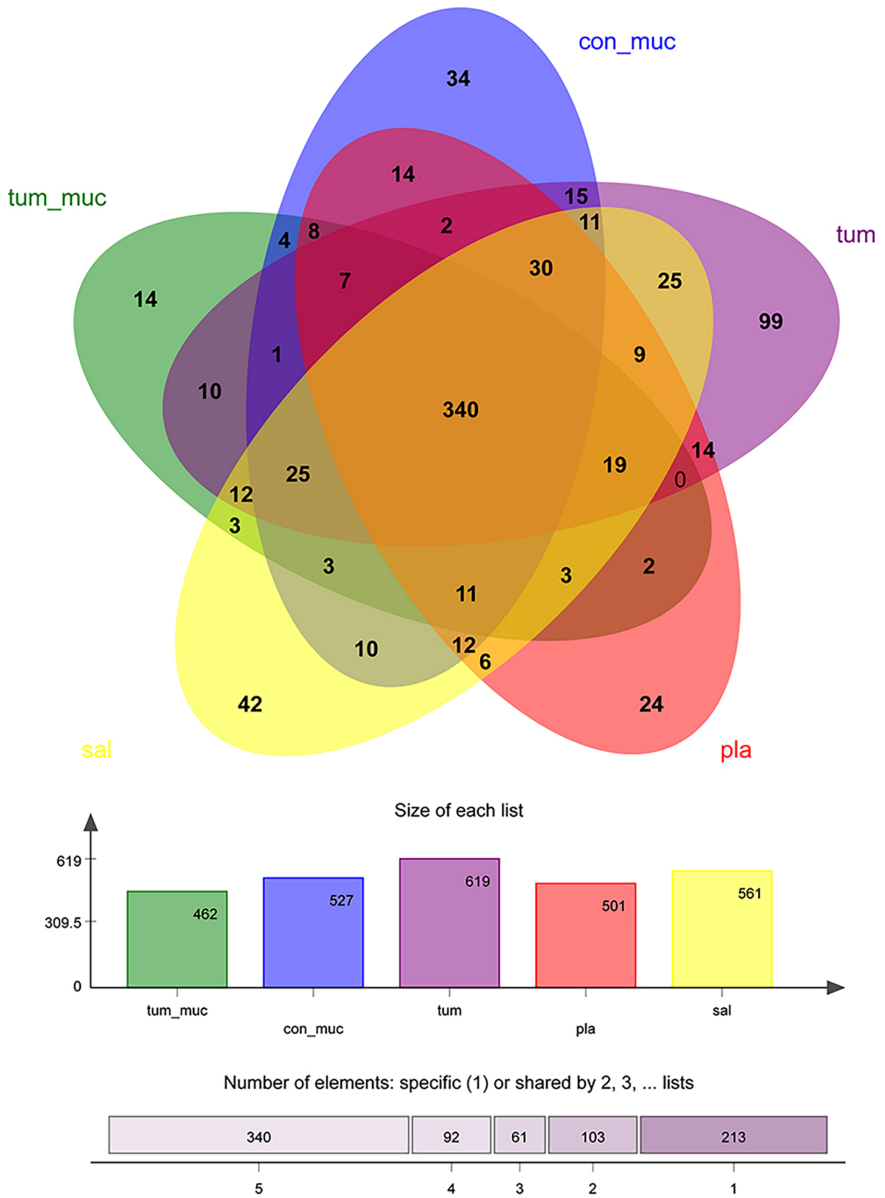


FIGURE 6 | Venn diagram of the number of operational taxonomic unit (OTU) among normal mucosa, subgingival plaque, saliva, intratumoral tissue, and surface of tumor.

promote immunoevasion of oral cancer by protecting cancer from macrophage attack and could facilitate cell migration, which was slightly enhanced by co-infection with *F. nucleatum* (Liu et al., 2020). *Prevotella* was found to have a close relationship with digestive tract cancers (Yang et al., 2009). Although other bacteria in the present study lacked in mechanism evidence, they provided clues for future studies to reveal the relationship between microorganisms and oral cancer.

In the present study, in one sample, the oral microbiome's characteristics were close to those of healthy individuals. Further analysis indicated that it might be because the sample was collected from a patient in the early stages of OSCC, confirming previous research (Chocolatewala et al., 2010; Perera et al., 2016; Mukherjee et al., 2017), in which the microbiome changes continued with cancer development. The microbiome will change with the pathological environment during carcinogenesis.

In conclusion, using random forests and cross-validations, this study provided a method to build a diagnostic model based on oral microbiota, which could be applied to the diagnosis of OSCC in large populations accurately and conveniently without radiation before invasive procedures. Furthermore, this study provided an application sample to develop diagnostic models as an auxiliary diagnostic tool not only for OSCC but also for various tumors.

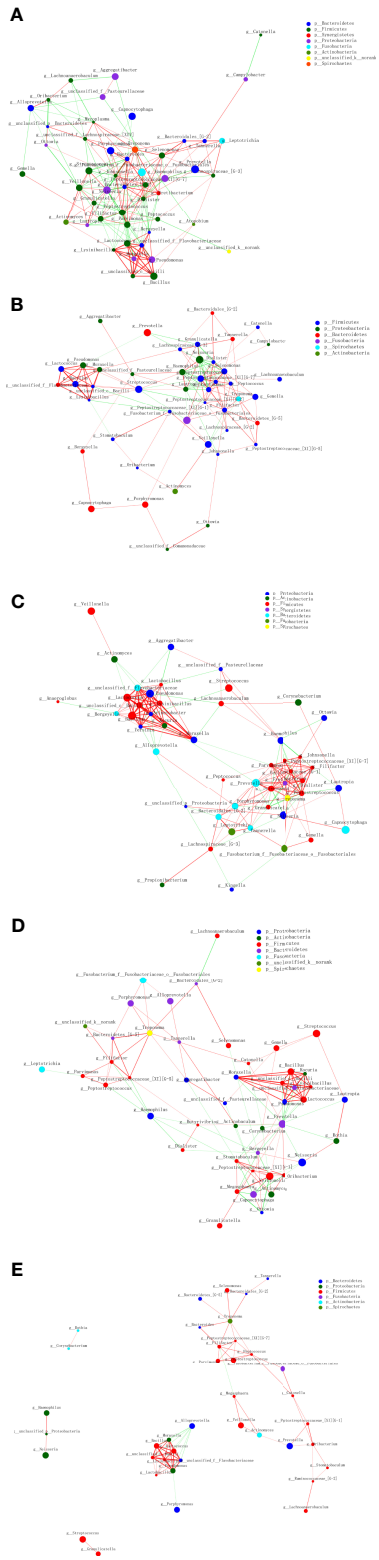


FIGURE 7 | Interconnection of the oral squamous cell carcinoma (OSCC) and salivary and other sites bacteria. Tumor tissue **(A)**, the tumor surface **(B)**, subgingival plaque **(C)**, healthy mucosa **(D)**, and saliva **(E)**.

TABLE 3 | Prediction and observation of OSCC in saliva samples.

Observed	Predicted		Total
	Healthy controls	OSCC patients	
Healthy controls	42	4	46
OSCC patients	0	47	47

OSCC, oral squamous cell carcinoma.

TABLE 4 | Prediction and observation of OSCC in saliva samples.

Observed	Predicted		Total
	Healthy controls	OSCC patients	(n = 1,512)
Healthy controls	659	97	756
OSCC patients	0	756	756

OSCC, oral squamous cell carcinoma.

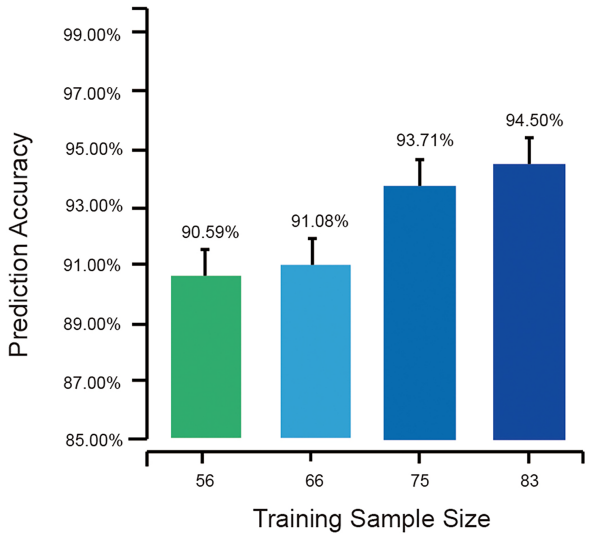


FIGURE 8 | The average external prediction accuracies of random forests with different training sample sizes. The bottom part of the figure is truncated to present the differences in the error bar.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, SRP119028.

ETHICS STATEMENT

The institutional review board of the West China Hospital Stomatology of Sichuan University approved the study (Approval number: WCHSIRB-D-2013-047). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

XXZ: validation, methodology, formal analysis, data curation, writing—original draft, and writing—review and editing. YH: validation, methodology, data curation, writing—original draft, and writing—review and editing. XP: methodology, data curation, and writing—original draft. BL: methodology, data curation, and writing—original draft. QH: methodology and data curation. BR: supervision. ML: supervision. LL: resources. YL: supervision. GC: formal analysis and supervision. JL: funding acquisition and supervision. YM: conceptualization, methodology, formal analysis, writing—review and editing, and funding acquisition. XDZ: conceptualization, methodology, and funding acquisition. LC: conceptualization, methodology, writing—review and editing, and funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Natural Science Foundation of China, 81870759 and 82071106 (LC), 81803332 (YM), and 81991500 and 81991501 (JL); the Sichuan Science & Technology Program 2018SZ0284 (YM); Innovative Research Team Program of Sichuan Province (LC); and the Chengdu Science & Technology Bureau 2018-YF05-01265-SN (YM).

ACKNOWLEDGMENTS

Thanks to all the volunteers who participated in this experiment and the technicians working in the State Key Laboratory of Oral Diseases.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2021.728933/full#supplementary-material

REFERENCES

Aas, J. A., Paster, B. J., Stokes, L. N., Olsen, I., and Dewhirst, F. E. (2005). Defining the Normal Bacterial Flora of the Oral Cavity. *J. Clin. Microbiol.* 43, 5721–5732. doi: 10.1128/jcm.43.11.5721-5732.2005

Ahn, J., Chen, C. Y., and Hayes, R. B. (2012). Oral Microbiome and Oral and Gastrointestinal Cancer Risk. *Cancer Causes Cont.* 23, 399–404. doi: 10.1007/s10552-011-9892-7

Al-Hebshi, N. N., Nasher, A. T., Idris, A. M., and Chen, T. (2015). Robust Species Taxonomy Assignment Algorithm for 16S rRNA NGS Reads: Application to Oral Carcinoma Samples. *J. Oral. Microbiol.* 7, 9. doi: 10.3402/jom.v7.28934

Avila, M., Ojcius, D. M., and Yilmaz, O. (2009). The Oral Microbiota: Living With a Permanent Guest. *DNA Cell Biol.* 28, 405–411. doi: 10.1089/dna.2009.0874

Bracci, P. M. (2017). Oral Health and the Oral Microbiome in Pancreatic Cancer: An Overview of Epidemiological Studies. *Cancer J.* 23, 310–314. doi: 10.1097/ppo.0000000000000287

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME Allows Analysis of High-Throughput Community Sequencing Data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Chang, C. R., Wang, H. Y., Liu, J. C., Pan, C. L., Zhang, D. M., Li, X., et al. (2019). Porphyromonas Gingivalis Infection Promoted the Proliferation of Oral Squamous Cell Carcinoma Cells Through the miR-21/PDCD4/AP-1 Negative Signaling Pathway. *ACS Infect. Dis.* 5, 1336–1347. doi: 10.1021/acsinfecdis.9b00032

Chao, A., Lee, S. M., and Jeng, S. L. (1992). Estimating Population Size for Capture-Recapture Data When Capture Probabilities Vary by Time and Individual Animal. *Biometrics* 48, 201–216. doi: 10.2307/2532750

Chao, A., and Shen, T.-J. (2003). Nonparametric Estimation of Shannon's Index of Diversity When There are Unseen Species in Sample. *Environ. Ecol. Stat.* 10, 429–443. doi: 10.1023/A:1026096204727

Chocolatewala, N., Chaturvedi, P., and Desale, R. (2010). The Role of Bacteria in Oral Cancer. *Indian J. Med. Paediatr. Oncol.* 31, 126–131. doi: 10.4103/0971-5851.76195

Cole, J. R., Chai, B., Farris, R. J., Wang, Q., Kulam, S. A., McGarrell, D. M., et al. (2005). The Ribosomal Database Project (RDP-II): Sequences and Tools for High-Throughput rRNA Analysis. *Nucleic Acids Res.* 33, D294–D296. doi: 10.1093/nar/gki038

Cong, J., and Zhang, X. (2018). How Human Microbiome Talks to Health and Disease. *Eur. J. Clin. Microbiol. Infect. Dis.* 37, 1595–1601. doi: 10.1007/s10096-018-3263-1

de Mendoza, I. L. I., Mendia, X. M., de la Fuente, A. M. G., Andres, G. Q., and Urizar, J. M. A. (2020). Role of Porphyromonas Gingivalis in Oral Squamous Cell Carcinoma Development: A Systematic Review. *J. Periodontal Res.* 55, 13–22. doi: 10.1111/jre.12691

Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C. R., Yu, W.-H., et al. (2010). The Human Oral Microbiome. *J. Bacteriol.* 192, 5002–5017. doi: 10.1128/jb.00542-10

Gallimidi, A. B., Fischman, S., Revach, B., Bulvik, R., Maliutina, A., Rubinstein, A. M., et al. (2015). Periodontal Pathogens Porphyromonas Gingivalis and Fusobacterium Nucleatum Promote Tumor Progression in an Oral-Specific Chemical Carcinogenesis Model. *Oncotarget* 6, 22613–22623. doi: 10.18632/oncotarget.4209

Groeger, S., Domann, E., Gonzales, J. R., Chakraborty, T., and Meyle, J. (2011). B7-H1 and B7-DC Receptors of Oral Squamous Carcinoma Cells are Upregulated by Porphyromonas Gingivalis. *Immunobiology* 216, 1302–1310. doi: 10.1016/j.imbio.2011.05.005

Guerrero-Preston, R., Godoy-Vitorino, F., Jedlicka, A., Rodriguez-Hilario, A., Gonzalez, H., Bondy, J., et al. (2016). 16s rRNA Amplicon Sequencing Identifies Microbiota Associated With Oral Cancer, Human Papilloma Virus Infection and Surgical Treatment. *Oncotarget* 7, 51320–51334. doi: 10.18632/oncotarget.9710

Ha, N. H., Woo, B. H., Kim, D. J., Ha, E. S., Choi, J. I., Kim, S. J., et al. (2015). Prolonged and Repetitive Exposure to Porphyromonas Gingivalis Increases Aggressiveness of Oral Cancer Cells by Promoting Acquisition of Cancer Stem Cell Properties. *Tumor Biol.* 36, 9947–9960. doi: 10.1007/s13277-015-3764-9

Hooper, S. J., Crean, S. J., Fardy, M. J., Lewis, M. A. O., Spratt, D. A., Wade, W. G., et al. (2007). A Molecular Analysis of the Bacteria Present Within Oral Squamous Cell Carcinoma. *J. Med. Microbiol.* 56, 1651–1659. doi: 10.1099/jmm.0.46918-0

Hu, Y. J., Wang, Q., Jiang, Y. T., Ma, R., Xia, W. W., Tang, Z. S., et al. (2013). Characterization of Oral Bacterial Diversity of Irradiated Patients by High-Throughput Sequencing. *Int. J. Oral Sci.* 5, 21–25. doi: 10.1038/ijos.2013.15

Kalogirou, E.-M., and Sklavounou-Andrikopoulou, A. (2017). Is Dental Implantation Indicated in Patients with Oral Mucosal Diseases. *Balkan J. Dent. Med.* 21, 83–92. doi: 10.1515/bjdm-2017-0013

Kitajima, K., Suenaga, Y., Minamikawa, T., Komori, T., Otsuki, N., Nibu, K., et al. (2015). Clinical Significance of SUVmax in F-18-FDG PET/CT Scan for Detecting Nodal Metastases in Patients With Oral Squamous Cell Carcinoma. *Springerplus* 4, 12. doi: 10.1186/s40064-015-1521-6

Krogh, P., Hald, B., and Holmstrup, P. (1987). Possible Mycological Etiology of Oral Mucosal Cancer - Catalytic Potential of Infecting Candida-Albicans and Other Yeasts in Production of N-Nitrosobenzylmethylamine. *Carcinogenesis* 8, 1543–1548. doi: 10.1093/carcin/8.10.1543

Kujan, O., Glenny, A. M., Duxbury, J., Thakker, N., and Sloan, P. (2005). Evaluation of Screening Strategies for Improving Oral Cancer Mortality: A Cochrane Systematic Review. *J. Dent. Educ.* 69, 255–265. doi: 10.1002/j.0022-0337.2005.69.2.tb03911.x

Li, B., Ge, Y., Cheng, L., Zeng, B., Yu, J., Peng, X., et al. (2019). Oral Bacteria Colonize and Compete With Gut Microbiota in Gnotobiotic Mice. *Int. J. Oral. Sci.* 11, 10. doi: 10.1038/s41368-018-0043-9

Li, F., Tao, D., Feng, X., Wong, M. C. M., and Lu, H. (2018). Establishment and Development of Oral Microflora in 12-24 Month-Old Toddlers Monitored by High-Throughput Sequencing. *Front. Cell. Infect. Microbiol.* 8, 422. doi: 10.3389/fcimb.2018.00422

Liao, C. T., Wang, H. M., Huang, S. F., Chen, I. H., Kang, C. J., Lin, C. Y., et al. (2011). PET and PET/CT of the Neck Lymph Nodes Improves Risk Prediction in Patients With Squamous Cell Carcinoma of the Oral Cavity. *J. Nucl. Med.* 52, 180–187. doi: 10.2967/jnumed.110.082370

Liu, S., Zhou, X., Peng, X., Li, M., Ren, B., Cheng, G., et al. (2020). Porphyromonas Gingivalis Promotes Immunoevasion of Oral Cancer by Protecting Cancer From Macrophage Attack. *J. Immunol.* 205, 282–289. doi: 10.4049/jimmunol.1901138

McInnes, P., and Cutting, M. (2010). *Core Microbiome Sampling Protocol A HMP Protocol# 07–001. Manual of Procedures*.

Mukherjee, P. K., Wang, H., Retuerto, M., Zhang, H., Burkey, B., Ghannoum, M. A., et al. (2017). Bacteriome and Mycobiome Associations in Oral Tongue Cancer. *Oncotarget* 8, 97273–97289. doi: 10.18632/oncotarget.21921

Ng, S. H., Yen, T. C., Liao, C. T., Chang, J. T. C., Chan, S. C., Ko, S. F., et al. (2005). F-18-FDG PET and CT/MRI in Oral Cavity Squamous Cell Carcinoma: A Prospective Study of 124 Patients With Histologic Correlation. *J. Nucl. Med.* 46, 1136–1143.

Park, D. G., Woo, B. H., Lee, B. J., Yoon, S., Cho, Y., Kim, Y. D., et al. (2019). Serum Levels of Interleukin-6 and Titers of Antibodies Against Porphyromonas Gingivalis Could Be Potential Biomarkers for the Diagnosis of Oral Squamous Cell Carcinoma. *Int. J. Mol. Sci.* 20, 12. doi: 10.3390/ijms20112749

Perera, M., Al-Hebshi, N. N., Speicher, D. J., Perera, I., and Johnson, N. W. (2016). Emerging Role of Bacteria in Oral Carcinogenesis: A Review With Special Reference to Perio-Pathogenic Bacteria. *J. Oral. Microbiol.* 8, 10. doi: 10.3402/jom.v8.32762

Pushalkar, S., Ji, X., Li, Y., Estilo, C., Yegnanarayana, R., Singh, B., et al. (2012). Comparison of Oral Microbiota in Tumor and Non-Tumor Tissues of Patients With Oral Squamous Cell Carcinoma. *BMC Microbiol.* 12, 144. doi: 10.1186/1471-2180-12-144

Quince, C., Lanzen, A., Curtis, T. P., Davenport, R. J., Hall, N., Head, I. M., et al. (2009). Accurate Determination of Microbial Diversity From 454 Pyrosequencing Data. *Nat. Methods* 6, 639–U27. doi: 10.1038/nmeth.1361

Sato, J., Goto, J., Murata, T., Kitamori, S., Yamazaki, Y., Satoh, A., et al. (2010). Changes in Saliva Interleukin-6 Levels in Patients With Oral Squamous Cell Carcinoma. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.* 110, 330–336. doi: 10.1016/j.tripleo.2010.03.040

Schmidt, B. L., Kuczynski, J., Bhattacharya, A., Huey, B., Corby, P. M., Queiroz, E. L. S., et al. (2014). Changes in Abundance of Oral Microbiota Associated With Oral Cancer. *PloS One* 9, e98741. doi: 10.1371/journal.pone.0098741

Segata, N., Haake, S. K., Mannon, P., Lemon, K. P., Waldron, L., Gevers, D., et al. (2012). Composition of the Adult Digestive Tract Bacterial Microbiome Based on Seven Mouth Surfaces, Tonsils, Throat and Stool Samples. *Genome Biol.* 13, R42. doi: 10.1186/gb-2012-13-6-r42

Simard, E. P., Torre, L. A., and Jemal, A. (2014). International Trends in Head and Neck Cancer Incidence Rates: Differences by Country, Sex and Anatomic Site. *Oral. Oncol.* 50, 387–403. doi: 10.1016/j.oraloncology.2014.01.016

Stuckensen, T., Kovacs, A. F., Adams, S., and Baum, R. P. (2000). Staging of the Neck in Patients With Oral Cavity Squamous Cell Carcinomas: A Prospective Comparison of PET, Ultrasound, CT and MRI. *J. Craniomaxillofac. Surg.* 28, 319–324. doi: 10.1054/jcms.2000.0172

Sturød, K., Dhariwal, A., Dahle, U. R., Vestrheim, D. F., and Petersen, F. C. (2020). Impact of Narrow-Spectrum Penicillin V on the Oral and Faecal Resistome in a Young Child Treated for Otitis Media. *J. Global Antimicrob. Resist.* 20, 290–297. doi: 10.1016/j.jgar.2019.08.004

Teng, F., Yang, F., Huang, S., Bo, C. P., Xu, Z. Z., Amir, A., et al. (2015). Prediction of Early Childhood Caries *Via* Spatial-Temporal Variations of Oral Microbiota. *Cell Host Microbe* 18, 296–306. doi: 10.1016/j.chom.2015.08.005

Verma, D., Garg, P. K., and Dubey, A. K. (2018). Insights Into the Human Oral Microbiome. *Arch. Microbiol.* 200, 525–540. doi: 10.1007/s00203-018-1505-3

Wang, A.-H., Li, M., Li, C.-Q., Kou, G.-J., Zuo, X.-L., and Li, Y.-Q. (2016). Human Colorectal Mucosal Microbiota Correlates With Its Host Niche Physiology Revealed by Endomicroscopy. *Sci. Rep.* 6, 21952. doi: 10.1038/srep21952

Wang, Y. X., Liu, S. Y., Li, B. L., Jiang, Y. L., Zhou, X. X., Chen, J., et al. (2019). Staphylococcus Aureus Induces COX-2-Dependent Proliferation and Malignant Transformation in Oral Keratinocytes. *J. Oral. Microbiol.* 11, 12. doi: 10.1080/20002297.2019.1643205

Xie, Z., Zhou, F., Yang, Y., Li, L., Lei, Y., Lin, X., et al. (2018). Lnc-PCDH9-13:1 Is a Hypersensitive and Specific Biomarker for Early Hepatocellular Carcinoma. *EBioMedicine* 33, 57–67. doi: 10.1016/j.ebiom.2018.06.026

Xun, Z., Zhang, Q., Xu, T., Chen, N., and Chen, F. (2018). Dysbiosis and Ecotypes of the Salivary Microbiome Associated With Inflammatory Bowel Diseases and the Assistance in Diagnosis of Diseases Using Oral Bacterial Profiles. *Front. Microbiol.* 9, 1136. doi: 10.3389/fmicb.2018.01136

Yang, L. Y., Lu, X. H., Nossa, C. W., Francois, F., Peek, R. M., and Pei, Z. H. (2009). Inflammation and Intestinal Metaplasia of the Distal Esophagus Are Associated With Alterations in the Microbiome. *Gastroenterology* 137, 588–597. doi: 10.1053/j.gastro.2009.04.046

Yin, X., Gu, X. H., Yin, T. T., Wen, H. Y., Gao, X. L., and Zheng, X. (2016). Study of Enteropathogenic Bacteria in Children With Acute Diarrhoea Aged From 7 to 10 Years in Xuzhou, China. *Microb. Pathog.* 91, 41–45. doi: 10.1016/j.micpath.2015.11.027

Zarco, M., Vess, T., and Ginsburg, G. (2012). The Oral Microbiome in Health and Disease and the Potential Impact on Personalized Dental Medicine. *Oral. Dis.* 18, 109–120. doi: 10.1111/j.1601-0825.2011.01851.x

Zhou, S., Cai, Y., Wang, M., Yang, W. D., and Duan, N. (2018). Oral Microbial Flora of Patients With Sicca Syndrome. *Mol. Med. Rep.* 18, 4895–4903. doi: 10.3892/mmr.2018.9520

Zhu, Y., Lin, X., Zhao, F., Shi, X., Li, H., Li, Y., et al. (2015). Meat, Dairy and Plant Proteins Alter Bacterial Composition of Rat Gut Bacteria. *Sci. Rep.* 5, 15220. doi: 10.1038/srep15220

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Cellular and Infection Microbiology

ORIGINAL RESEARCH
published: 26 August 2021
doi: 10.3389/fcimb.2021.720790

Oral Phenotype and Salivary Microbiome of Individuals With Papillon–Lefèvre Syndrome

*Giulia Melo Lettieri[1], Luander Medrado Santiago[1], Giancarlo Crosara Lettieri[2], Luiz Gustavo dos Anjos Borges[3,4], Letícia Marconatto[3], Laudimar Alves de Oliveira[1], Nailê Damé-Teixeira[1] and Loise Pedrosa Salles[1]**

[1] Department of Dentistry, Faculty of Health Sciences, University of Brasilia, Brasília, Brazil, [2] Periodontology Research Group, Specialized Center in Periodontology and Implantology, Brasília, Brazil, [3] Microbial Interactions and Processes Research Group, Helmholtz Centre for Infection Research, Braunschweig, Germany, [4] Institute of Petroleum and Natural Resources, Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Brazil

OPEN ACCESS

Edited by:
George Seghal Kiran,
Pondicherry University, India

Reviewed by:
J Christopher Fenno,
University of Michigan, United States
Thinesh Thangadurai,
Pondicherry University, India

***Correspondence:**
Loise Pedrosa Salles
loise@unb.br

Specialty section:
This article was submitted to Microbiome in Health and Disease, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: 05 June 2021
Accepted: 05 August 2021
Published: 26 August 2021

Citation:
Lettieri GM, Santiago LM, Lettieri GC, Borges LGdA, Marconatto L, de Oliveira LA, Damé-Teixeira N and Salles LP (2021) Oral Phenotype and Salivary Microbiome of Individuals With Papillon–Lefèvre Syndrome. Front. Cell. Infect. Microbiol. 11:720790. doi: 10.3389/fcimb.2021.720790

Papillon–Lefèvre syndrome (PLS) is an autosomal recessive rare disease, main characteristics of which include palmoplantar hyperkeratosis and premature edentulism due to advanced periodontitis (formerly aggressive periodontitis). This study aimed to characterize the oral phenotype, including salivary parameters, and the salivary microbiome of three PLS sisters, comparatively. Two sisters were toothless (PLSTL1 and PLSTL2), and one sister had most of the teeth in the oral cavity (PLST). Total DNA was extracted from the unstimulated saliva, and the amplicon sequencing of the 16S rRNA gene fragment was performed in an Ion PGM platform. The amplicon sequence variants (ASVs) were obtained using the DADA2 pipeline, and the taxonomy was assigned using the SILVA v.138. The main phenotypic characteristics of PLS were bone loss and premature loss of primary and permanent dentition. The PLST sister presented advanced periodontitis with gingival bleeding and suppuration, corresponding to the advanced periodontitis as a manifestation of systemic disease, stage IV, grade C. All three PLS sisters presented hyposalivation as a possible secondary outcome of the syndrome. Interestingly, PLST salivary microbiota was dominated by the uncultured bacteria *Bacterioidales* (F0058), *Fusobacterium*, *Treponema*, and *Sulfophobococcus* (*Archaea* domain). *Streptococcus*, *Haemophilus*, and *Caldivirga* (*Archaea*) dominated the microbiome of the PLSTL1 sister, while the PLSTL2 had higher abundances of *Lactobacillus* and *Porphyromonas*. This study was the first to show a high abundance of organisms belonging to the *Archaea* domain comprising a core microbiome in human saliva. In conclusion, a PLST individual does have a microbiota different from that of the periodontitis' aggressiveness previously recognized. Due to an ineffective cathepsin C, the impairment of neutrophils probably provided a favorable environment for the PLS microbiome. The interactions of *Bacteroidales* F0058, *Caldivirga*, and *Sulfophobococcus* with the microbial consortium of PLS deserves future investigation. Traditional periodontal therapy is not efficient in PLS patients. Unraveling the PLS

microbiome is essential in searching for appropriate treatment and avoiding early tooth loss.

Keywords: Papillon–Lefèvre disease, cathepsin C, periodontal infection, saliva, microbiology, periodontitis

INTRODUCTION

Papillon and Lefèvre first described the Papillon–Lefèvre syndrome (PLS) in 1924 (Papillon and Lefevre, 1924). PLS is a hereditary autosomal recessive and rare condition that affects one to four people per million (Hart and Shapira, 1994). Generally, consanguineous marriages are PLS individuals' origin (Bhavsar et al., 2013; Bullón et al., 2018). It was estimated that over one billion people live in countries where consanguineous marriages are customary (Hamamy et al., 2011). Among them, one in every three marriages is between cousins. The main impact of consanguinity is the increased expression of multiple mutations encoding rare autosomal recessive genetic disorders, with an increased risk for first cousin couples to bear affected children. The PLS's main phenotypic characteristics are palmoplantar hyperkeratosis and the premature loss of deciduous and permanent teeth. The edentulism process in PLS starts with aggressive periodontitis (AlBarrak et al., 2016). In the period of dental exfoliation, the gingival tissue of PLS patients becomes hyperplastic and hemorrhagic, and there is an extended significant bone loss of the maxilla and mandible, cement exposure, and tooth mobility that culminates in loss of teeth (Robertson et al., 2001; Jordan, 2004; Tumen et al., 2015). Therefore, poor quality of life is expected in childhood, as it corresponds with the most destructive period of the disease. The periodontal condition and tooth loss generate high sensitivity in patients with PLS and poor diet quality (Hattab, 2019). These subjects present an increased incidence of skin and oral infections, which led to a substantial immunological disorder hypothesis at the first line of cellular defense. The immune system impairment may explain the predisposition to oral infection and periodontitis in PLS as a primary etiological component.

The PLS results from mutations in the cathepsin C gene (*CTSC*), also known as dipeptidyl peptidase 1 (*DPPI*), located on chromosome 11q14. The *CTSC* mutations produce an inactive cysteine protease or reduce its function (Fischer et al., 1997; Hart et al., 1999; Toomes et al., 1999). Currently, 113 *CTSC* variants have been reported in ethnically various populations, including a novel missense variant in exon 6 of the cathepsin C gene of a PLS Chinese individual (Yu et al., 2021). Over 90% of the variants were missense variants, nonsense variants, or frameshift variants, and most of them were in exons 5–7 of *CTSC*. Cathepsin C is an essential lysosomal enzyme in the cascade of activation of immune and inflammatory cell serine proteases and other cell lineages. The majority of proteins that demand cathepsin C processing are essential for the innate immune system's proper function. Neutrophil elastase, proteinase-3, and granzymes A, B, and C are examples of proteins that depend on cathepsin C-mediated cleavage for activation (Kaplan and Radic, 2012). Bullón et al. (2018) showed autophagosome accumulation in mutant fibroblasts from a PLS patient's skin. The autophagosome accumulation was associated with alterations in oxidative/antioxidative status, reduced oxygen consumption, and a marked autophagic dysfunction. Immune and inflammatory cells also showed dysfunctional behavior in PLS. For instance, neutrophils demonstrated hyperactivity with increased oxidative stress and reduced capacity to form neutrophils' extracellular trap structures (NETs) (Sørensen et al., 2014; Roberts et al., 2016). NETs are important defensive structures composed of DNA, chromatin, and bactericidal proteins (Kaplan and Radic, 2012). Sørensen et al. (2014) showed that PLS patients neutrophils lack or had significantly reduced amounts of neutrophils elastase (NE), cathepsin G (CTSG), proteinase 3 (PR3), and azurocidin (CAP37). Azurocidin is a member of the neutrophil serine proteases, with intense chemotactic activity toward monocytes, and the formation of NETs depends on the presence of NE. Excessive or diminished NET production may lead to autoimmune and inflammatory disorders, like inflammasome activation, interfering significantly in PLS patients' defensive mechanisms. Scientific reports of $CTSC^{-/-}$ mice show that neutrophil granulocyte is altered in the absence of CTSC (John et al., 2019). Moreover, the neutrophil serine protease elastase (NE) activity was markedly reduced by approximately 50% in the knockout granulocytes. Other neutrophil serine proteases, cathepsin G (CTSG) and proteinase 3 (PR3), were strongly reduced as well. The cleavage of the cell–cell contact molecule E-cadherin was also impaired in the absence of CTSC, suggesting that the impaired tissue infiltration of $CTSC^{-/-}$ neutrophils are caused by reduced E-cadherin cleavage at adherens junctions rather than by reduced motility of neutrophils (John et al., 2019). Taken together, the mutations in the cathepsin-C gene and the immune/inflammatory cell dysfunction may explain PLS patients being more prone to oral dysbiosis and proliferation of periodontal biofilms. More critical, PLS individuals may present a diverse biofilm challenging to control with the conventional treatment.

The oral microbiota in dysbiosis is a relevant factor in several oral conditions such as dental caries, periodontitis (PD), apical lesions, alveolar osteitis, and tonsillitis. It also plays a role in certain systemic diseases such as cardiovascular disease, diabetes mellitus, pneumonia, and premature births (Dewhirst et al., 2010). Nonetheless, the knowledge of the microbial community is essential to evaluate the effects of these microorganisms in the host and might help the evolution of treatments for oral disorders (Socransky and Haffajee, 2005). However, the oral microbiome in PLS is not fully described or related to the phenotype of the syndrome. This knowledge gap might explain why the periodontitis (PD) treatment is not efficient in those patients (Albandar et al., 2012). The hypothesis for the unsuccessful PD therapy in PLS is the combination of the

existence of pathogenic species, the onset inflammation (Kilian et al., 2016; Hoare et al., 2018), and the immune disturbance caused by cathepsin C mutation that is not well understood (Albandar et al., 2012; Kilian et al., 2016). Therefore, understanding the triad "oral microbiota, immune/inflammatory response, and cathepsin-C mutation" is mandatory to search for a successful periodontal and systemic treatment of PLS, currently unavailable. This study aimed to characterize the oral phenotype, including salivary parameters, and the salivary microbiome of three PLS sisters, comparatively. Two of them were already edentulous, while the younger one had 15 teeth in the oral cavity. To achieve our purpose, we chose a combination of universal primer pairs for the 16S rRNA gene with suitable putative coverage for archaea and the analysis of the amplicon sequence variants (ASVs) followed by taxonomy assignment using the SILVA v.138. Sequence variations in 16S and other metagenomic loci contain phylogenetic information that can be used to infer the taxonomic relationships of the microbial hosts (Fricker et al., 2019). Analysis of ASVs provides improved sensitivity and specificity and reduces the problem of inflated microbiota datasets due to falsely identified distinct operational taxonomic units (OTUs) originating from misclustered sequences (Callahan et al., 2017). According to the authors, the ultimate reference-free statistical denoising methods such as Dada2 overcome the non-reproducibility of OTU clustering results with modified or expanded datasets by recovering independent biological sequences as ASVs, promoting reproducibility and comparability of amplicon-based microbiome analysis.

MATERIALS AND METHODS

Participants and Clinical and Radiographic Examination

This research was approved by the Research Ethics Committee of the School of Health Sciences of the University of Brasília (FS-UnB; no. 2.974.167) and performed under the World Medical Association (WMA) Declaration of Helsinki. Individuals were informed verbally about the study's objective and signed a consent form. Three sisters from a consanguineous marriage of first cousin couples and diagnosed with PLS were selected for this study. Characteristics preintervention included advanced periodontitis, palmoplantar hyperkeratosis, and a long story of failed dental treatment. The sisters were submitted to clinical, photographic, and radiographic examinations. Two PLS sisters had lost all their teeth approximately 2 years before the study enrolment, except for the impacted third molars: Papillon–Lefèvre syndrome toothless 1 (PLSTL1) and Papillon–Lefèvre syndrome, toothless 2 (PLSTL2). One sister had 15 teeth in her mouth [Papillon–Lefèvre syndrome toothed (PLST)]. The sisters were 14, 16, and 18 years old at the baseline on August 3, 2018, non-smokers, and had no other systemic diseases apart from PLS. After the saliva samples collection, the PLST patient was submitted to oral hygiene instruction and motivation. An experienced periodontist treated the PLST patient with subgingival scaling and root planning under local anesthesia. The probing depth technique consisted of 4 points measurements to detect the most profound penetration areas. The disease classification was established following the 2017 World Workshop of Periodontal and Peri-implant Disease and Conditions (Caton et al., 2018). Re-evaluations were carried out 1 month later and every 3 months after the study enrolment.

Salivary Characteristics

Salivary Flow and Reliability Assays

Saliva samples were collected in the morning on of August 3, 2018. Patients were asked to refrain from eating and drinking for 2 h before the sampling. None of the sisters were using systemic antibiotics or local antimicrobials or submitted to periodontal treatment for nine months previously to the saliva collection. The unstimulated salivary flowrate was performed by passive drooling, with patients seated. The collection was carried out for 5 min. The volume of saliva was measured (mL/min). The saliva reliability was analyzed during the transfer of saliva to a microtube. The stimulated salivary flow was performed by 1 min chewing a rubber dam (Madeitex, São José dos Campos, SP, Brazil). The saliva during the stimulation was discarded, and then, the flow was measured by 5 min (Navazesh and Kumar, 2008). For the phenotypic analysis, the PLS salivary samples were collected in three different sessions, with intervals of approximately 1 year. No medication or periodontal treatment was taking place 3 months before or during the saliva collection period. The classifications regarding resting salivary flow were assialia (0.00 mL/min), hyposialia (0.1–0.29 mL/min), and ideal (0.3–0.4 mL/min). Regarding stimulated salivary flow, the classifications were asialia (0.00 mL/min), severe hyposalivation (0.1–0.4 mL/min), moderate hyposalivation (0.5–0.9 mL/min), mild hyposalivation (1.0–1.4 mL/min), ideal salivation (1.5–2.5 mL/min), and sialorrhea (>2.5 mL/min). Regarding the reliability, the classifications were serous (does not form a string), fluid (a string of 2 cm), or viscous (a string of ≥5 cm).

Salivary pH and Buffering Capacity

A volume of 1 mL of stimulated saliva was transferred to 1.5 mL tubes, and the pH was measured using the tape method (pH-Fix 0-14, Macherey-Nagel, Düren, NRW, Germany) by 30 s. Another 1 mL of stimulated saliva was added to 3 ml of hydrochloric acid PA 37% (0.005M) (Dinâmica, Indaiatuba, Brazil) to evaluate the buffer capacity and the pH measured again after 2 min (mColorpHast MilliporeSigma, Burlington, MA, USA).

Salivary Glucose

The salivary glucose was analyzed using the Glucose Liquiform kit (Labtest Diagnóstica, Lagoa Santa, MG, Brazil). The glucose stabilizer Glistab (Labtest Diagnóstica, Lagoa Santa, MG, Brazil) was added to the unstimulated saliva in the proportion of 30 μL for each 3 mL of saliva. After that, the samples were centrifuged for 1 min, and 150 μL of the supernatant was mixed to 500 μL of the reagent 1 (phosphate buffer, 30 mmol/L, pH 7.5; phenol, ≥1 mmoL/L; glucose oxidase, ≥12,500 U/L; peroxidases, ≥800 U/L; 4-aminoanthypyrine, ≥290 μmol/L; sodium azide, 7.5 mmol/L;

and surfactants). The samples were incubated at 37°C for 10 min, and the absorbance was measured at 505 nm using the spectrophotometer Spectramax (Molecular Devices LCC, San Jose, CA, USA). The capillary glucose (mg/dL) was evaluated for comparative purposes with Accu-Check and test strips (Roche, Basel, BS, Swiss).

Salivary Amylase

This assay was performed using an adapted protocol of amylase test (Labtest Diagnóstica, Lagoa Santa, MG, Brazil). The saliva samples were centrifuged for 1 min to 20,000 rpm, and the supernatant diluted 300 times with 0.85% NaCl. After this, 2 µl of the prepared saliva samples was added into 50 µl of substrate 1 (0.4 g/L starch, pH 7.0; phosphate buffer; and stabilizer), previously incubated at 37°C for 2 min, and incubated for additional 7 min and 30 s at 37°C. Then, 50 µL of the color reagent (potassium iodate, 16.7 mmol/L; potassium iodide, 271 mmol/L; and hydrochloric acid, 112 mmol/L) and 400 µL of distilled water were added, and after 5 min at room temperature, the absorbance was measured at 660 nm using the spectrophotometer Spectramax (Molecular Devices LCC, San Jose, CA, USA).

Salivary DNA Extraction, Amplicon Sequencing, and Bioinformatics

The DNA was extracted using phenyl-chloroform as successfully described by Smalla et al. (1993) for prokaryotes cells. After DNA extraction, PCR was performed for partial amplification of 16S rRNA gene using universal primers 515F (5′-GTGCCAGCMGCCGCGGTAA-3′) and 806R (5′-GGACTACVSGGGTATCTAAT-3′) (Bates et al., 2011). The PCR conditions were as follows: 50 µL mixture, consisting of 1.5 mM $MgCl_2$, 0.2 µM of each primer, 0.2 mM of each dNTP, 1 U Platinum Taq DNA polymerase, 1× PCR reaction buffer, and approximately 10 ng of genomic DNA. The PCR cycles were one initial denaturation step at 95°C/3 min, 25 cycles including denaturation at 95°C/30s, annealing at 52°C/1 min, and extension at 72°C/1min plus one final extension step at 72°C/7min. The PCR amplicons were purified using Agencount AMPure Beads (Beckman Coulter, Indianapolis, IN, USA). Library preparation was performed as described in the Ion Plus Fragment Library from an initial amount of 100 ng of DNA and sequenced at Ion PGM System (Thermo Fisher, Waltham, MA, USA) using an Ion 316 chip, following the manufacturer's instructions. The raw dataset is deposited at the National Biotechnology Information Center (NCBI) under the BioProject PRJNA558499. The 16S rRNA gene reads were submitted to the DADA2 version 1.18 (Callahan et al., 2016) in R version 3.6.3 (R Core Team, 2021) to obtain the amplicon sequence variants (ASVs). Reads were filtered, retrieving reads longer than 100 bp and allowing a maximum of two errors per read. The error rates were estimated, and reads were dereplicated to remove redundancy. Chimeras were removed, and taxonomy was assigned using the Silva v.138 databases (Quast et al., 2013). ASVs assigned to eukaryote, chloroplast, or mitochondria were removed using phyloseq (McMurdie and Holmes, 2013) before further analysis.

RESULTS

Oral Phenotype

In general, the phenotypic characteristics observed in the three patients were bone loss and the early loss of primary and permanent dentition, particularly for the PLSTL sisters (PLSTL1 and PLSTL2). The clinical examination of the patient who still had teeth (PLST) revealed severe gingival inflammation and hyperplasia (enlargement), spontaneous bleeding, tooth mobility, heavy dental calculus accumulation, and deep periodontal pockets (**Figures 1A–C**). The mean of PLST probing pocket depth was 8.82 mm (± 3.46), which did not regress with previous regular periodontal treatment (**Table 1**). The panoramic X-ray of the PLST sister showed loss of the permanent incisors and first molars. The remaining teeth had normal anatomy; however, they presented a "floating" appearance because of periodontal ligament and bone loss (**Figure 2**). The parents reported that the first symptoms for all sisters started in the primary dentition with teeth mobility and gingival bleeding. According to the new classification, the PLST condition was periodontitis as a manifestation of systemic disease (ICD-10 Q82.8), stage IV, generalized, and grade C.

Salivary Characteristics

Table 1 shows the systemic and the salivary characteristics of the three PLS sisters: pH, buffering capacity, amylase activity, and glucose. The patients with PLS presented hyposialia at an unstimulated salivary flow. The stimulated salivary flow resulted in severe or moderate hyposalivation. The pH was 6.97 ± 0.21 for PLST, 7.1 ± 0.17 for PLSTL1, and 6.7 ± 0.57 for PLSTL2. The PLST salivary glycemia was 1.79 ± 2.50 mg/dL, that of PLSTL1 was 10.11 ± 2.36 mg/dL, and that of PLSTL2 was 1.26 ± 1.99 mg/dL. The PLST amylase was 11,278.19 ± 2,040.13 U/dL, while that of PLSTL1 was 114,008.15 ± 2,422 U/dL and PLSTL2 was 16,016.74 ± 2,399.40U/dL (**Table 1**).

Salivary Microbiome

Both domains, *Bacteria* (90.16%) and *Archaea* (9.84%), were present in the samples. PLS salivary microbiome presented different profiles at the phyla level, with higher proportions of Fusobacteria at the PLST sister and Firmicutes and Proteobacteria at the PLSTL1 sister and Bacteroidetes and Firmicutes at the PLSTL2, as shown in **Figure 3**. The heatmap (**Figure 4**) shows the genera relative abundances of prokaryotic taxa. *Bacteroidales* F0058 (25.88%) and *Fusobacterium* (34.64%) dominated the microbiome of the PLST sister. The PLST microbiome also presented high abundances of *Tannarella* (1.12%), *Treponema* (14%), *Campylobacter* (5.53%), and *Aggregatibacter* (4.61%). *Streptococcus* dominated the microbiome of the PLSTL1 sister, comprising 32% of the total microbiome, with *Haemophilus* and *Caldivirga* in a relative abundance higher than 10%, while the PLSTL2 had higher abundances of *Lactobacillus* (8%) and *Porphyromonas* (3%). The *Archaea* domain corresponded to 10 genus-level taxa, all from the phyla Crenoarchaeota and Halobacteriota. The genus *Caldivirga* (family Thermoproteaceae) and *Sulfophobococcus* (family Desulfurococcaceae) comprised the core microbiome of

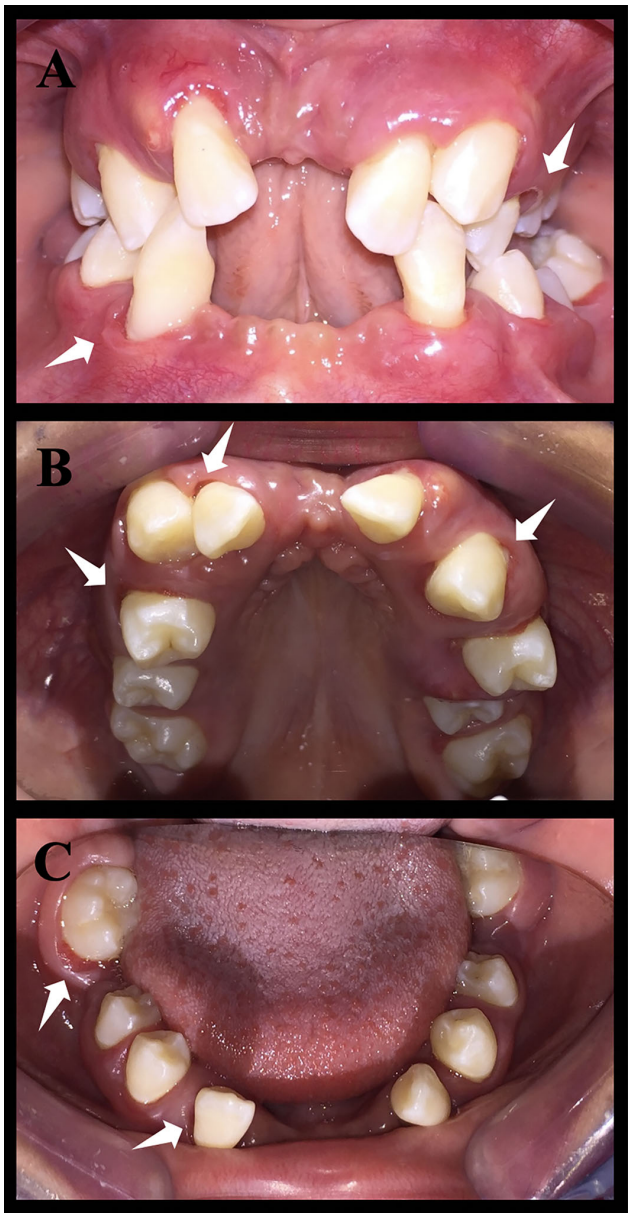


FIGURE 1 | Intraoral photos of patient Papillon–Lefèvre syndrome toothed (PLST). The white arrows indicate the periodontal pockets. It is possible to notice a significant loss of dental elements. **(A)** Occlusion photo of the patient. **(B)** Upper arch photo. **(C)** Lower arch photo.

Archaea among PLS sisters. *Caldivirga*'s relative abundance was 0.8% in PLST, 13.27% in PLSTL1, and 6.15% in PLSTL2. *Sulfophobococcus*' relative abundance was 4.45% in PLST, 0.95% in PLSTL1, and 0.15% in PLSTL2. The Venn diagram (**Figure 5**) shows that the PLSTL microbiomes shared higher ASVs than the PLST. A core microbiome could be identified, including the organisms *Prevotella*, *Fusobacterium* (high abundance in all samples, and dominant in the toothed sister), *Caldivirga* (family Thermoproteaceae, belonging to the *Archaea* domain, high abundance in PLSTL), *Streptococcus* (low abundance in PLST, 0.7%), *Oribacterium*, *Aggregatibacter*, *Dehalococcoidia* 1226B1H1-22-FL, *Haemophilus*, *Actinomyces*, *Campylobacter*, *Gemella*, *Sulfophobococcus* (*Archaea*), *Capnocytophaga* (ubiquity, but in low abundance), and *Peptococcus*.

DISCUSSION

The oral phenotype of the PLS individuals in this study followed the syndrome profile with bone loss and early edentulism for the oldest sisters (PLSTL1 and PLSTL2). All of them had palmoplantar hyperkeratosis as well. The PLST (younger sister) presented tooth mobility, bone loss, deep probing depth, inflamed gum, and feeding difficulty. The evaluation of the probing depth was complex in the PLST individual because of the spontaneous gingival bleeding, gingival hyperplasia, intense tooth mobility, and the young age of the patient. The attempts to treat periodontitis in PLS by conventional methods are frustrated in most cases, which was the case of these sisters. This study shows differences in the salivary microbiome between PLST and PLSTL (PLSTL1 and PLSTL2). For the first time, the high relative abundance of uncultivated bacteria *Bacteroidales* F0058, *Caldivirga*, and *Sulfobococcus* (the last two from the *Archaea* domain) were detected in PLS, which was possible due to the next-generation sequencing method.

The salivary characterization of the patients showed that the PLS patients have normal pH, buffering capacity, and reliability. The three PLS patients presented severe or moderate hyposalivation, and hyposialia was found in PLSTL1 and PLSTL2. This finding suggests hyposalivation as a possible characteristic of the syndrome and corroborates the study of Lundgren et al. (1996). More studies are necessary to confirm this hypothesis, as hyposalivation is associated with a higher risk of oral infections (Pedersen et al., 2018; Shimomura-Kuroki et al., 2020). Salivary hypofunction may also play a role in the diet change, resulting in malnutrition and or weight loss, affecting life quality (Pedersen et al., 2018). It is essential to highlight that saliva plays an essential role in maintaining a balanced microbiota (Pedersen et al., 2018) and promoting oral health (Kilian et al., 2016). Curiously, the level of salivary glucose was high in PLSTL1. The salivary microbiome was altered in the presence of high salivary glucose concentration, showing a decrease in bacterial load (Goodson et al., 2017). In the study of Goodson et al., the order of the microbiome reduction was in the direction of aciduric strength of bacterial species, with *Prevotella* spp. being more sensitive and *Streptococcus mutans* among the most resistant. The relative abundance of *Streptococcus*, and *Prevotella* in salivary samples of PLSTL1 and PLSTL2 followed a similar profile in our study. *Streptococcus* dominated the PLSTL1 microbiome (salivary glucose >10mg/dl) and *Prevotella* dominated the microbiome in PLSTL2 (salivary glucose, <2 mg/dl). A high abundance of *Streptococcus*, *Treponema*, and *Campylobacter* in dental biofilms of PLS individuals was described by Albandar et al. (2012), using 16S ribosomal DNA cloning and the Human Oral Microbe

TABLE 1 | Baseline characteristics of the PLS individuals and salivary parameters, pH, buffering capacity, reliability, salivary flow, amylase activity, and glucose in Papillon–Lefèvre syndrome.

Individual	PLST	PLSTL1	PLSTL2
Gender	Female	Female	Female
Age (years)	14	18	16
Number of erupted teeth	15	Edentulous	Edentulous
Probing Depth (mm)	8.82 (± 3.46)	–	–
Gingival Index	2.76 (± 0.30)	–	–
Unstimulated Saliva (mL/min)	0.35 (± 0.17)	0.21 (± 0.05)	0.26 (± 0.06)
Classification	Ideal	Hypossialia	Hypossialia
Stimulated Saliva (mL/min)	0.73 (± 0.34)	0.51 (± 0.36)	0.26 (± 0.06)
Classification	Moderate Hyposalivation	Moderate Hyposalivation	Severe Hyposalivation
pH	6.97 (± 0.21)	7.1 (± 0.17)	6.7 (± 0.57)
Buffering Capacity	4.67 (± 1.15)	4.0 (± 1.0)	4.67 (± 1.15)
Reliability	Serous	Fluid	Fluid
Glucose (mg/dL)	1.79 (± 2.50)	10.11 (± 2.35)	1.26 (± 1.99)
Amylase (U/dL)	11,278.19 (± 2,040.13)	114,008.15 (± 2,422.31)	16,016.74 (± 2,399.40)
Capillary Glucose (mg/dL)	96	106	92

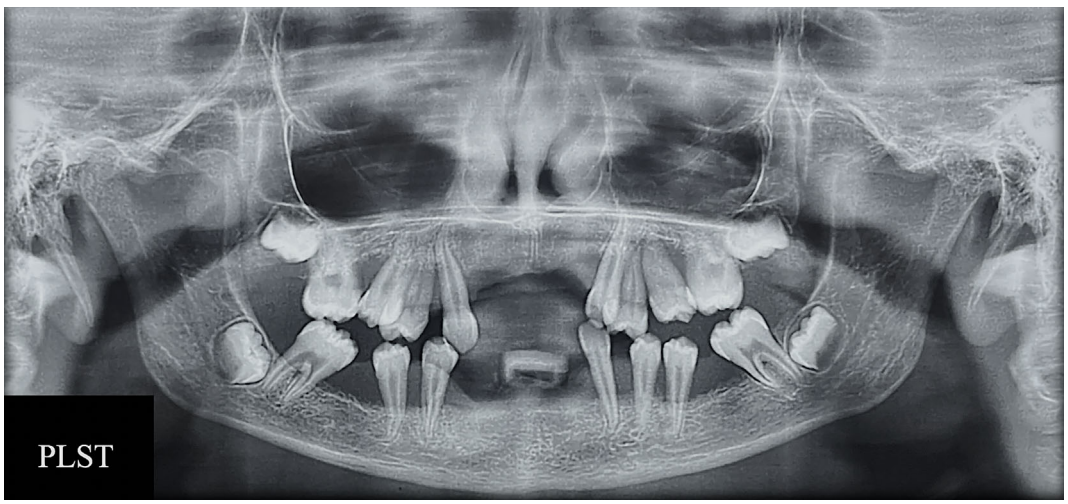


FIGURE 2 | RX exam of patient Papillon–Lefèvre syndrome toothed (PLST) showing bone loss and the appearance of teeth "floating".

Identification Microarray (HOMIM). However, our results showed fewer *Streptococcus* in the PLST sample using an next-generation sequencing (NGS) method. Albandar et al. (2012) used a cloning method, which leads to loss of diversity, particularly for uncultured microorganisms. Due to the methodological differences, we could better describe the abundance of other genera that should be involved in the dysbiosis of PLS.

Despite the high relative abundance of acidogenic *Firmicutes*, the PLST1 and PLST2 saliva pH was neutral. The mean salivary amylase concentration in PLSTL2 samples was 16,016.74 ± 2,399.40 U/dl, while PLSTL1 and PLST amylase concentrations were approximately 11,000 U/dl. An interesting study demonstrated that the copy number of the salivary amylase gene AMY was correlated with oral and gut microbiome composition and function (Poole et al., 2019). Amylase is a crucial salivary enzyme that hydrolyzes alpha bonds of starch and glycogen, beginning starch degradation in the mouth. According to the authors, the microbiomes of AMY1 low-copy individuals (AMY1 L) had enhanced capacity to break down complex carbohydrates. AMY1 high-copy subjects (AMY1 H) had a higher abundance of salivary *Porphyromonas*. Their gut microbiota had an increased abundance of resistant starch-degrading microbes. The OTUs that significantly discriminated the AMY1 H and AMY1 L groups belonged to the genera *Prevotella* (AMY1 H), *Haemophilus* (AMY1 L), *Neisseria* (AMY1 L), and *Porphyromonas* (AMY1 H). In this study, PLSTL2 had a higher abundance of *Prevotella* (23.56%) and *Porphyromonas* (2.68%) than the other PLS sisters, which corroborates the AMY H hypothesis of Poole et al. (2019). *Neisseria* was found only in PLSTL1 samples (3.28%), and *Haemophilus* had a higher relative abundance in PLSTL1 (17.12%) than in the saliva of the other PLS sisters. The amylase activity and concentration do not seem to be affected by cathepsin C deletion in mice ($CTSC^{-/-}$), neither cathepsin-C

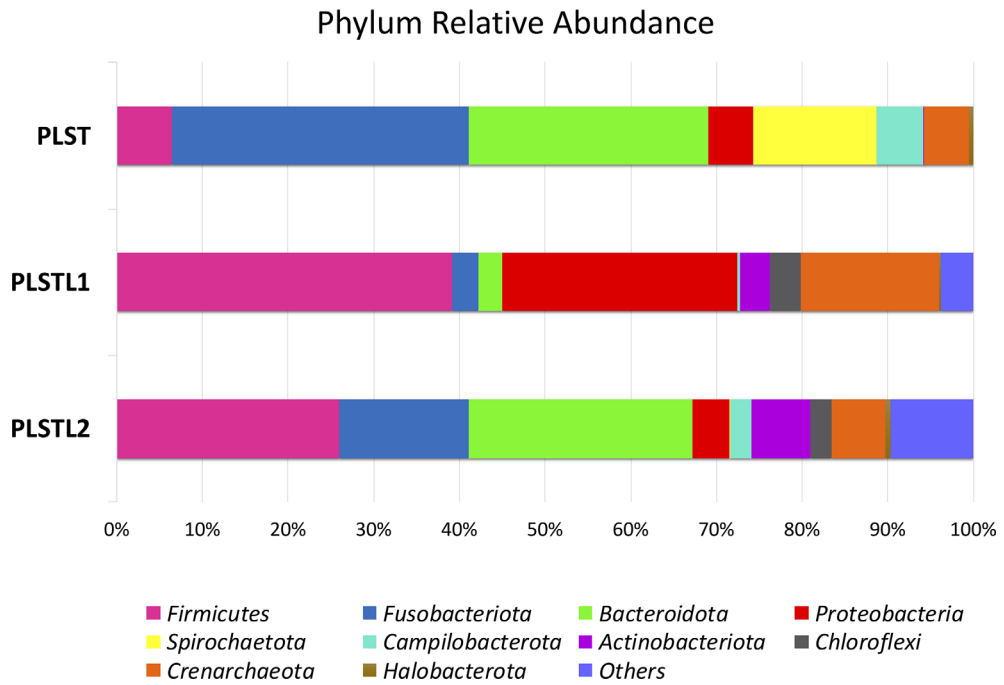


FIGURE 3 | Distribution of phyla in the salivary microbiome of the sisters with Papillon–Lefèvre syndrome. Papillon–Lefèvre syndrome toothed (PLST), Papillon–Lefèvre syndrome toothless 1 (PLSTL1), and Papillon–Lefèvre syndrome toothless 2 (PLSTL2).

colocalized in the zymogen secretory compartments (John et al., 2019). Therefore, the high amylase concentration in PLSTL2 is probably an isolated observation with no relationship with a cathepsin C mutation in PLS but deserved analysis as it can affect the microbiome. It is essential to highlight the high abundance of *Granulicatella* and *Veillonella* at the PLSTL sisters. Those genera are part of the oral mucosa specialized microbiome. Comparing PLSTL sisters with PLST, we must consider that the difference observed in the microbiome of PLST saliva is strongly related to the presence of teeth surfaces as physical supports for supra and subgingival plaques attachment.

The salivary microbial changes after PLS patients losing their teeth are expected since, it represents a loss in bacteria adherence surface (Aas et al., 2005; Socransky and Haffajee, 2005; Gazdeck et al., 2019). According to Zaura et al. (2009), the communities obtained in saliva samples are closer to mucosa communities than dental sites. The teeth loss cause changes in the oral cavity that result in the loss of bacterial taxa in soft tissues (Gazdeck et al., 2019). The biofilm that formed on teeth develops quickly and has a more significant proportion of species than edentulous individuals (O'Donnell et al., 2015). The literature suggests the demand for hard surfaces for the colonization of some species and the gingival crevice fluid from gingival sulci or periodontal pockets for the colonization of others, which can explain the main differences found in the microbiome of the partially edentulous and edentulous PLS sisters. However, the edentulous subjects (PLSTL1 and PLSTL2) that wore full dentures at the baseline of this study (for almost 2 years and 6 months, respectively) also showed striking differences in their microbiome. Socransky and Haffajee (2005) showed that, in general, the microbial profile was more similar between healthy and periodontitis subjects than edentulous subjects. The data of Socransky and Haffajee study (2005) supported the concept that the nature of the hard tissue surface in the oral cavity impacts the nature of colonizing species on hard tissue surfaces and soft tissue surfaces. Individuals who were fully edentulous and had been wearing dentures for at least 1 year showed differences in their microbiome compared with those present in samples of supragingival plaque from subjects who were periodontally healthy or had chronic periodontitis. They corroborated the denture surfaces as support for recolonization by the abundance of *S. mutans* and lactobacilli in the mouths of edentulous individuals wearing dentures. Their data indicated that *Porphyromonas gingivalis*, *Aggregatibacter actinomycetemcomitans*, and *Tannerella forsythia* could be detected in edentulous subjects 1 year or longer after all teeth have been extracted. The environmental characteristics as a consequence of the high salivary glucose of PLSTL1, the severe hyposalivation and high amylase level of PLSTL2, and other factors like the time of dentures usage and oral hygiene profiles can be responsible for the differences found in PLSTL sisters microbiome.

The PLS individuals of this study had a genetic predisposition of ineffective cathepsin C that led to inflammation and stage IV periodontitis, which suits the inflammation-mediated plymicrobial-emergence and dysbiotic exacerbation (IMPEDE) model (Van Dyke et al., 2020). The periodontitis inflammation

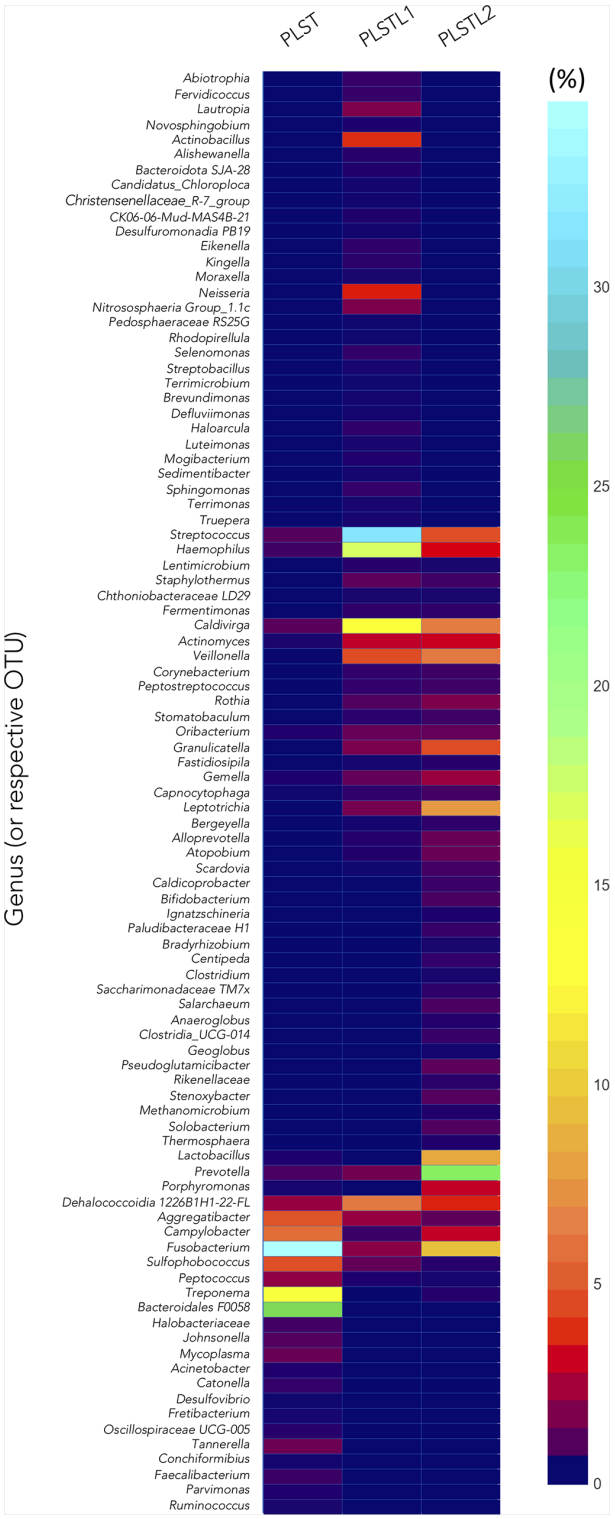


FIGURE 4 | Heatmap comparing taxa composition of saliva of PLST patients. Papillon–Lefèvre syndrome toothed (PLST), Papillon–Lefèvre syndrome toothless 1 (PLSTL1), and Papillon–Lefèvre syndrome toothless 2 (PLSTL2).

process can trigger dysbiosis, and the own dysbiosis enhances the inflammatory response (Van Dyke et al., 2020). The dysbiosis theory in periodontitis is related to the microbiological community's transition from Gram-positive commensal to Gram-negative-enriched "inflammatogenic" community (Kirst et al., 2015). Interestingly, *Bacteroidales* F0058 (clone AU126, OT274, or HOT274), presumptively a Gram-negative bacterium, massively dominated the microbiome of the PLST sister as an exclusive ASV related to the presence of teeth, with no reads in PLSTL microbiomes. In the Human Microbiome Project (HMP) cohort of 210 adults, *Bacteroidales* F0058 [provisionally assigned *Bacteroidales Neisseriaceae* (G-1) *bacterium* HMT-274] was common in adult nostrils (Escapa et al., 2018). The investigation of Kumar et al. (2003) on potential periodontal pathogens in 66 individuals showed that *Bacteroidales bacterium* HMT-274 were among the most strongly associated with the formerly chronic periodontitis, comparable or more significant than species *P. gingivalis* and *T. forsythia*. The prevalence of *B. bacterium* HMT-274 was significantly high in the study population with periodontitis (82%). Li et al. (2006) also found a significantly high prevalence of *B. bacterium* HMT-274 in subgingival plaque in formerly chronic periodontitis (77.1%) and plaque-induced gingivitis (61.5%). *Bacteroidales bacterium* HMT-274 (F0058), as-yet-uncultivated bacterium, has no potential described growth partners. Escapa et al. (2018), in their ANCOM analysis, detected only the group-specific taxon and no other species with differential relative abundance related to *B. bacterium* HMT-274. However, according to Oliveira et al. (2016), the new species *B. bacterium* HMT-274 is not a consequence of periodontitis but is likely to play a crucial role in initiating the disease.

Another striking result of our study was the presence of the *Archaea* domain, corresponding to 24 ASVs and 10 genus-level taxons, all from the phylum *Crenoarchaeota* and *Halobacteriota*. The *Haloarcula* was recently found to be part of the gastrointestinal tract microbiome (Kim et al., 2020). Our study is the first report on the presence of high abundance of *Archaea* in the human salivary microbiome. Although mostly methanogen *Archaea* has been detected at the oral cavity so far (Belmok et al., 2020), and other signs of archaeal presence in oral samples were associated with samples contamination, the high abundance of *Caldivirga* and *Sulfophobococcus* represents evidence of the importance of those organisms (or their taxonomic-related), possibly associated with dysbiotic sites. *Crenarchaeota* was already detected in human fecal samples, suggesting their presence in the microbiota of the human digestive ecosystem (Rieu-Lesme et al., 2005). In the salivary samples of the PLS sisters, both genus *Caldivirga* (family Thermoproteaceae) and *Sulfophobococcus* (family Desulfurococcaceae) comprised the core microbiome, ranging from 0.8% to 13% of relative abundance. Some studies explained part of the indirect/direct role of archaea in inflammation in specific sites and the proinflammatory potential of some species (Borrel et al., 2020). When there is a dysbiosis or infection, various body sites are known to have a higher prevalence of archaea, especially skin and oral cavity. In severe periodontitis, it was shown that the shift to anaerobic fermentative bacteria is accompanied by an increase in *Archaea* (Dabdoub et al., 2016). The PLST individual presented

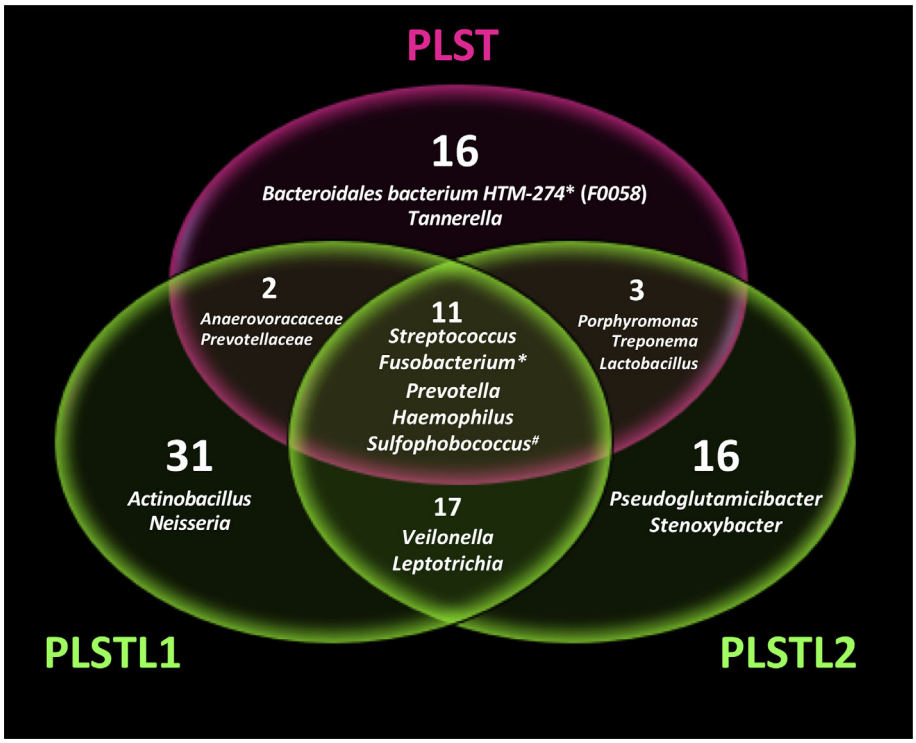


FIGURE 5 | Venn diagram showing the shared microbial genus in Papillon–Lefèvre syndrome toothed patient (PLST) and toothless (PLSTL1 and PLSTL2) sisters. Asterisk indicates high abundance in PLST sister (>20%). Number symbol indicates the highest abundance of *Archaea* domain in PLST.

over 5% relative abundance of *Crenarchaeota* (4.45% *Sulfophobococcus* and 0.8% *Caldivirga*). PLSTL1 and PLSTL2 presented 0.95% and 0.15% *of Sulfophobococcus*, while *Caldivirga* abundance was 13.27% and 6.15%, respectively. This result also suggests an archaeome-related dysbiosis in PLS that changes with edentulism. Pausan et al. (2019) discussed the importance of archaea-specific procedures, as universal approaches fail to picture the diversity of archaeal signatures in the previous analysis. Although their combination of universal primers 515F-806R showed coverage of 94.6% of archaea *in silico* (95.90% of Euryarchaeota, 94.60% of Thaumarchaeota, and 89.10% of Nanoarchaeota), the authors did not find similar detection levels experimentally. In their study, a nested PCR approach based on a first PCR with primer pair 344F-1041R, followed by a second PCR with 519F-806R, was superior for analyzing the archaeome of the gastrointestinal tract, oral cavity, and skin. In our study, the combination of universal primers 515F-806R showed a good coverage for the *Archaea* domain in the salivary samples of the PLS individuals, corroborating the *in silico* findings of Pausan et al. (2019). The high levels of archaea detection in our study may reflect a specific feature of the syndrome, probably revealed due to the advances in the archaeal sequences in databases. We believe that the oral archaeome will be easily characterized following technological advances in the future.

Dissimilatory and assimilatory sulfate reduction taxons dominated the *Bacteria* and *Archaea* core microbiome in PLS, and they were massively abundant in PLST (Kanehisa and Goto, 2000). *Desulfurococcales* and *Sulfophobococcus* are anaerobes acidophilic that ferment sugars and peptides, using organic compounds to reduce sulfur and generate hydrogen sulfide (H_2S), organic acids, and alcohols (DasSarma et al., 2009). The H_2S from *Archaea* metabolism may serve as an electron donor for a diversity of aerobic chemotrophic and anoxygenic microorganisms, forming microbiological communities in sulfidic habitats (Blohs et al., 2019). The sulfate-reducing bacteria tended to prevail in pockets with bleeding on probing, and their presence significantly correlated with pocket depth (Langendijk et al., 2000 and Dabdoub et al., 2016). Possibly, *Sulfophobococcus* and *Caldivirga* have a similar correlation in PLS and developed an essential role in coaggregation to form an aggressive subgingival plaque. *Archaea* may indirectly participate in periodontal disease by serving as a hydrogen sink, thereby facilitating the proliferation of pathogenic secondary fermenters to levels beyond the one in its absence (Dabdoub et al., 2016). Therefore, *Archaea* may have contributed to a favorable environment for the sulfate-reducing bacteria in PLST. *Staphylococcus aureus*, *Prevotella intermedia*, and *Fusobacterium nucleatum* are organic sulfate-reducing bacteria (Kushkevych et al., 2020). *Fusobacterium* dominated the PLST microbiome (34.64%), consistent with a sulfidic habitat and with the crucial role of this bacterium in oral biofilm structure and ecology. *Fusobacterium nucleatum*, for example, was found to act as a bridge between early and late periodontal biofilms colonizers (Thurnheer et al., 2019). Moreover, *F. nucleatum* triggers the production of matrix metalloproteinases by the host and has

enhanced hemolytic activity, and the production of H_2S is a key virulence trait of this bacteria in periodontitis (Thurnheer et al., 2019). Members of the genus *Treponema* are also capable of homoacetogenesis (hydrogen-consuming process), including the periodontitis pathogen, *Treponema denticola*. PLST presented high abundances of *Treponema* (14%), *Tannarella* (1.12%), and *Campylobacter* (5.53%), frequently enriched in the periodontitis-associated microbiome (Socransky and Haffajee, 2005, Hajishengallis et al., 2012). Those results support the hypothesis of possible syntrophic interactions between *Archaea, Treponema*, and other members of the red complex (Dabdoub et al., 2016), in a highly "inflammatogenic" subgingival community.

Over the years, the microbiota and host response interactions as the initial causal agent of periodontitis became more evident. However, the pathological shift from localized and contained to progressive and destructive periodontitis has not been clarified yet (Bartold and Van Dyke, 2019; Van Dyke et al., 2020). The development of aggressive periodontium destruction in PLS individuals may be a model to understand how genetics, cytokines, immunological factors, and microbiome interact for the periodontitis outcomes. In PLS, the genetic disorder of cathepsin C (CTSC) came first and led to dysbiosis in the individuals of this study. Taken together the previous literature and our findings, we hypothesized a subgingival microbiome biogeography in PLST that begins with the teeth eruption (**Figure 6**). As soon as the teeth begin to erupt in the dental arch of PLST individual, the salivary pellicle covers them providing attachment support and substrate for the ubiquitous commensal microorganisms. Thereafter, some microorganisms comprise the biofilm's biogeography on a micrometer scale, as proposed by Mark Welch et al. (2016); *Streptococcus*, *Haemophilus*, *Actinomyces*, *Capnocytophaga*, and *Fusobacterium* can colonize the tooth surface (**Figure 6**). The impaired immune system cannot provide the first line of cellular defense due to an ineffective cathepsin C. Neutrophils elastase (NE) is not active or reduced, making the infiltration of neutrophils to the adjacent tissues impossible, impairing phagocytosis, cytonemes, and NETs formation (John et al., 2019). The lack of neutrophil's defense can reduce the control of the microbial load, allowing a massive biofilm accumulation (**Figure 6**). The first colonizers proliferate, decreasing the level of O_2. The environment is highly favorable to other microorganisms coadhere or coaggregate, shifting from Gram-positive to Gram-negative anaerobes (Kirst et al., 2015). For instance, *Fusobacterium*, which abundantly dominates the microbiome of PLST, creates an anoxic lay and dictates the biofilm structure and ecology (Mark Welch et al., 2016). *Fusobacterium nucleatum*, for example, overgrew in a sulfidic environment and showed to be a high producer of hydrogen sulfide (H_2S) (Thurnheer et al., 2019). The same characteristic applies to another producer of H_2S, the archaea *Sulfophobococcus* (Blohs et al., 2019). Anaerobes that use sulfate as a terminal electron acceptor for anaerobiosis in the

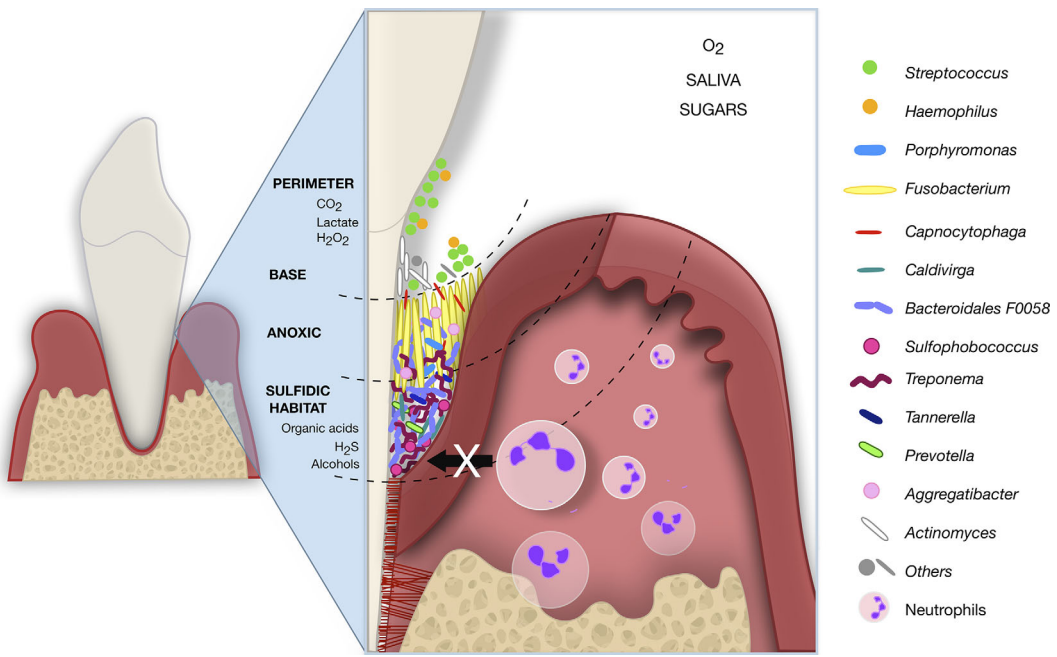


FIGURE 6 | Hypothetical subgingival microbiome biogeography of PLST. The just erupted teeth serve as attachment support for ubiquitous commensals that form an initial pellicle (*Streptococcus* and *Actinomyces*). The neutrophils infiltration and first line of defense are impaired due to an ineffective cathepsin C. The attached aerobic microbes then serve as a substrate for further colonizers. *Fusobacterium*, also a ubiquitous bacterium, plays a crucial role as a bridge between the first and late colonizers. *Fusobacterium* produces hydrogen sulfate (H_2S), overgrows, and creates a sulfidic habitat ideal for anaerobic such as *Treponema*, *Porphyromonas*, *Prevotella*, *Tannerella*, *Sulfophobococcus*, *Caldivirga*, and *Bacteroidales* F0058. The schematic image represents a hypothesis for PLST and was inspired by Mark Welch et al. (2016).

dissimilatory reduction process of sulfur (DRS) and organic compounds as source of energy find a perfect sulfidic habitat (Thurnheer et al., 2019; Kushkevych et al., 2020). The environment conditions, the synergy and microbial competition together, can favor the proliferation of other Gram-negative and anaerobic microorganisms such as *Archaea*, *Treponema*, *Bacteroidales* F0058, *Tannerella*, and *Porphyromonas* mimicking a formal aggressive periodontitis consortium in PLST (**Figure 6**).

This study suggests that host response and the resident microbiome are linked to a bidirectional imbalance between health and disease in PLS. The PLST individual does have a microbiota different from that of the periodontitis's aggressiveness previously recognized. A microbiome possibly favored by a sulfidic habitat, capable of resisting conventional periodontal treatment and antibiotic therapy. For instance, archaea are recognized as resistant to antimicrobial agents that interfere with peptidoglycan biosynthesis since their cell wall lacks peptidoglycan. The three PLS sisters were provided with intensive periodontal treatment, mechanical therapy, and oral hygiene instructions, including methods for denture cleaning before and after the study. The attempts of treatment prior to the study were unsuccessful, even conventional antibiotic therapy. All consequences of an environment where neutrophils and their frontline defensive mechanisms are inexistent or substantially reduced. The phenotype of Papillon–Lefèvre Syndrome highlights the microbiome dysbiosis and the fundamental role that neutrophils play in maintaining oral and skin health. Considering the high relative abundance in PLST, the genus *Fusobacterium*, *Treponema*, *Tannerella*, and *Sulfophobococcus* are possible candidates to form a consortium with *Bacteroidales* F0058; this is a hypothesis that deserves future investigation. Undeniable, the archaeome is ubiquitous in the salivary microbiome and more complex than previously thought. The knowledge about the PLS microbiome is exceptionally relevant in the search for specific successful treatments. New broad-spectrum antimicrobial agents and second-generation retinoid acitretin (vitamin A derivative) are potential alternatives to provide a better prognosis for the outcomes of PLS.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Faculty of Health Sciences of the University of Brasília (FS-UnB; no. 2.974.167). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

GL conceived ideas for this manuscript and was involved in data interpretation and manuscript drafting. LMS conducted experiments. GL was involved in the patient's diagnosis, treatment, and manuscript drafting. LB conducted experiments and taxonomy assessment. LM conducted experiments and taxonomy assessment. LO was involved in study design, experiments, and manuscript drafting. ND-T conceived ideas for this manuscript and was involved in study design, data interpretation, and manuscript drafting. LPS conceived ideas for this manuscript and was involved in study design, experiments, data interpretation, and manuscript drafting. All authors contributed to the article and approved the submitted version.

FUNDING

The study was supported by Institutional funds from the University of Brasília (PROAP-DGP/UnB No. 02/2019).

ACKNOWLEDGMENTS

We thank Dr. Floyd E. Dewhirst from The Forsyth Institute (Boston, MA, USA) that kindly provided important information about the *Bacteroidales bacterium* HMT-274 and shared relevant scientific literature on this subject. We also thank High Performance Computing Lab—LAD/PUCRS for allowing access to run the high-throughput sequences analyses. LB thanks PEGA/PUCRS.

REFERENCES

Aas, J. A., Paster, B. J., Stokes, L. N., Olsen, I., and Dewhirst, F. E. (2005). Defining the Normal Bacterial Flora of the Oral Cavity. *J. Clin. Microbiol.* 43 (11), 5721–5732. doi: 10.1128/JCM.43.11.5721-5732.2005

AIBarrak, Z. M., Alqarni, A. S., Chalisserry, E. P., and Anil, S. (2016). Papillon–Lefèvre Syndrome: A Series of Five Cases Among Siblings. *J. Med. Case Rep.* 10 (1), 260. doi: 10.1186/s13256-016-1051-z

Albandar, J. M., Khattab, R., Monem, F., Barbuto, S. M., and Paster, B. J. (2012). The Subgingival Microbiota of Papillon-Lefèvre Syndrome. *J. Periodontol.* 83 (7), 902–908. doi: 10.1902/jop.2011.110450

Bartold, P. M., and Van Dyke, T. E. (2019). An Appraisal of the Role of Specific Bacteria in the Initial Pathogenesis of Periodontitis. *J. Clin. Periodontol.* 46 (6), 6–11. doi: 10.1111/jcpe.13046

Bates, S. T., Berg-Lyons, D., Caporaso, J. G., Walters, W. A., Knight, R., and Fierer, N. (2011). Examining the Global Distribution of Dominant Archaeal Populations in Soil. *ISME. J.* 5 (5), 908–917. doi: 10.1038/ismej.2010.171

Belmok, A., de Cena, J. A., Kyaw, C. M., and Damé-Teixeira, N. (2020). The Oral Archaeome: A Scoping Review. *J. Dent. Res.* 99 (6), 630–643. doi: 10.1177/0022034520910435

Bhavsar, M. V., Brahmbhatt, N. A., Sahayata, V. N., and Bhavsar, N. V. (2013). Papillon-Lefevre Syndrome: Case Series and Review of Literature. *J. Indian. Soc. Periodontol.* 17 (6), 806. doi: 10.4103/0972-124X.124530

Blohs, M., Moissl-Eichinger, C., Mahnert, A., Spang, A., Dombrowski, N., Krupovic, M., et al. (2019). "Archaea – An Introduction," in *Encyclopedia of Microbiology, 4th ed.* Ed. T. M. Schmidt (Cambridge, MA: Academic Press), 243–252.

Borrel, G., Brugère, J.-F., Gribaldo, S., Schmitz, R. A., and Moissl-Eichinger, C. (2020). The Host-Associated Archaeome. *Nat. Rev. Microbiol.* 18 (11), 622–636. doi: 10.1038/s41579-020-0407

Bullón, P., Castejón-Vega, B., Román-Malo, L., Jimenez-Guerrero, M. P., Cotán, D., Forbes-Hernandez, T. Y., et al. (2018). Autophagic Dysfunction in Patients With Papillon-Lefevre Syndrome is Restored by Recombinant Cathepsin C Treatment. *J. Allergy Clin. Immunol.* 142 (4), 1131–1143. doi: 10.1016/j.jaci.2018.01.018

Callahan, B., McMurdie, P., and Holmes, S. (2017). Exact Sequence Variants Should Replace Operational Taxonomic Units in Marker-Gene Data Analysis. *ISME. J.* 11, 2639–2643. doi: 10.1038/ismej.2017.119

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., and Holmes, S. P. (2016). DADA2: High-Resolution Sample Inference From Illumina Amplicon Data. *Nat. Methods* 13 (7), 581–583. doi: 10.1038/nmeth.3869

Caton, J. G., Armitage, G., Berglundh, T., Chapple, I. L., Jepsen, S., Kornman, K. S., et al. (2018). A New Classification Scheme for Periodontal and Peri-Implant Diseases and Conditions–Introduction and Key Changes From the 1999 Classification. *J. Periodontol.* 89, S1–S8. doi: 10.1111/jcpe.12935

Dabdoub, S., Ganesan, S., and Kumar, P. (2016). Comparative Metagenomics Reveals Taxonomically Idiosyncratic Yet Functionally Congruent Communities in Periodontitis. *Sci. Rep.* 6, 38993. doi: 10.1038/srep38993

DasSarma, S., Coker, J. A., and DasSarma, P. (2009). "Archaea (Overview)," in *Encyclopedia of Microbiology*, *3rd ed.* Ed. T. M. Schmidt (Cambridge, MA: Academic Press), 1–23.

Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C., Yu, W.-H., et al. (2010). The Human Oral Microbiome. *J. Bacteriol.* 192 (19), 5002–5017. doi: 10.1128/JB.00542-10

Escapa, I. F., Chen, T., Huang, Y., Gajare, P., Dewhirst, F. E., and Lemon, K. P. (2018). New Insights Into Human Nostril Microbiome From the Expanded Human Oral Microbiome Database (eHOMD): A Resource for the Microbiome of the Human Aerodigestive Tract. *Msystems* 3 (6), e00187–18. doi: 10.1128/mSystems.00187-18

Fischer, J., Blanchet-Bardon, C., Prud'homme, J.-F., Pavek, S., Steijlen, P. M., Dubertret, L., et al. (1997). Mapping of Papillon-Lefevre Syndrome to the Chromosome 11q14 Region. *Eur. J. Hum. Genet.* 5, 156–160. doi: 10.1159/000484751

Fricker, A. M., Podlesny, D., and Fricke, W. F. (2019). What is New and Relevant for Sequencing-Based Microbiome Research? A Mini-Review. *J. Adv. Res.* 19, 105–112. doi: 10.1016/j.jare.2019.03.006

Gazdeck, R. K., Fruscione, S. R., Adami, G. R., Zhou, Y., Cooper, L. F., and Schwartz, J. L. (2019). Diversity of the Oral Microbiome Between Dentate and Edentulous Individuals. *Oral. Dis.* 25 (3), 911–918. doi: 10.1111/odi.13039

Goodson, J. M., Hartman, M.-L., Shi, P., Hasturk, H., Yaskell, T., Vargas, J., et al. (2017). The Salivary Microbiome is Altered in the Presence of a High Salivary Glucose Concentration. *PloS One* 12 (3), e0170437. doi: 10.1371/journal.pone.0170437

Hajishengallis, G., Darveau, R. P., and Curtis, M. A. (2012). The Keystone-Pathogen Hypothesis. *Nat. Rev. Microbiol.* 10 (10), 717–725. doi: 10.1038/nrmicro2973

Hamamy, H., Antonarakis, S. E., Cavalli-Sforza, L. L., Temtamy, S., Romeo, G., Kate, L. P. T., et al. (2011). Consanguineous Marriages, Pearls and Perils: Geneva International Consanguinity Workshop Report. *Genet. Med.* 13, 841–847. doi: 10.1097/GIM.0b013e318217477f

Hart, T. C., Hart, P. S., Bowden, D. W., Michalec, M. D., Callison, S. A., Walker, S. J., et al. (1999). Mutations of the Cathepsin C Gene Are Responsible for Papillon-Lefevre Syndrome. *J. Med. Genet.* 36 (12), 881–887.

Hart, T. C., and Shapira, L. (2000). Papillon-Lefèvre Syndrome. *Periodontol* 6 (1), 88–100. doi: 10.1111/j.1600-0757.1994.tb00029.x

Hattab, F. N. (2019). Papillon-Lefèvre Syndrome: From Then Until Now. *Stomatol. Dis. Sci.* 3, 1. doi: 10.20517/2573-0002.2018.22

Hoare, A., Marsh, P. D., and Diaz, P. I. (2018). Ecological Therapeutic Opportunities for Oral Diseases. Bugs as Drugs: Therapeutic Microbes for the Prevention and Treatment of Disease. *Microbiol. Spectr.* 5 (4), 235–265. doi: 10.1128/microbiolspec.BAD-0006-2016

John, D. S., Aschenbach, J., Krüger, B., Sendler, M., Weiss, F. U., Mayerle, J., et al. (2019). Deficiency of Cathepsin C Ameliorates Severity of Acute Pancreatitis by Reduction of Neutrophil Elastase Activation and Cleavage of E-Cadherin. *J. Bio. Chem.* 294 (2), 697–707. doi: 10.1074/jbc.RA118.004376

Jordan, R. C. K. (2004). Diagnosis of Periodontal Manifestations of Systemic Diseases. *Periodontol* 34 (1), 217–229. doi: 10.1046/j0906-6713.2002.003433.x

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

Kaplan, M. J., and Radic, M. (2012). Neutrophil Extracellular Traps: Double-Edged Swords of Innate Immunity. *J. Immunol.* 189 (6), 2689–2695. doi: 10.4049/jimmunol.1201719

Kilian, M., Chapple, I., Hannig, M., Marsh, P., Meuric, V., Pedersen, A., et al. (2016). The Oral Microbiome–an Update for Oral Healthcare Professionals. *Br. Dent. J.* 221 (10), 657–666. doi: 10.1038/sj.bdj.2016.865

Kim, J. Y., Whon, T. W., Lim, M. Y., Kim, Y. B., Kim, N., Kwon, M.-S., et al. (2020). The Human Gut Archaeome: Identification of Diverse Haloarchaea in Korean Subjects. *Microbiome* 8, 114. doi: 10.1186/s40168-020-00894-x

Kirst, M. E., Li, E. C., Alfant, B., Chi, Y.-Y., Walker, C., Magnusson, I., et al. (2015). Dysbiosis and Alterations in Predicted Functions of the Subgingival Microbiome in Chronic Periodontitis. *App. Environ. Microbio.* 81 (2), 783–793. doi: 10.1128/AEM.02712-14

Kumar, P., Griffen, A., Barton, J., Paster, B., Moeschberger, M., and Leys, E. (2003). New Bacterial Species Associated With Chronic Periodontitis. *J. Den. Res.* 82 (5), 338–344. doi: 10.1177/154405910308200503

Kushkevych, I., Cejnar, J., Treml, J., Dordevic´, D., Kollar, P., and Vítezová, M. (2020). Recent Advances in Metabolic Pathways of Sulfate Reduction in Intestinal Bacteria. *Cells* 9, 698. doi: 10.3390/cells9030698

Langendijk, P. S., Hanssen, J. T., and van der Hoeven, J. S. (2000). Sulfate-Reducing Bacteria in Association With Human Periodontitis. *J. Clin. Periodontol.* 27, 943–950. doi: 10.1034/j.1600-051z.2000.027012943.x

Li, C., Liang, J., and Jiang, Y. (2006). Association of Uncultivated Oral Phylotypes AU126 and X112 With Periodontitis. *Oral. Dis.* 12 (4), 371–374. doi: 10.1111/j.1601-0825.2005.01205.x

Lundgren, T., Twetman, S., Johansson, I., Crossner, C.-G., and Birkhed, D. (1996). Saliva Composition in Children and Young Adults With Papillon-Lefèvre Syndrome. *J. Clin. Periodontol.* 23, 1068–1072. doi: 10.1111/j.1600-051X.1996.tb01805.x

Mark Welch, J. L., Rossetti, B. J., Rieken, C. W., Dewhirst, F. E., and Borisy, G. G. (2016). Biogeography of a Human Oral Microbiome at the Micron Scale. *Proc. Natl. Acad. Sci.* 113, E791. doi: 10.1073/pnas.1522149113

McMurdie, P. J., and Holmes, S. (2013). Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PloS One* 8 (4), e61217. doi: 10.1371/journal.pone.0061217.Print2013

Navazesh, M., and Kumar, S. K. (2008). Measuring Salivary Flow: Challenges and Opportunities. *J. Am. Dent. Assoc.* 139, 35S–40S. doi: 10.14219/jada.archive.2008.0353

O'Donnell, L. E., Robertson, D., Nile, C. J., Cross, L. J., Riggio, M., Sherriff, A., et al. (2015). The Oral Microbiome of Denture Wearers is Influenced by Levels of Natural Dentition. *PloS One* 10 (9), e0137717. doi: 10.1371/journal.pone.0137717

Oliveira, R., Fermiano, D., Feres, M., Figueiredo, L., Teles, F., Soares, G., et al. (2016). Levels of Candidate Periodontal Pathogens in Subgingival Biofilm. *J. Dent. Res.* 95 (6), 711–718. doi: 10.1177/0022034516634619

Papillon, M., and Lefevre, P. (1924). Two Cases of Symmetrically Familial Palmar and Plantar Hyperkeratosis (Meleda Disease) Within Brother and Sister Combined With Severe Dental Alterations in Both Cases. *Bull. Soc. Fr. Dermatol. Syphiligr.* 31 (2), 82–87.

Pausan, M. R., Csorba, C., Singer, G., Till, H., Schöpf, V., Santigli, E., et al. (2019). Exploring the Archaeome: Detection of Archaeal Signatures in the Human Body. *Front. Microbiol.* 10, 2796. doi: 10.3389/fmicb.2019.02796

Pedersen, A. M. L., Sørensen, C. E., Proctor, G., Carpenter, G., and Ekström, J. (2018). Salivary Secretion in Health and Disease. *J. Oral. Rehabil.* 45 (9), 730–746. doi: 10.1111/joor.12664

Poole, A. C., Goodrich, J. K., Youngblut, N. D., Luque, G. G., Ruaud, A., Sutter, J. L., et al. (2019). Human Salivary Amylase Gene Copy Number Impacts Oral and Gut Microbiomes. *Cell Host Microbe* 25 (4), 553–564. e557. doi: 10.1016/j.chom.2019.03.001

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools. *Nucleic Acids Res.* 41 (Database issue), D590–D596. doi: 10.1093/nar/gks1219

R Core Team (2021) *R: A Language and Environment for Statistical Computing* (Vienna, Austria: R Foundation for Statistical Computing). Available at: https://www.R-project.org/ (Accessed Accessed: April 05, 2021).

Rieu-Lesme, F., Delbès, C., and Sollelis, L. (2005). Recovery of Partial 16s rDNA Sequences Suggests the Presence of Crenarchaeota in the Human Digestive Ecosystem. *Curr. Microbiol.* 51, 317–321. doi: 10.1007/s00284-005-0036-8

Robertson, K., Drucker, D., James, J., Blinkhorn, A., Hamlet, S., and Bird, P. (2001). A Microbiological Study of Papillon-Lefevre Syndrome in Two Patients. *J. Clin. Pathol.* 54 (5), 371–376. doi: 10.1136/jcp.54.5.371

Roberts, H., White, P., Dias, I., McKaig, S., Veeramachaneni, R., Thakker, N., et al. (2016). Characterization of Neutrophil Function in Papillon-Lefèvre Syndrome. *J. Leukoc. Biol.* 100 (2), 433–444. doi: 10.1189/jlb.5A1015-489R

Sørensen, O. E., Clemmensen, S. N., Dahl, S. L., Østergaard, O., Heegaard, N. H., Glenthøj, A., et al. (2014). Papillon-Lefevre Syndrome Patient Reveals Species-Dependent Requirements for Neutrophil Defenses. *J. Clin. Investig.* 124 (10), 4539–4548. doi: 10.1172/JCI76009

Shimomura-Kuroki, J., Nashida, T., Miyagawa, Y., Morita, T., and Hayashi-Sakai, S. (2020). Analysis of Salivary Factors Related to the Oral Health Status in Children. *J. Oral. Sci.* 62 (2), 226–230. doi: 10.2334/josnusd.18-0293

Smalla, K., Cresswell, N., Mendonca-Hagler, L. C., Wolters, A., and Elsas, J. V. (1993). Rapid DNA Extraction Protocol From Soil for Polymerase Chain Reaction-Mediated Amplification. *J. Appl. Bacteriol.* 74 (1), 78–85. doi: 10.1111/j.1365-2672.1993.tb02999.x

Socransky, S. S., and Haffajee, A. D. (2005). Periodontal Microbial Ecology. *Periodontol* 38 (1), 135–187. doi: 10.1111/j.1600-0757.2005.00107.x

Thurnheer, T., Karygianni, L., Flury, M., and Belibasakis, G. N. (2019). Fusobacterium Species and Subspecies Differentially Affect the Composition and Architecture of Supra- and Subgingival Biofilms Models. *Front. Microbiol.* 10:1716. doi: 10.3389/fmicb.2019.01716

Toomes, C., James, J., Wood, A. J., Wu, C. L., McCormick, D., Lench, N., et al. (1999). Loss-Of-Function Mutations in the Cathepsin C Gene Result in Periodontitis and Palmoplantar Keratosis. *Nat. Genet.* 23 (4), 421–424. doi: 10.1038/70525

Tumen, D. S., Tumen, E. C., Gunay, A., Lacin, N., and Cetin, S. G. (2015). The Typical Appearance and CBCT Images of the Patient With Papillon-Lefevre Syndrome: A Case Report. *J. Int. Dent. Med. Res.* 8 (3), 128.

Van Dyke, T. E., Bartold, P. M., and Reynolds, E. C. (2020). The Nexus Between Periodontal Inflammation and Dysbiosis. *Front. Immunol.* 11, 511. doi: 10.3389/fimmu.2020.00511

Yu, H., He, X., Liu, X., Zhang, H., Shen, Z., Shi, Y., et al. (2021). A Novel Missense Variant in Cathepsin C Gene Leads to PLS in a Chinese Patient: A Case Report and Literature Review. *Mol. Genet. Genomic* 9 (7), e1686. doi: 10.1002/mgg3.1686

Zaura, E., Keijser, B. J., Huse, S. M., and Crielaard, W. (2009). Defining the Healthy Core Microbiome" of Oral Microbial Communities. *BMC Microbiol.* 9 (1), 1–12. doi: 10.1186/1471-2180-9-259

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Cellular and Infection Microbiology

ORIGINAL RESEARCH
published: 08 October 2021
doi: 10.3389/fcimb.2021.727732

Comparison of Red-Complex Bacteria Between Saliva and Subgingival Plaque of Periodontitis Patients: A Systematic Review and Meta-Analysis

Yaling Jiang[1,2], Bingqing Song[1], Bernd W. Brandt[2], Lei Cheng[1], Xuedong Zhou[1], Rob A. M. Exterkate[2], Wim Crielaard[2] and Dong Mei Deng[2]**

[1] State Key Laboratory of Oral Diseases & National Clinical Research Center for Oral Diseases & Department of Cariology and Endodontics, West China Hospital of Stomatology, Sichuan University, Chengdu, China, [2] Department of Preventive Dentistry, Academic Center for Dentistry Amsterdam (ACTA), University of Amsterdam and Vrije Universiteit Amsterdam, Amsterdam, Netherlands

OPEN ACCESS

Edited by:
Ulvi Kahraman Gürsoy, University of Turku, Finland

Reviewed by:
Chen Li, China Medical University, China
Nursen Topcuoglu, Istanbul University, Turkey

***Correspondence:**
Lei Cheng
chenglei@scu.edu.cn
Dong Mei Deng
d.deng@acta.nl

Specialty section:
This article was submitted to Microbiome in Health and Disease, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: *19 June 2021*
Accepted: *14 September 2021*
Published: *08 October 2021*

Citation:
Jiang Y, Song B, Brandt BW, Cheng L, Zhou X, Exterkate RAM, Crielaard W and Deng DM (2021) Comparison of Red-Complex Bacteria Between Saliva and Subgingival Plaque of Periodontitis Patients: A Systematic Review and Meta-Analysis. Front. Cell. Infect. Microbiol. 11:727732. doi: 10.3389/fcimb.2021.727732

The development of periodontitis is associated with an imbalanced subgingival microbial community enriched with species such as the traditionally classified red-complex bacteria (*Porphyromonas gingivalis*, *Tannerella forsythia*, and *Treponema denticola*). Saliva has been suggested as an alternative to subgingival plaque for the microbial analysis due to its easy and non-invasive collection. This systematic review aims to determine whether the levels of red-complex bacteria assessed using saliva reflect those in subgingival plaque from periodontitis patients. The MEDLINE, EMBASE, and Cochrane Library databases were searched up to April 30, 2021. Studies were considered eligible if microbial data of at least one of the red-complex species were reported in both saliva and subgingival plaque from periodontitis patients, based on DNA-based methods. Of the 17 included studies, 4 studies used 16S rRNA gene sequencing techniques, and the rest used PCR-based approaches. The detection frequency of each red-complex species in periodontitis patients was reported to be > 60% in most studies, irrespective of samples types. Meta-analyses revealed that both detection frequencies and relative abundances of red-complex bacteria in saliva were significantly lower than those in subgingival plaque. Moreover, the relative abundances of all 3 bacterial species in saliva showed significantly positive correlation with those in subgingival plaque. In conclusion, current evidence suggests that one-time saliva sampling cannot replace subgingival plaque for microbial analysis of the red-complex bacteria in periodontitis patients. Given the positive microbial associations between saliva and subgingival plaque, a thorough review of longitudinal clinical studies is needed to further assess the role of saliva.

Keywords: periodontitis, *Porphyromonas gingivalis*, *Tannerella forsythia*, *Treponema denticola*, 16S rRNA gene amplicon sequencing, real-time PCR

1 INTRODUCTION

Periodontitis is one of the most prevalent oral infectious diseases, affecting over 740 million people worldwide (Kassebaum et al., 2014). It is a chronic inflammation, associated with dysbiotic subgingival biofilms, resulting in progressive and irreversible destruction of tooth supporting tissues (Vieira Colombo et al., 2016; Tonetti et al., 2017). Although it is not clear whether dysbiotic biofilms initiate the disease or are a consequence of the disease, it is well established that at diseased status, the subgingival microbiota is enriched with gram-negative, proteolytic bacteria, while in healthy situation, the microbiota is mainly composed of gram-positive bacteria (Curtis et al., 2020; Van Dyke et al., 2020). A recent review collected existing evidence and proposed an "Inflammation-Mediated Polymicrobial-Emergence and Dysbiotic-Exacerbation" (IMPEDE) model, which included the microbial element and complemented the current clinical classification of periodontitis (Van Dyke et al., 2020). According to this model, local inflammation drives an initial shift in microbial composition and the formation of periodontal pocket exacerbates this microbial shift by further enriching disease-associated species.

To determine the compositional shift of periodontitis-related microbiota, subgingival plaque obtained from diseased pockets has been considered as the most representative sample. However, collecting subgingival plaque is invasive and requires specialized training for proper sampling. Moreover, reports have shown that the quality and quantity of collected plaque samples may be greatly influenced by the collection methods (Renvert et al., 1992; van der Horst et al., 2013). Compared to subgingival plaque, saliva is much better accessible and can be collected noninvasively in larger quantity. Since the collection of saliva does not require special sampling tools, sample quality could be less influenced by sampling methods or operators. Saliva has been proposed as an alternative to subgingival plaque for studying the association between oral microbes and periodontal disease since 1998 (Umeda et al., 1998). It was hypothesized that microbes residing in a periodontal pocket could be spread, washed out or spilt over into saliva (Haririan et al., 2014; Li et al., 2015). Multiple clinical studies have compared the levels of important periodontal microbes using samples collected from subgingival pockets and saliva, but the results were inconsistent. For example, Umeda et al. (1998) reported that *Porphyromonas gingivalis* and *Treponema denticola* were detected more often in saliva than in subgingival plaque samples, whereas Nickles et al. (2017) reported the opposite. Moreover, the open-ended DNA sequencing techniques developed in the past decades have revealed that different niches in the oral cavity harbor considerably different microbial communities with distinct microbial composition (Huttenhower et al., 2012; Mark Welch et al., 2019). Differential microbial profiles of saliva and subgingival plaque have been demonstrated by several studies (Segata et al., 2012; Simón-Soro et al., 2013). Therefore, a systematic review that includes recent studies using sequencing techniques is needed to objectively assess microbial compositions of saliva and subgingival plaque.

So far, the most frequently examined microbes in clinical samples obtained from periodontitis patients are *P. gingivalis*, *Tannerella forsythia* and *T. denticola*. These 3 species were grouped as red-complex bacteria in 1998 (Socransky et al., 1998) based on the evidence that they were frequently isolated together and were strongly associated with periodontitis. Members of the red-complex group have been the main target in many clinical studies which performed the comparison between saliva and subgingival plaque samples. In the past, techniques employed to examine the levels of these bacteria were mainly targeted approaches, such as bacterial culture, immunological assays, PCR and quantitative real-time PCR (qPCR) (Suchett-Kaye et al., 2001). Among them, PCR and qPCR were used most frequently, since these techniques have better sensitivity and specificity as compared to other methods (Loesche, 1992; Boutaga et al., 2003). In addition, the abovementioned 16S rRNA gene sequencing techniques have been increasingly applied in clinical studies (Li et al., 2015; Belstrøm et al., 2017), since they offer the possibility to profile the entire microbiota besides specific bacterial species of interest.

This systematic review aimed to evaluate clinical evidence on the levels of red-complex bacteria in saliva and subgingival plaque samples collected from patients with periodontitis. We focused on studies which used DNA-based (targeted and open-ended) methods for microbial identification.

2 MATERIALS AND METHODS

This systematic review was performed following the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) statement (Moher et al., 2009), and was registered at the National Institute for Health Research PROSPERO, International Prospective Register of Systematic Reviews (registration number: CRD42020219510).

2.1 Search Strategy

The MEDLINE (*via* PubMed), EMBASE, and Cochrane Library databases were searched up to April 30, 2021 by two independent researchers (YJ and BS), using the search strategy described in **Supplementary Table S1**. In addition, a manual search of the reference list of the included studies was conducted.

2.2 Study Selection

Studies that met the following criteria were included: 1) population: humans with periodontitis; 2) exposure: saliva samples; 3) comparison: subgingival plaque samples; 4) outcome: microbial data of at least one of the red-complex bacteria that was obtained using a DNA-based method; 5) study design: clinical studies of any design, except case report and case series.

Exclusion criteria were: 1) studies without periodontal diagnosis of the participants; 2) studies including participants who had an explicit diagnosis of any systemic disease or systemic condition, such as pregnancy; 3) studies including participants who used medication (e.g., antibiotics) or the medication status

of participants was not mentioned; 4) studies with no baseline data in case of a prospective or interventional design; 5) not full-text publications (e.g., conference abstracts); 6) studies published in languages other than English. In addition, publications with overlapped data were identified at the phase of full-text screening. These studies were conducted by the same group of authors, and the data (a part or all) reported in different publications were obtained from the same group of subjects. In order to avoid duplicate data extraction, only the publication on a larger number of subjects was included.

The selection of studies was performed in two steps based on the above inclusion and exclusion criteria. In the first step, articles were screened on the basis of title and abstract, using a Web platform (rayyan.qcri.org) (Ouzzani et al., 2016). In the second step, the selected studies underwent full-text evaluation.

2.3 Data Extraction and Methodological Quality Assessment

A customized data extraction form was used to collect the following information from each included study: 1) characteristics of the study (e.g., author, year of publication, and study location); 2) characteristics of the participants (e.g., number, periodontal diagnosis, and clinical parameters); 3) methodological features of the study (e.g., method of saliva and subgingival plaque sample collection, method to evaluate red-complex bacteria); 4) microbial outcomes. For prospective and interventional studies, only data from the baseline measurement were extracted for analysis. When needed, the corresponding author(s) were contacted for the missing data.

The methodological qualities of all included studies were assessed using the 8-item "Critical Appraisal Checklist for Analytical Cross-Sectional Studies" by the Joanna Briggs Institute (JBI) (Moola et al., 2020). The answer to each question was "Yes", "No", or "Unclear", and an overall rating score was given to each study which equals to the total number of "Yes" answers given, ranging from 0 to 8. The scores of 0–3, 4–5, and 6–8 were classified as low, medium, and high quality of studies, respectively (Yazdanian et al., 2020).

The steps mentioned above, including study selection, data extraction and methodological quality assessment were conducted independently by two researchers (YJ and BS), and any disagreement was resolved through discussion. If disagreement persisted, another researcher (DD) was consulted to achieve consensus.

2.4 Summary Outcome Measures and Statistical Analysis

Based on the results of included studies, we summarized 3 types of microbial outcomes: detection frequency, bacterial count and/or relative abundance of each red-complex bacteria.

Detection frequency refers to the percentage of subjects positive for a specific microorganism. It was reported in studies using either targeted PCR-based approaches or 16S rRNA gene sequencing techniques. Bacterial count provides the (semi-) quantity of a specific microorganism in a sample. This outcome parameter was reported in the studies using qPCR techniques. For studies using 16S rRNA gene sequencing techniques, two types of data were extracted: detection frequency and relative abundance.

Meta-analyses were performed to assess the statistical differences in the detection frequency and relative abundance of each red-complex species between saliva and subgingival plaque samples, using RevMan software version 5.4 (The Cochrane Collaboration; Copenhagen, Denmark). Heterogeneity among studies was assessed by chi-squared test and inconsistency index I^2. Values of $I^2 < 30\%$, 30%–60%, and > 60% were considered as low, moderate and large heterogeneity, respectively (Higgins et al., 2003). When the heterogeneity was significant ($p < 0.1$), a random-effects model (DerSimonian-Laird method) was applied to examine the overall effect; otherwise, a fixed-effects model (Mantel-Haenszel method) was used. The odds ratio (OR) for detection frequency and mean difference (MD) for relative abundance were calculated at 95% confidence intervals (CI). Differences were considered statistically significant if $p < 0.05$.

3 RESULTS

3.1 Results of Search and Study Selection

The literature search of three electronic databases identified 2520 records in total (**Figure 1**). After removing duplicates, 1680 articles were retained for title and abstract screening, from which 1634 articles were excluded and the remaining 46 were assessed further in full-text reading. Of these 46 studies, 30 were excluded based on the eligibility criteria (**Supplementary Table S2**). One study (Umeda et al., 1998) was identified additionally from the manual search of the reference lists of the selected studies. Therefore, 17 studies were finally included in this systematic review.

3.2 Quality Assessment

Figure 2 shows the methodological quality assessment of the included studies, which presented the answer to each appraisal criteria as well as an overall rating of each study. Nine out of 17 studies had high quality, 8 studies had medium quality and no study had low quality.

3.3 Characteristics of the Included Studies

The main characteristics of the included studies are presented in **Table 1**. Most studies had a cross-sectional design (n = 14); the rest had a prospective (n = 2) or interventional (n = 1) design.

Out of the 17 studies, 14 studies reported the classification of periodontitis: 7 studies included patients with chronic periodontitis (CP) only, 1 study with aggressive periodontitis (AgP) only, and 5 studies with both CP and AgP. Different from the other 13 studies, Choi et al. (2020) classified the disease as moderate and severe periodontitis. Among these 14 studies, only 10 studies specified their diagnostic criteria: 5 studies (Takeuchi et al., 2001; Feng et al., 2014; Haririan et al., 2014; Li et al., 2015; Yang et al., 2016) followed the criteria of the 1999 International Classification of Periodontal Diseases and Conditions (Armitage, 1999), 2 studies (Belstrøm et al., 2017; Belstrøm et al., 2018) used the

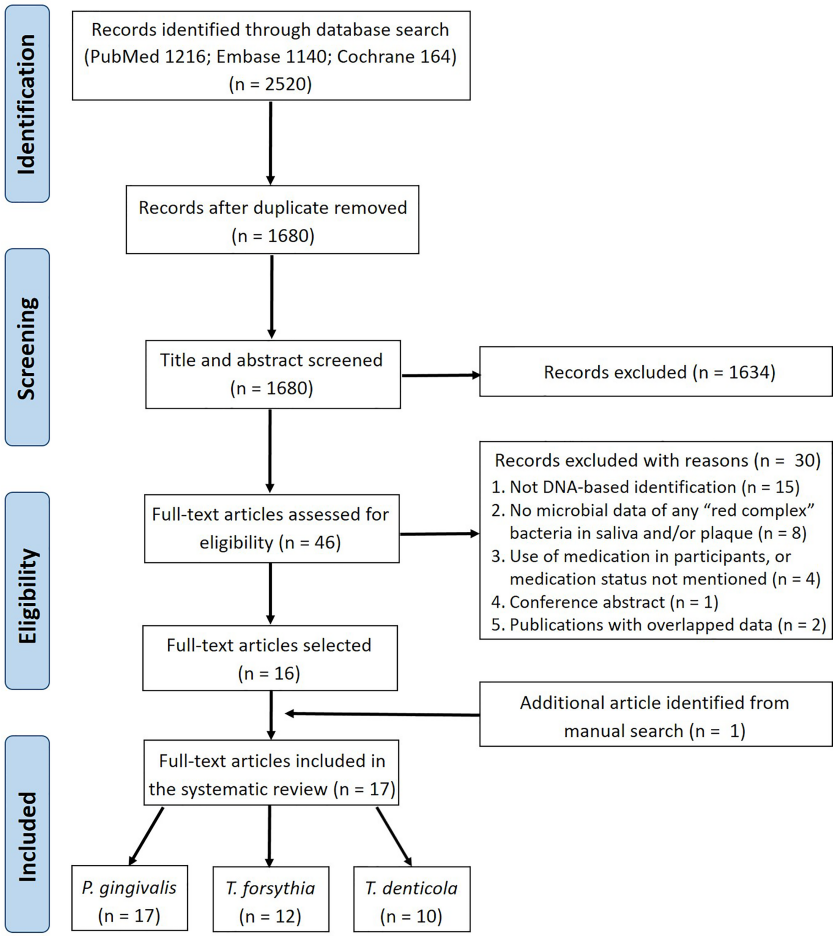


FIGURE 1 | Flow diagram of the literature search and study selection.

task force report by the American Academy of Periodontology (Geurs, 2015), 1 study (Choi et al., 2020) used a criteria modified from the case definition by the US Centers for Disease Control and Prevention and the American Academy of Periodontology (Page and Eke, 2007), and 2 studies (He et al., 2012; Nickles et al., 2017) used self-defined criteria. Three studies (Amano et al., 1999; Boutaga et al., 2007; Chen et al., 2015) did not specify the type of periodontitis and the diagnostic criteria.

For sample collection, saliva was collected as unstimulated (10 studies), stimulated (3 studies) or oral rinse sample (4 studies); subgingival plaque was collected by paper point in 11 studies and by curette in 6 studies. Most studies (n = 14) analyzed subgingival plaque samples pooled from multiple periodontal pockets (2 to full mouth). Only 2 studies analyzed plaque samples from individual pocket (Takeuchi et al., 2001; O'Brien-Simpson et al., 2017) and 1 study analyzed both pooled and individual plaque samples (Belstrøm et al., 2017).

With regard to the methods used for microbial identification, 4 studies used open-ended 16S rRNA gene sequencing techniques (Chen et al., 2015; Li et al., 2015; Belstrøm et al., 2017; Belstrøm et al., 2018), and the remaining 13 studies used targeted PCR-based approaches, including PCR, qPCR, and microarray techniques.

3.4 Clinical Data

The most frequently reported clinical data in the included studies, mean probing pocket depth (PPD), clinical attachment loss (CAL) and bleeding on probing (BOP) of full mouth and/or the sampled sites, are summarized in **Table 2**. Two studies did not report any clinical data (Amano et al., 1999; Estrela et al., 2010). Overall, there were no big variations among the reported mean PPD and/or CAL, except that 2 studies reported a mean PPD less than 3 mm (He et al., 2012; Choi et al., 2020). The mean PPD of the sampled sites generally ranged from 4 to 8 mm.

3.5 Microbial Data

3.5.1 Detection Frequency of the Red-Complex Bacteria

The detection frequency of at least one of the red-complex bacteria could be extracted from 16 out of the 17 included studies. One study (Yang et al., 2016) did not report detection frequency, but bacterial counts only. **Figure 3** presents an overview of bacterial detection

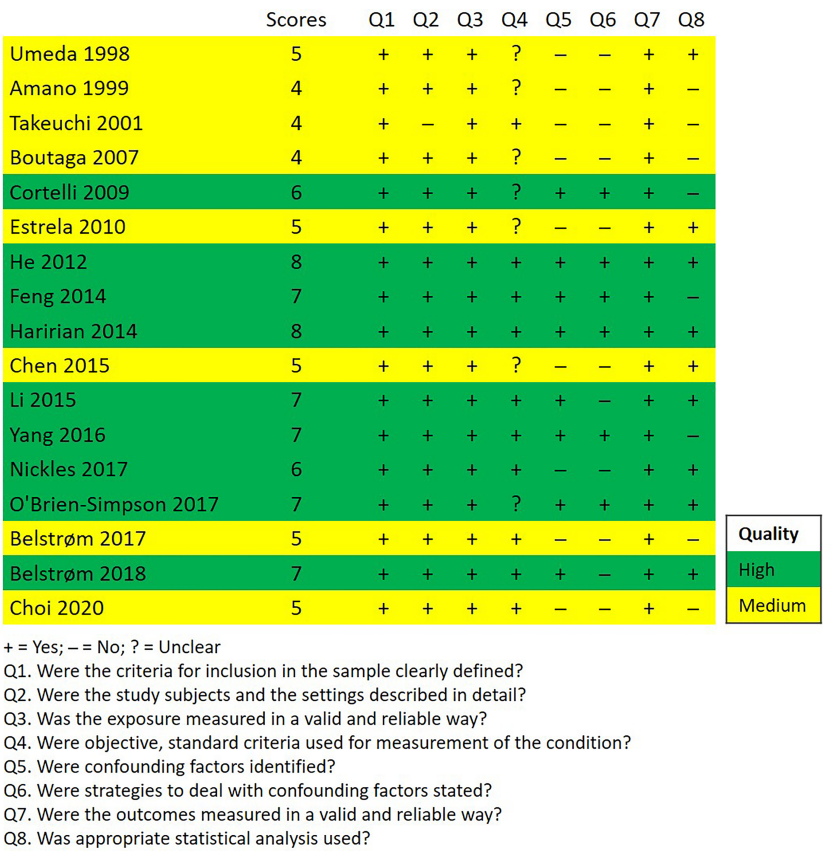


FIGURE 2 | Quality assessment of the included studies according to the JBI Critical Appraisal Checklist for Analytical Cross-Sectional Studies.

frequencies in saliva and subgingival plaque samples. Most studies reported the detection frequencies of the red-complex bacteria were more than 60% in both saliva and subgingival samples of periodontitis patients. The only study which reported detection frequencies of less than 60% for all 3 species combined the data of healthy subjects and periodontitis patients (Umeda et al., 1998). Four studies also included healthy subjects, which showed varied detection frequencies of the red-complex bacteria, ranging from 2% to 45%. However, in each study, the detection frequencies of the red-complex bacteria in healthy group were much lower than those in the corresponding periodontitis group, irrespective of the sample types (Takeuchi et al., 2001; He et al., 2012; Feng et al., 2014; O'Brien-Simpson et al., 2017). Generally, most studies reported that the detection frequency of a specific red-complex species in periodontitis patients was lower in saliva than in subgingival plaque. But a few studies reported opposite trends. These studies are Boutaga et al. (2007) and Choi et al. (2020) for *P. gingivalis*, Boutaga et al. (2007), Choi et al. (2020) and Belstrøm et al. (2017) for *T. forsythia*, and Choi et al. (2020) for *T. denticola*.

Next, meta-analyses were performed by summarizing the results from different studies, in order to assess the differences between saliva and subgingival plaque samples statistically. In total, 10 out of 16 studies were included in the meta-analyses. Six studies were excluded due to the following reasons: 1. Umeda et al. (1998) did not report data per healthy, gingivitis and periodontitis group. Only the average data of 3 groups were given; 2. O'Brien-Simpson et al. (2017) analyzed subgingival plaque samples from 6 periodontal pockets individually while other studies used pooled subgingival plaque samples; 3. Since the pocket depth was believed to affect the profile of subgingival microbiota (Van Dyke et al., 2020), studies which did not report PPD (Amano et al., 1999; Estrela et al., 2010) and reported PPD < 3 mm (He et al., 2012; Choi et al., 2020) were excluded. As shown in **Figure 4**, meta-analyses revealed that the heterogeneity among studies was low for all 3 red-complex bacteria, with I^2 ranging from 6% to 17%. The detection frequencies of *P. gingivalis* [**Figure 4A**; OR = 0.64, 95% CI: (0.43, 0.93), $p = 0.02$], *T. forsythia* [**Figure 4B**; OR = 0.53, 95% CI: (0.30, 0.95), $p = 0.03$], and *T. denticola* [**Figure 4C**; OR = 0.45, 95% CI: (0.28, 0.72), $p = 0.001$] were all significantly higher in subgingival plaque samples than that in saliva samples.

3.5.2 Bacterial Counts

Six out of the 17 studies reported bacterial counts of at least one of the red-complex species. Meta-analyses could not be performed on the data of bacteria counts due to the varied data format reported among studies (e.g., bacterial cell numbers and bacterial DNA copy numbers). Therefore, only a descriptive summary of the data is presented. As shown in **Table 3**, 4 studies (Boutaga et al., 2007;

TABLE 1 | Main characteristics of all included studies.

Author (Country, Year)	Design	Patients	Diagnostic criteria	Saliva	Subgingival plaque			Method	Red complex reported
					Type	Sample site	Sample method		
Umeda et al. (USA, 1998)	Cross-sectional	CP (130); AgP (3)	NA	Unstimulated	Pooled (4 sites)	Deepest pockets	Paper point	PCR	*Pg; Tf; Td*
Amano et al. (Japan, 1999)	Cross-sectional	P (93)	NA	Unstimulated	Pooled (2 sites)	Deepest pockets	Curette	PCR	*Pg*
Takeuchi et al. (Japan, 2001)	Cross-sectional	CP (65); AgP (38)	International classification 1999[a]	Unstimulated	Individual (4 sites)	Deepest pockets	Paper point	PCR	*Pg; Td*
Boutaga et al. (Netherlands, 2007)	Cross-sectional	P (21)	NA	Oral rinse	Pooled (4 sites)	Deepest pockets	Paper point	qPCR	*Pg; Tf*
Cavalca Cortelli et al. (Brazil, 2009)	RCT	CP (20)	NA	Unstimulated	Pooled (8 sites)	Two sites of PPD ≥ 5 mm with BOP and CAL per quadrant	Paper point	PCR	*Pg; Tf*
Estrela et al. (Brazil, 2010)	Cross-sectional	CP (30)	NA	Unstimulated	Pooled (2 sites)	NA	Paper point	Multiplex PCR	*Pg*
He et al. (China, 2012)	Cross-sectional	CP (60)	≥ 4 teeth with BOP, CAL and radiographic alveolar bone loss, and PPD ≥ 4 mm in ≥ 4 sites not on the same tooth	Unstimulated	Pooled (4 sites)	Deepest pockets	Paper point	qPCR	*Pg*
Feng et al. (China, 2014)	Cross-sectional	AgP (81)	International classification 1999[a]	Unstimulated	Pooled (4 sites)	Site of PPD ≥ 4 mm and CAL ≥ 2 mm of the first molars	Curette	PCR	*Pg; Tf; Td*
Haririan et al. (Austria, 2014)	Cross-sectional	CP (43); AgP (33)	International classification 1999[a]	Oral rinse	Pooled (4 sites)	Deepest pockets	Paper point	microarray technique	*Pg; Tf; Td*
Chen et al. (China, 2015)	Cross-sectional	P (30)	NA	Unstimulated	Pooled (4 sites)	Deepest pockets	Curette	16S rRNA gene sequencing (454 GS FLX)	*Pg; Tf*
Li et al. (China, 2015)	Cross-sectional	CP (10); AgP (10)	International classification 1999[a]	Unstimulated	Pooled (4 sites)	Site of PPD ≥ 4mm and CAL ≥ 2mm of the first molars	Curette	16S rRNA gene sequencing (454 GS FLX)	*Pg; Tf; Td*
Yang et al. (China, 2016)	Prospective	CP (45)	International classification 1999[a]	Unstimulated	Pooled	All teeth	Paper point	qPCR	*Pg; Tf; Td*
Nickles et al. (Germany, 2017)	Cross-sectional	CP (27); AgP (23)	Vertical CAL ≥ 5 mm at > 30% sites and age > 35 y (CP); clinically healthy, radiographic bone loss ≥ 50% on ≥ 2 different teeth and age < 35 y (AgP)	Oral rinse	Pooled (4 sites)	Deepest pockets	Paper point	qPCR	*Pg; Tf; Td*
O'Brien-Simpson et al. (Australia, 2017)	Cross-sectional	CP (50)	NA	Stimulated	Individual (6 sites)	Deepest pockets	Curette	qPCR	*Pg*
Belstrøm et al. (Denmark, 2017)	Cross-sectional	CP (18)	Task force report by the AAP[b]	Stimulated	Pooled; individual (3 sites)[c]	Deepest pockets	Paper point	16S rRNA gene sequencing (Illumin a MiSeq)	*Pg; Tf; Td*

(Continued)

TABLE 1 | Continued

Author (Country, Year)	Design	Patients	Diagnostic criteria	Saliva	Subgingival plaque			Method	Red complex reported
					Type	Sample site	Sample method		
Belstrøm et al. (Denmark, 2018)	Prospective	CP (24)	Task force report by the AAP[b]	Stimulated	Pooled (4 sites)	Deepest pockets	Curette	16S rRNA gene sequencing (Illumina MiSeq)	*Pg; Tf; Td*
Choi et al. (Korea, 2020)	Cross-sectional	MP (38); SP (38)[d]	Modification of CDC-AAP case definitions[d]	Oral rinse	Pooled (3 sites)	Deepest pockets	Paper point	qPCR	*Pg; Tf; Td*

NA, not available.

RCT, randomized clinical trial; CP, chronic periodontitis; AgP, aggressive periodontitis; P, periodontitis unclassified; SP, severe periodontitis; MP, moderate periodontitis; PPD, probing pocket depth; CAL, clinical attachment loss; BOP, bleeding on probing; PCR, polymerase chain reaction; qPCR, quantitative real-time PCR; Pg, Porphyromonas gingivalis; Tf, Tannerella forsythia; Td, Treponema denticola.

[a]*1999 International Classification of Periodontal Diseases and Conditions.*

[b]*American Academy of Periodontology (AAP) Task Force Report on the Update to the 1999 Classification of Periodontal Diseases and Conditions.*

[c]*Only data from pooled subgingival plaque samples were used for analysis in this systematic review, in order to be comparable with data from other studies.*

[d]*Case definition introduced by the US Centers for Disease Control and Prevention and the American Academy of Periodontology (CDC-AAP). MP, moderate periodontitis; SP, severe periodontitis.*

Yang et al., 2016; Nickles et al., 2017; O'Brien-Simpson et al., 2017) reported higher red-complex counts in subgingival plaque than in saliva, while the other 2 studies reported the opposite results (He et al., 2012; Choi et al., 2020). Only 1 study (Nickles et al., 2017) performed statistical analysis to confirm the reported higher counts in subgingival plaque.

3.5.3 Relative Abundance of the Red-Complex Bacteria

The relative abundance data were extracted from 4 sequencing studies in the following ways: from the published data (Belstrøm et al., 2017), from the data provided by the authors upon request (Li et al., 2015; Belstrøm et al., 2018), and from the raw sequence

TABLE 2 | Clinical parameters (PPD, CAL and BOP) of periodontitis patients in the included studies.

Study	Classification of Periodontitis	Full mouth			Sampled sites		
		Mean PPD (mm)	Mean CAL (mm)	Mean BOP (% sites)	Mean PPD (mm)	Mean CAL (mm)	Mean BOP (% sites)
Umeda et al., 1998	CP	–	–	–	5.10 ± 1.50	–	–
	AgP	–	–	–	5.60 ± 0.70	–	–
Amano et al., 1999	P	–	–	–	–	–	–
Takeuchi et al., 2001	CP	–	–	–	5.82 ± 2.21	6.66 ± 2.51	57.30
	AgP	–	–	–	5.84 ± 2.40	6.20 ± 2.70	77.60
Boutaga et al., 2007	P	–	–	–	6.48 ± 1.04	4.45 ± 3.68	–
Cavalca Cortelli et al., 2009	CP	4.96 ± 0.48	–	–	–	–	–
Estrela et al., 2010	CP	–	–	–	–	–	–
He et al., 2012	CP	2.70 ± 0.70	2.40 ± 1.80	41.00 ± 18.70	3.90 ± 1.30	3.90 ± 2.70	–
Feng et al., 2014	AgP	5.02 ± 1.08	4.67 ± 1.53	–	6.85 ± 1.47	6.03 ± 1.86	–
Haririan et al., 2014	CP	4.01 ± 0.93	4.54 ± 1.21	40.82 ± 23.64	7.19 ± 1.12	–	–
	AgP	3.87 ± 0.91	4.39 ± 0.95	46.22 ± 24.82	7.52 ± 1.13	–	–
Chen et al., 2015	P	4.80 ± 0.96	4.30 ± 1.43	–	–	–	–
Li et al., 2015	CP	4.50 ± 1.24	4.40 ± 1.05	100	5.47 ± 1.24	5.85 ± 2.47	–
	AgP	4.84 ± 0.91	4.28 ± 1.33	100	6.95 ± 0.74	5.92 ± 0.91	–
Yang et al., 2016	CP	3.21 ± 0.86	2.09 ± 1.32	–	–	–	–
Nickles et al., 2017	CP	–	–	–	8.61 ± 1.32	8.99 ± 1.28	–
	AgP	–	–	–	7.96 ± 1.97	8.15 ± 2.40	–
O'Brien-Simpson et al., 2017	CP	3.60 ± 1.00	7.00 ± 2.10	60.80 ± 25.30	–	–	–
Belstrøm et al., 2017	CP	–	–	–	7.00	8.00	–
Belstrøm et al., 2018	CP	3.40	4.10	56.00	6.40	7.00	–
Choi et al., 2020	MP	2.49	2.65	47.13	–	–	–
	SP	2.89	3.82	53.91	–	–	–

Data are presented as mean or mean ± SD of full mouth and/or sampled sites where the subgingival plaque samples were collected.

PPD, probing pocket depth; CAL, clinical attachment loss; BOP, bleeding on probing; CP, chronic periodontitis; AgP, aggressive periodontitis; P, periodontitis unclassified; MP, moderate periodontitis; SP, severe periodontitis.

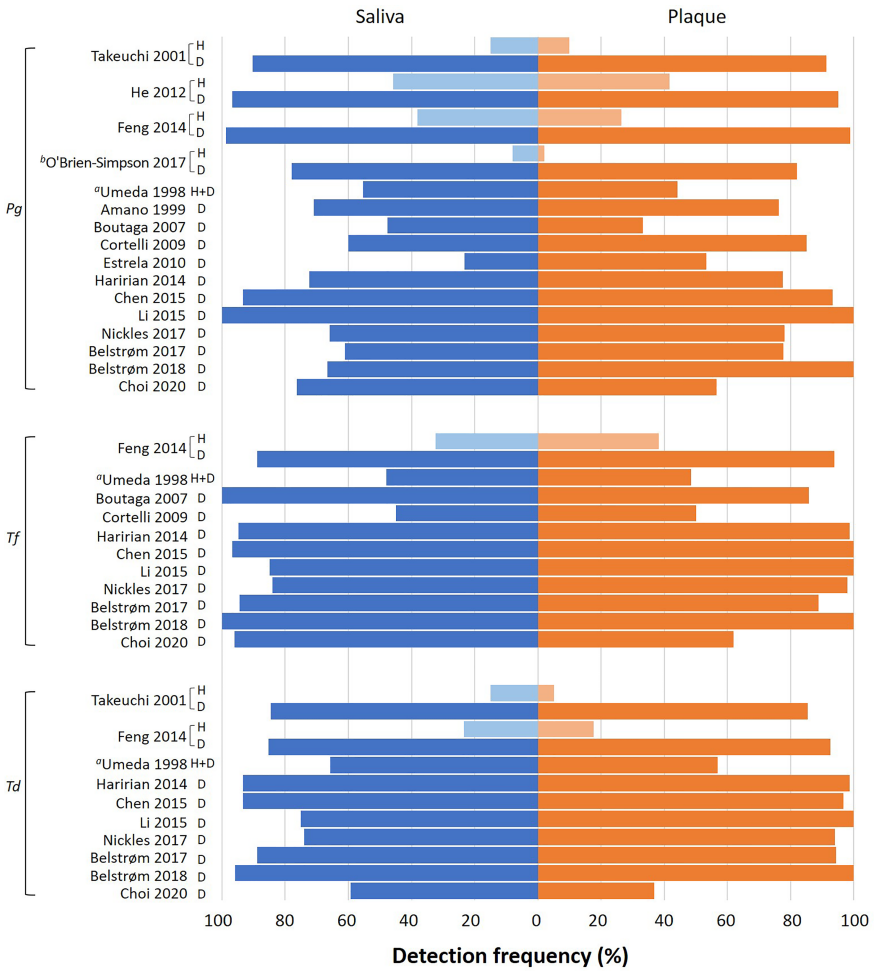


FIGURE 3 | Overview of the detection frequency of *P. gingivalis* (*Pg*), *T. forsythia* (*Tf*) and *T. denticola* (*Td*) in saliva and subgingival plaque samples reported in each study. H: healthy subjects; D: periodontitis patients. Pooled subgingival plaque samples were used for analysis unless specified otherwise. [a]This study mentioned 3 groups of subjects (health, gingivitis and periodontitis), but the data of different subject groups were reported together. [b]This study analyzed the subgingival plaque samples from 6 periodontal pockets individually.

data that have been uploaded to the Sequence Read Archive (Chen et al., 2015).

As shown in **Figure 5**, meta-analyses revealed that the relative abundances of *P. gingivalis* [**Figure 5A**; MD = -10.27, 95% CI: (-18.15, -2.38), $p < 0.00001$], *T. forsythia* [**Figure 5B**; MD = -1.85, 95% CI: (-2.57, -1.12), $p < 0.00001$] and *T. denticola* [**Figure 5C**; MD = -1.20, 95% CI: (-1.74, -0.66), $p < 0.0001$] in subgingival plaque were all significantly higher than in saliva. However, considerably high heterogeneities among 4 studies were observed for the data of all 3 bacteria. Taking *P. gingivalis* as an example, the reported mean relative abundance ranged from 2.5 to 25% in subgingival plaque and from 0.1% to 5% in saliva, with an I^2 index of 93%.

Since the relative abundance data of each patient were available from all 4 sequencing studies, we conducted Spearman's rank correlation analysis in SPSS version 25 (SPSS Inc., Chicago, IL, USA) using the paired data obtained from saliva and subgingival plaque per patient. **Figure 6** shows that positive correlations in relative abundance between saliva and subgingival plaque samples were observed for all 3 red-complex species, with a correlation coefficient of 0.73 for *P. gingivalis* ($p < 0.0001$), 0.28 for *T. forsythia* ($p = 0.006$) and 0.27 for *T. denticola* ($p = 0.01$).

3.5.4 Predominant Bacterial Genera in Saliva and Plaque Samples

The open-ended sequencing techniques allow the detection of more bacterial species than the targeted approach. Hence, the microbial compositions were summarized from 4 sequencing studies. The top 5 most abundant bacterial genera of each study are shown in **Figure 7**. Within each study, the top 5 most abundant bacterial genera in saliva were generally different from those in subgingival plaque, suggesting a major

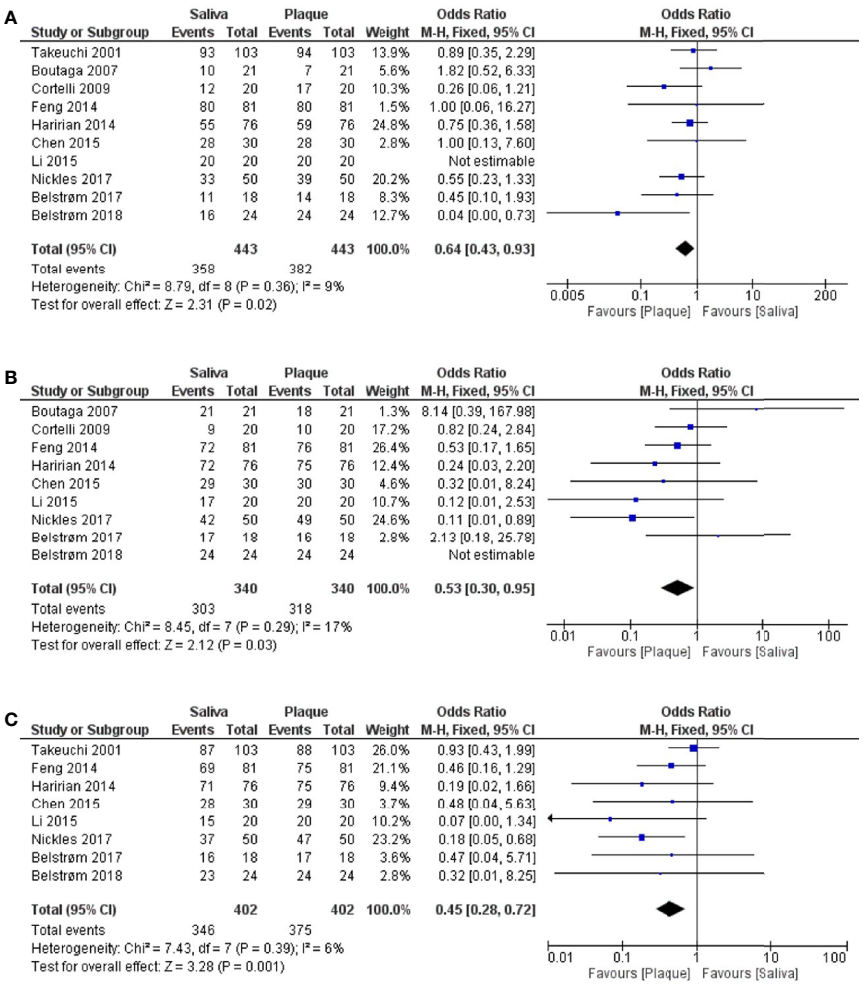


FIGURE 4 | Forest plots of meta-analyses comparing the detection frequency of: **(A)** *P. gingivalis*; **(B)** *T. forsythia*; **(C)** *T. denticola* between saliva and subgingival plaque samples from patients with periodontitis.

compositional difference between these two sample types. Among the most abundant genera, 4 bacterial genera were shared by all 4 studies: *Streptococcus* and *Prevotella* in saliva samples and *Porphyromonas* and *Fusobacterium* in subgingival plaque samples.

4 DISCUSSION

It has been widely accepted that periodontitis is an inflammatory disease caused by a dysbiotic microbial community (Hajishengallis, 2015). Representative microbial sampling is

TABLE 3 | Comparisons of bacterial counts of the red complex species between saliva and subgingival plaque samples from periodontitis patients.

Study	*P. gingivalis*		*T. forsythia*		*T. denticola*	
	saliva *vs* plaque		saliva *vs* plaque		saliva *vs* plaque	
He et al., 2012	↑		–	–	–	–
O'Brien-Simpson et al., 2017		↑	–	–	–	–
Boutaga et al., 2007		↑		↑	–	–
Yang et al., 2016		↑		↑		↑
Nickles et al., 2017		↑*		↑*		↑*
Choi et al., 2020	↑		↑		↑	

↑*Higher in the corresponding sample.*
**Data presented in the study were evaluated statistically.*

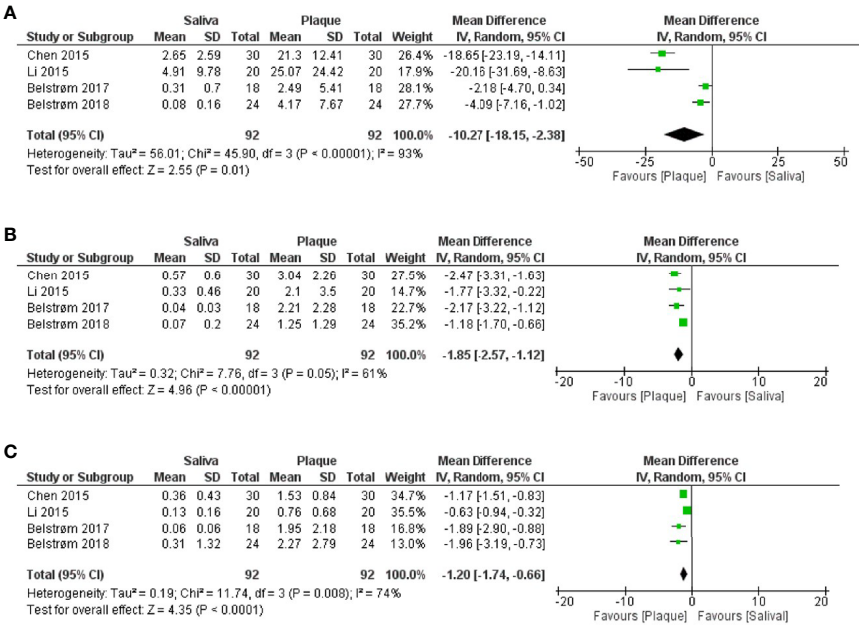


FIGURE 5 | Forest plots of meta-analyses comparing the relative abundances (%) of: **(A)** *P. gingivalis*; **(B)** *T. forsythia*; **(C)** *T. denticola* between saliva and subgingival plaque samples from patients with periodontitis, as determined in studies using 16S rRNA gene sequencing techniques.

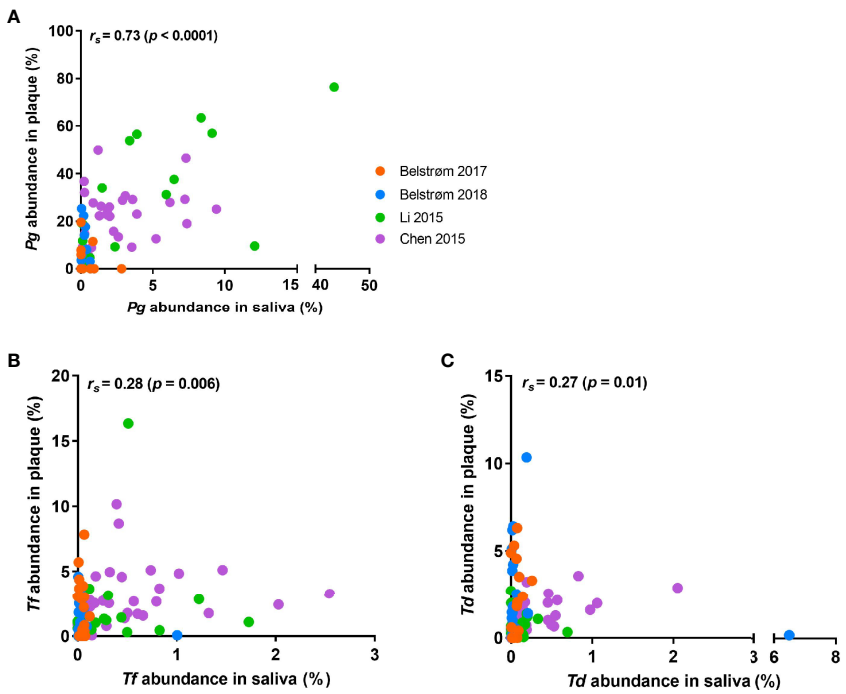


FIGURE 6 | Scatter plots showing the distributions of the relative abundance of: **(A)** *P. gingivalis*; **(B)** *T. forsythia*; **(C)** *T. denticola* in saliva and subgingival plaque samples per patient, as determined in studies using 16S rRNA gene sequencing techniques. Each dot represents one patient, and patients from different studies are indicated by different colors of the dots. Spearman's rank correlation coefficient (r_s) and p value are shown in each plot.

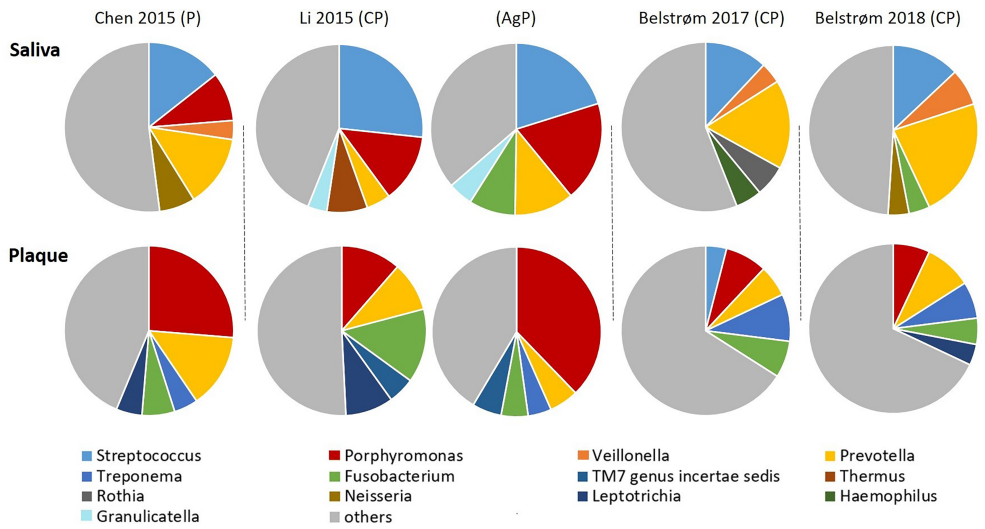


FIGURE 7 | Approximate relative abundance of the top 5 most abundant genera identified in saliva and subgingival plaque from patients with periodontitis, as determined in studies using 16S rRNA gene sequencing techniques. CP, chronic periodontitis; AgP, aggressive periodontitis; P, periodontitis unclassified.

crucial for disease prevention, diagnosis and treatment. Subgingival plaque in periodontal pockets represents the onset and development of periodontitis the best. However, due to its complicated sampling process, the use of saliva as an alternative has been an interest in many clinical studies. This systematic review identified 17 studies that reported the levels of red-complex bacteria in both saliva and subgingival plaque in periodontitis patients. Three types of outcome parameters, detection frequency, bacterial count and relative abundance, were examined. The meta-analyses on both detection frequency and relative abundance revealed that the levels of the red-complex bacteria in saliva were significantly lower than those in subgingival plaque.

Previously various researchers have claimed, based on their own data, that saliva could be a potential alternative to subgingival plaque for microbiologic analysis in periodontitis patients (Boutaga et al., 2007; Haririan et al., 2014; Belstrøm et al., 2017). Our meta-analysis summarized the results based on the samples obtained from 443 periodontitis patients in 10 studies (**Figure 4**). We found that saliva samples cannot represent the levels (i.e., detection frequency and relative abundance) of red-complex bacteria in subgingival plaque accurately in these patients. Since all the data analyzed in this review were obtained from samples taken at one time point, our finding indicates that one-time saliva sampling cannot be used to screen patients for the red-complex bacteria. We also examined other factors which might influence the comparison between saliva and subgingival plaque samples. Interestingly, the subgroup analysis (**Supplementary Material**; **Figure S1**) based on the collection methods of subgingival plaque, paper point or curette, showed that the results of studies using curette were in line with the finding mentioned above. However, in the studies using paper point, the detection frequencies of *P. gingivalis* and *T. forsythia* in saliva and subgingival plaque samples were similar. Possibly, a paper point collects unattached microbes in the periodontal pocket, which are likely spilt over to saliva; whereas a curette collects firmly attached biofilms (Jervøe-Storm et al., 2007). Clinical findings on the collection methods of subgingival plaque are inconsistent: one study (Renvert et al., 1992) stated that paper point sampling presented different microbial information as compared to curette sampling; whereas another study claimed a good agreement for the results of two sampling methods. Moreover, the open-ended sequencing method revealed DNA contamination in paper points, making this collection method unsuitable for sequencing analysis (van der Horst et al., 2013). Taken together, the methods used for collecting subgingival plaque may potentially influence microbial composition comparisons between saliva and subgingival plaque.

It is worth noting that our data analyses demonstrated positive associations between saliva and subgingival plaque in terms of red-complex levels despite limited data. Among the 17 included studies, 4 studies not only examined samples from periodontitis groups, but also from periodontal healthy groups. Although the reported detection frequencies in the healthy subjects varied considerably among studies (2% to 45%), all 4 studies showed that within one study, the detection frequency of each red-complex bacteria was much higher in periodontitis group than that in healthy group, irrespective of the sample type. Moreover, the relative abundances of the red-complex bacteria in saliva were also significantly correlated to that in subgingival plaque (**Figure 6**). Hence, it is possible that once the red-complex bacteria are enriched in subgingival plaque at diseased state, they could be spilt over or washed out into saliva, which consequently increase their levels in saliva. To this end, sampling saliva at multiple time points might help to trace the compositional shift of subgingival microbiota towards a disease provoking state. The study of Belstrøm et al. (2018) showed the correlation of the

levels of salivary red-complex bacteria to that of subgingival plaque before and after treatment, indicating the possibility of such application of saliva sample. However, a thorough review on longitudinal clinical studies is needed to confirm this.

The magnitude and presence of statistical heterogeneity in a meta-analysis are usually explained by heterogeneity of methodological and/or clinical sources (Deeks et al., 2021). In our results, high heterogeneities of the relative abundance data among the 4 sequencing studies were observed, irrespective of the targeted bacterial species (I^2: 61%–93%). Since the sequencing method-related heterogeneity in the meta-analysis of microbiome data has been reported before (Lozupone et al., 2013; Duvallet, 2018), the source of high heterogeneity here is likely related to the sequencing techniques. All 4 studies varied in many steps in sample processing as well as analysis, such as the DNA extraction methods, targeted 16S rRNA gene region for sequencing, sequencing platform used and sequencing depth. Lozupone et al. (2013) conducted meta-analysis on human microbiome data extracted from 12 different studies, and revealed that the technical variations in different studies could obscure biologically meaningful compositional differences. However, when the studied parameter had a large effect size (e.g., the body sites), the bias caused by variation in sequencing methodology could be outweighed by the real difference. In our case, despite the high heterogeneity, the relative abundance of red-complex bacteria in saliva was consistently lower than that in subgingival plaque.

During data analysis, we identified a clinical parameter, periodontal pocket depth, which potentially contributed to the high heterogeneity in a meta-analysis. In addition to the results on the basis of 10 studies reported in **Figure 4**, we also performed a meta-analysis on the detection frequencies reported in 14 studies, where an additional 4 studies with unknown or low periodontal pocket depth were included (**Supplementary Material**; **Figure S2**). From these meta-analyses, we observed substantially high heterogeneities (I^2: 49%–73%) as compared to those in the 10-study based meta-analyses (I^2: 6%–17%). The 14-study based meta-analyses showed no significant difference in the detection frequency of all 3 red-complex bacteria between saliva and subgingival plaque. We suspected that the high heterogeneity of the 14-study based meta-analyses was, at least partly, caused by variations in pocket depths. Van Dyke et al. (2020) stated that the depth of periodontal pocket was one of the crucial elements which determined the dysbiosis of subgingival microbiota, as differentiate microbial profiles were observed in pockets with different depths. For example, the abundance of *Bacteriodetes* significantly increased as the pocket deepened (Kirst et al., 2015). Our observation showed that the periodontal pocket depth might be an important confounding factor for microbial analysis. Interestingly, in the summary of bacterial count data, the 2 studies that reported a low pocket depth were also the only 2 studies which showed higher salivary red-complex bacterial counts in saliva as compared to subgingival plaque (He et al., 2012; Choi et al., 2020). Likely, the differential levels of red-complex bacteria in saliva and subgingival plaque are associated with the depth of periodontal pockets. Unfortunately, most studies included this review obtained samples from deep pockets, and we could not further illustrate the potential influence of this confounding factor.

In conclusion, this systematic review shows that the levels of red-complex bacteria in saliva were significantly lower than those in subgingival plaque in patients with periodontitis, in terms of the detection frequency and relative abundance. This finding is based on the meta-analyses on the data obtained from 443 patients at one sampling time point. In addition, our analyses reveal positive associations in the levels of red-complex bacteria between saliva and subgingival plaque despite limited data. Therefore, we recommend a thorough review of longitudinal clinical studies to further assess the role of saliva in detecting periodontitis-related microorganisms.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

YJ and DD conceived and designed the study. YJ and BS performed the study and collected the data. YJ, BB, and DD analyzed and interpreted the data. YJ drafted the manuscript. BB, LC, XZ, RE, WC, and DD revised the manuscript critically. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the ACTA Research Institute and the National Natural Science Foundation of China grant 81870759 (to LC).

ACKNOWLEDGMENTS

We would like to thank dr. Daniel Belstrøm and dr. Xianghui Feng who kindly provided us the requested data in their studies, and dr. Zisheng Tang who kindly responded to our email but was not able to provide the requested data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2021.727732/full#supplementary-material

REFERENCES

Amano, A., Nakagawa, I., Kataoka, K., Morisaki, I., and Hamada, S. (1999). Distribution of Porphyromonas Gingivalis Strains With fimA Genotypes in Periodontitis Patients. *J. Clin. Microbiol.* 37 (5), 1426–1430. doi: 10.1128/JCM.37.5.1426-1430.1999

Armitage, G. C. (1999). Development of a Classification System for Periodontal Diseases and Conditions. *Ann. Periodontol.* 4 (1), 1–6. doi: 10.1902/annals.1999.4.1.1

Belstrøm, D., Grande, M. A., Sembler-Møller, M. L., Kirkby, N., Cotton, S. L., Paster, B. J., et al. (2018). Influence of Periodontal Treatment on Subgingival and Salivary Microbiotas. *J. Periodontol.* 89 (5), 531–539. doi: 10.1002/JPER.17-0377

Belstrøm, D., Sembler-Møller, M. L., Grande, M. A., Kirkby, N., Cotton, S. L., Paster, B. J., et al. (2017). Microbial Profile Comparisons of Saliva, Pooled and Site-Specific Subgingival Samples in Periodontitis Patients. *PloS One* 12 (8), e0182992–e0182992. doi: 10.1371/journal.pone.0182992

Boutaga, K., Savelkoul, P. H. M., Winkel, E. G., and van Winkelhoff, A. J. (2007). Comparison of Subgingival Bacterial Sampling With Oral Lavage for Detection and Quantification of Periodontal Pathogens by Real-Time Polymerase Chain Reaction. *J. Periodontol.* 78 (1), 79–86. doi: 10.1902/jop.2007.060078

Boutaga, K., van Winkelhoff, A. J., Vandenbroucke-Grauls, C. M. J. E., and Savelkoul, P. H. M. (2003). Comparison of Real-Time PCR and Culture for Detection of Porphyromonas Gingivalis in Subgingival Plaque Samples. *J. Clin. Microbiol.* 41 (11), 4950–4954. doi: 10.1128/jcm.41.11.4950-4954.2003

Chen, H., Liu, Y., Zhang, M., Wang, G., Qi, Z., Bridgewater, L., et al. (2015). A Filifactor Alocis-Centered Co-Occurrence Group Associates With Periodontitis Across Different Oral Habitats. *Sci. Rep.* 5, 9053–9053. doi: 10.1038/srep09053

Cavalca Cortelli, S., Cavallini, F., Regueira Alves, M. F., et al. (2009). Clinical and Microbiological Effects of an Essential-Oil-Containing Mouth Rinse Applied in the "One-Stage Full-Mouth Disinfection" Protocol—A Randomized Doubled-Blinded Preliminary Study. *Clin. Oral. Invest.* 13 (2), 189–94. doi: 10.1007/s00784-008-0219-3

Choi, J. U., Lee, J. B., Kim, K. H., Kim, S., Seol, Y. J., Lee, Y. M., et al. (2020). Comparison of Periodontopathic Bacterial Profiles of Different Periodontal Disease Severity Using Multiplex Real-Time Polymerase Chain Reaction. *Diagn. (Basel Switzerland)* 10 (11), 965. doi: 10.3390/diagnostics10110965

Curtis, M. A., Diaz, P. I., and Van Dyke, T. E. (2020). The Role of the Microbiota in Periodontal Disease. *Periodontol. 2000* 83 (1), 14–25. doi: 10.1111/prd.12296

Deeks, J. J., Higgins, J. P. T., and Altman, D. G. (2021). "Chapter 10: Analysing Data and Undertaking Meta-Analyses," in *Cochrane Handbook for Systematic Reviews of Interventions Version 6-2 (Updated February 2021)*. Eds. J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page and V. A. Welch (Cochrane). Available at: www.training.cochrane.org/handbook.

Duvallet, C. (2018). Meta-Analysis Generates and Prioritizes Hypotheses for Translational Microbiome Research. *Microb. Biotechnol.* 11 (2), 273–276. doi: 10.1111/1751-7915.13047

Estrela, C., Pimenta, F. C., Alencar, A. H. G. D., Ruiz, L. F. N., and Estrela, C. (2010). Detection of Selected Bacterial Species in Intraoral Sites of Patients With Chronic Periodontitis Using Multiplex Polymerase Chain Reaction. *J. Appl. Oral. Sci.* 18 (4), 426–431. doi: 10.1590/s1678-77572010000400018

Feng, X., Zhang, L., Xu, L., Meng, H., Lu, R., Chen, Z., et al. (2014). Detection of Eight Periodontal Microorganisms and Distribution of Porphyromonas Gingivalis fimA Genotypes in Chinese Patients With Aggressive Periodontitis. *J. Periodontol.* 85 (1), 150–159. doi: 10.1902/jop.2013.120677

Geurs, N. (2015). American Academy of Periodontology Task Force Report on the Update to the 1999 Classification of Periodontal Diseases and Conditions. *J. Periodontol.* 86 (7), 835–838. doi: 10.1902/jop.2015.157001

Hajishengallis, G. (2015). Periodontitis: From Microbial Immune Subversion to Systemic Inflammation. *Nat. Rev. Immunol.* 15 (1), 30–44. doi: 10.1038/nri3785

Haririan, H., Andrukhov, O., Bertl, K., Lettner, S., Kierstein, S., Moritz, A., et al. (2014). Microbial Analysis of Subgingival Plaque Samples Compared to That of Whole Saliva in Patients With Periodontitis. *J. Periodontol.* 85 (6), 819–828. doi: 10.1902/jop.2013.130306

He, J., Huang, W., Pan, Z., Cui, H., Qi, G., Zhou, X., et al. (2012). Quantitative Analysis of Microbiota in Saliva, Supragingival, and Subgingival Plaque of Chinese Adults With Chronic Periodontitis. *Clin. Oral. Investig.* 16 (6), 1579–1588. doi: 10.1007/s00784-011-0654-4

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., and Altman, D. G. (2003). Measuring Inconsistency in Meta-Analyses. *BMJ (Clin. Res. ed.)* 327 (7414), 557–560. doi: 10.1136/bmj.327.7414.557

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., et al. (2012). Structure, Function and Diversity of the Healthy Human Microbiome. *Nature* 486 (7402), 207–214. doi: 10.1038/nature11234

Jervøe-Storm, P.-M., AlAhdab, H., Koltzscher, M., Fimmers, R., and Jepsen, S. (2007). Comparison of Curet and Paper Point Sampling of Subgingival Bacteria as Analyzed by Real-Time Polymerase Chain Reaction. *J. Periodontol.* 78 (5), 909–917. doi: 10.1902/jop.2007.060218

Kassebaum, N. J., Bernabé, E., Dahiya, M., Bhandari, B., Murray, C. J. L., and Marcenes, W. (2014). Global Burden of Severe Periodontitis in 1990-2010: A Systematic Review and Meta-Regression. *J. Dent. Res.* 93 (11), 1045–1053. doi: 10.1177/0022034514552491

Kirst, M. E., Li, E. C., Alfant, B., Chi, Y.-Y., Walker, C., Magnusson, I., et al. (2015). Dysbiosis and Alterations in Predicted Functions of the Subgingival Microbiome in Chronic Periodontitis. *Appl. Environ. Microbiol.* 81 (2), 783–793. doi: 10.1128/AEM.02712-14

Li, Y., Feng, X., Xu, L., Zhang, L., Lu, R., Shi, D., et al. (2015). Oral Microbiome in Chinese Patients With Aggressive Periodontitis and Their Family Members. *J. Clin. Periodontol.* 42 (11), 1015–1023. doi: 10.1111/jcpe.12463

Loesche, W. J. (1992). DNA Probe and Enzyme Analysis in Periodontal Diagnostics. *J. Periodontol.* 63 (12S), 1102–1109. doi: 10.1902/jop.1992.63.12s.1102

Lozupone, C. A., Stombaugh, J., Gonzalez, A., Ackermann, G., Wendel, D., Vázquez-Baeza, Y., et al. (2013). Meta-Analyses of Studies of the Human Microbiota. *Genome Res.* 23 (10), 1704–1714. doi: 10.1101/gr.151803.112

Mark Welch, J. L., Dewhirst, F. E., and Borisy, G. G. (2019). Biogeography of the Oral Microbiome: The Site-Specialist Hypothesis. *Annu. Rev. Microbiol.* 73, 335–358. doi: 10.1146/annurev-micro-090817-062503

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G.Group, P (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PloS Med.* 6 (7), e1000097–e1000097. doi: 10.1371/journal.pmed.1000097

Moola, S., Munn, Z., Tufanaru, C., Aromataris, E., Sears, K., Sfetcu, R., et al. (2020). "Chapter 7: Systematic Reviews of Etiology and Risk," in *JBI Manual for Evidence Synthesis*. Eds. E. Aromataris and Z. Munn (JBI).

Nickles, K., Scharf, S., Röllke, L., Dannewitz, B., and Eickholz, P. (2017). Comparison of Two Different Sampling Methods for Subgingival Plaque: Subgingival Paper Points or Mouthrinse Sample? *J. Periodontol.* 88 (4), 399–406. doi: 10.1902/jop.2016.160249

O'Brien-Simpson, N. M., Burgess, K., Lenzo, J. C., Brammar, G. C., Darby, I. B., and Reynolds, E. C. (2017). Rapid Chair-Side Test for Detection of Porphyromonas Gingivalis. *J. Dent. Res.* 96 (6), 618–625. doi: 10.1177/0022034517691720

Ouzzani, M., Hammady, H., Fedorowicz, Z., and Elmagarmid, A. (2016). Rayyan—A Web and Mobile App for Systematic Reviews. *Syst. Rev.* 5 (1), 210. doi: 10.1186/s13643-016-0384-4

Page, R. C., and Eke, P. I. (2007). Case Definitions for Use in Population-Based Surveillance of Periodontitis. *J. Periodontol.* 78 (7S), 1387–1399. doi: 10.1902/jop.2007.060264

Renvert, S., Wikström, M., Helmersson, M., Dahlén, G., and Claffey, N. (1992). Comparative Study of Subgingival Microbiological Sampling Techniques. *J. Periodontol.* 63 (10), 797–801. doi: 10.1902/jop.1992.63.10.797

Segata, N., Haake, S. K., Mannon, P., Lemon, K. P., Waldron, L., Gevers, D., et al. (2012). Composition of the Adult Digestive Tract Bacterial Microbiome Based on Seven Mouth Surfaces, Tonsils, Throat and Stool Samples. *Genome Biol.* 13 (6), R42. doi: 10.1186/gb-2012-13-6-r42

Simón-Soro, Á., Tomás, I., Cabrera-Rubio, R., Catalan, M. D., Nyvad, B., and Mira, A. (2013). Microbial Geography of the Oral Cavity. *J. Dent. Res.* 92 (7), 616–621. doi: 10.1177/0022034513488119

Socransky, S. S., Haffajee, A. D., Cugini, M. A., Smith, C., and Kent, J. R. L. (1998). Microbial Complexes in Subgingival Plaque. *J. Clin. Periodontol.* 25 (2), 134–144. doi: 10.1111/j.1600-051X.1998.tb02419.x

Suchett-Kaye, G., Morrier, J.-J., and Barsotti, O. (2001). Clinical Usefulness of Microbiological Diagnostic Tools in the Management of Periodontal Disease. *Res. Microbiol.* 152 (7), 631–639. doi: 10.1016/S0923-2508(01)01242-6

Takeuchi, Y., Umeda, M., Sakamoto, M., Benno, Y., Huang, Y., and Ishikawa, I. (2001). Treponema Socranskii, Treponema Denticola, and Porphyromonas Gingivalis Are Associated With Severity of Periodontal Tissue Destruction. *J. Periodontol.* 72 (10), 1354–1363. doi: 10.1902/jop.2001.72.10.1354

Tonetti, M. S., Jepsen, S., Jin, L., and Otomo-Corgel, J. (2017). Impact of the Global Burden of Periodontal Diseases on Health, Nutrition and Wellbeing of Mankind: A Call for Global Action. *J. Clin. Periodontol.* 44 (5), 456–462. doi: 10.1111/jcpe.12732

Umeda, M., Contreras, A., Chen, C., Bakker, I., and Slots, J. (1998). The Utility of Whole Saliva to Detect the Oral Presence of Periodontopathic Bacteria. *J. Periodontol.* 69 (7), 828–833. doi: 10.1902/jop.1998.69.7.828

van der Horst, J., Buijs, M. J., Laine, M. L., Wismeijer, D., Loos, B. G., Crielaard, W., et al. (2013). Sterile Paper Points as a Bacterial DNA-Contamination Source in Microbiome Profiles of Clinical Samples. *J. Dent.* 41 (12), 1297–1301. doi: 10.1016/j.jdent.2013.10.008

Van Dyke, T. E., Bartold, P. M., and Reynolds, E. C. (2020). The Nexus Between Periodontal Inflammation and Dysbiosis. *Front. Immunol.* 11, 511. doi: 10.3389/fimmu.2020.00511

Vieira Colombo, A. P., Magalhães, C. B., Hartenbach, F. A. R. R., Martins do Souto, R., and Maciel da Silva-Boghossian, C. (2016). Periodontal-Disease-Associated Biofilm: A Reservoir for Pathogens of Medical Importance. *Microb. Pathog.* 94, 27–34. doi: 10.1016/j.micpath.2015.09.009

Yang, X., Li, C., and Pan, Y. (2016). The Influences of Periodontal Status and Periodontal Pathogen Quantity on Salivary 8-Hydroxydeoxyguanosine and Interleukin-17 Levels. *J. Periodontol.* 87 (5), 591–600. doi: 10.1902/jop.2015.150390

Yazdanian, M., Armoon, B., Noroozi, A., Mohammadi, R., Bayat, A.-H., Ahounbar, E., et al. (2020). Dental Caries and Periodontal Disease Among People Who Use Drugs: A Systematic Review and Meta-Analysis. *BMC Oral. Health* 20 (1), 44. doi: 10.1186/s12903-020-1010-3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Cellular and Infection Microbiology

ORIGINAL RESEARCH
published: 12 October 2021
doi: 10.3389/fcimb.2021.673465

Analysis of Salivary Mycobiome in a Cohort of Oral Squamous Cell Carcinoma Patients From Sudan Identifies Higher Salivary Carriage of *Malassezia* as an Independent and Favorable Predictor of Overall Survival

*Nazar Mohamed[1,2], Jorunn Litlekalsøy[1], Israa Abdulrahman Ahmed[1,3], Einar Marius Hjellestad Martinsen[4], Jessica Furriol[5], Ruben Javier-Lopez[6], Mariam Elsheikh[2,7], Nuha Mohamed Gaafar[1,2], Luis Morgado[8], Sunil Mundra[8,9], Anne Christine Johannessen[1,10], Tarig Al-Hadi Osman[1], Elisabeth Sivy Nginamau[1,10], Ahmed Suleiman[2,7] and Daniela Elena Costea[1,10]**

[1] Gade Laboratory for Pathology, Department of Clinical Medicine, and Center for Cancer Biomarkers CCBIO, University of Bergen, Bergen, Norway, [2] Department of Oral and Maxillofacial Surgery/Department of Basic Sciences, University of Khartoum, Khartoum, Sudan, [3] Department of Operative Dentistry, University of Science & Technology, Omdurman, Sudan, [4] Department of Clinical Science, University of Bergen, Bergen, Norway, [5] Department of Nephrology, Haukeland University Hospital, Bergen, Norway, [6] Department of Biological Sciences, The Faculty of Mathematics and Natural Sciences, University of Bergen, Bergen, Norway, [7] Department of Oral & Maxillofacial Surgery, Khartoum Dental Teaching Hospital, Khartoum, Sudan, [8] Section for Genetics and Evolutionary Biology (EvoGene), Department of Biosciences, The Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway, [9] Department of Biology, College of Science, United Arab Emirates University, Al Ain, Abu Dhabi, United Arab Emirates, [10] Department of Pathology, Laboratory Clinic, Haukeland University Hospital, Bergen, Norway

Background: Microbial dysbiosis and microbiome-induced inflammation have emerged as important factors in oral squamous cell carcinoma (OSCC) tumorigenesis during the last two decades. However, the "rare biosphere" of the oral microbiome, including fungi, has been sparsely investigated. This study aimed to characterize the salivary mycobiome in a prospective Sudanese cohort of OSCC patients and to explore patterns of diversities associated with overall survival (OS).

Materials and Methods: Unstimulated saliva samples (n = 72) were collected from patients diagnosed with OSCC (n = 59) and from non-OSCC control volunteers (n = 13). DNA was extracted using a combined enzymatic–mechanical extraction protocol. The salivary mycobiome was assessed using a next-generation sequencing (NGS)-based methodology by amplifying the ITS2 region. The impact of the abundance of different fungal genera on the survival of OSCC patients was analyzed using Kaplan–Meier and Cox regression survival analyses (SPPS).

Results: Sixteen genera were identified exclusively in the saliva of OSCC patients. *Candida*, *Malassezia*, *Saccharomyces*, *Aspergillus*, and *Cyberlindnera* were the most

OPEN ACCESS

Edited by:
Thuy Do,
University of Leeds, United Kingdom

Reviewed by:
Hubert Low,
Chris O'Brien Lifehouse, Australia
Maha Abdelkawy,
Beni-Suef University, Egypt

***Correspondence:**
Daniela Elena Costea
daniela.costea@uib.no

Specialty section:
This article was submitted to Microbiome in Health and Disease, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: 27 February 2021
Accepted: 27 August 2021
Published: 12 October 2021

Citation:
Mohamed N, Litlekalsøy J, Ahmed IA, Martinsen EMH, Furriol J, Javier-Lopez R, Elsheikh M, Gaafar NM, Morgado L, Mundra S, Johannessen AC, Osman TA-H, Nginamau ES, Suleiman A and Costea DE (2021) Analysis of Salivary Mycobiome in a Cohort of Oral Squamous Cell Carcinoma Patients From Sudan Identifies Higher Salivary Carriage of Malassezia as an Independent and Favorable Predictor of Overall Survival. Front. Cell. Infect. Microbiol. 11:673465. doi: 10.3389/fcimb.2021.673465

relatively abundant fungal genera in both groups and showed higher abundance in OSCC patients. Kaplan–Meier survival analysis showed higher salivary carriage of the *Candida* genus significantly associated with poor OS of OSCC patients (Breslow test: $p = 0.043$). In contrast, the higher salivary carriage of *Malassezia* showed a significant association with favorable OS in OSCC patients (Breslow test: $p = 0.039$). The Cox proportional hazards multiple regression model was applied to adjust the salivary carriage of both *Candida* and *Malassezia* according to age ($p = 0.029$) and identified the genus *Malassezia* as an independent predictor of OS (hazard ratio = 0.383, 95% CI = 0.16–0.93, $p = 0.03$).

Conclusion: The fungal compositional patterns in saliva from OSCC patients were different from those of individuals without OSCC. The fungal genus *Malassezia* was identified as a putative prognostic biomarker and therapeutic target for OSCC.

Keywords: oral squamous cell carcinoma (OSCC), mycobiome, *toombak*, biomarker, overall survival (OS), *Malassezia*, *Candida*

INTRODUCTION

The oral cavity is a habitat for a diverse and fluctuating collection of microorganisms (Aas et al., 2005; Nasidze et al., 2009; Yang et al., 2016). The oral microbiome, which includes, in addition to complex bacterial communities, oral fungi, viruses, and phages (Baker et al., 2017), is one of the most diverse microbial communities in the human body (Dewhirst et al., 2010; Huttenhower et al., 2012), and this is related to its multiple ecosystems (Arweiler et al., 2016). The oral microbiota represents a critical component of health and diseases (Jenkinson and Lamont, 2005; Avila et al., 2009), and balance is maintained by a continuous interplay with the host (Vasquez et al., 2018). Dysbiosis of the oral microbiome has been proposed as a marker, initiator, or modifier of oral diseases (Ghannoum et al., 2010; Hooks and O'Malley, 2017; Iliev and Leonardi, 2017; Rosier et al., 2018).

Recent advances in microbial detection techniques allowed the transition from culture-dependent studies of a single species to complex *in vitro* multispecies community detection and characterization studies (Baker et al., 2017). Large next-generation sequencing (NGS)-based projects, such as the Human Microbiome (Huttenhower et al., 2012), the Integrative Human Microbiome Project with a focus on the mechanisms of host–microbiome interactions (Proctor et al., 2019), and the Human Oral Microbiome Database (Proctor et al., 2019), give deeper insights into the human microbiome. Despite advances in the understanding of the microbiome, majority of the studies have focused on the bacterial part of the microbiome. Little is known about the fungal part of the human microbiome, recently defined as the mycobiome (Ghannoum et al., 2010; Cui et al., 2013; Chandra et al., 2016).

The few existing studies have revealed that the diversity of the oral mycobiota is lower when compared to that of the oral bacteriome (Iliev and Leonardi, 2017), and it is dominated by members of the phylum Ascomycota, mainly *Candida* spp., with *Candida albicans* as the dominant species. The other commonly identified fungi in the oral mycobiome are *Cladosporium*, *Aureobasidium*, *Saccharomycetales*, *Aspergillus*, *Fusarium*, *Cryptococcus*, and *Malassezia* (Ghannoum et al., 2010; Dupuy et al., 2014).

Evidence is accumulating on the role of fungi in neoplasia (Rindum et al., 1994; McCullough et al., 2002; Barrett et al., 2008; Hebbar et al., 2013; Berkovits et al., 2016; Zhu et al., 2017; Conche and Greten, 2018; Al-Hebshi et al., 2019; Aykut et al., 2019). Some earlier studies have suggested a possible role of *Candida* in the initiation of carcinogenesis (Field et al., 1989; Krogh, 1990). *Candida* may have a causal role in oral precancer and cancer, albeit an indirect one, implying that *Candida*, along with other cofactors, e.g., tobacco consumption, is involved in the initiation and promotion of carcinogenesis (Bakri et al., 2010; Sanjaya et al., 2011). Some *C. albicans* strains may contribute to oral carcinogenesis by producing endogenous nitrosamine (Krogh et al., 1987). An immune-mediated role in the acceleration of pancreatic ductal adenocarcinoma has also been suggested recently for another genus, namely, *Malassezia* (Aykut et al., 2019).

There are only sparse reports in the literature on the mycobiome in oral squamous cell carcinoma (OSCC). Perera et al. revealed a dysbiotic mycobiome characterized by lower species diversity and increased relative abundance of *C. albicans* in tissue biopsies of OSCC in a cohort of patients from Sri Lanka (Perera et al., 2017). Berkovits et al. (2016) used cultivation techniques coupled with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and identified a more diverse mycobiome associated with OSCC, mainly consisting of *Candida* species in addition to *Rhodotorula*, *Saccharomyces*, and *Kloeckera*.

The oral microbiota is dynamic and responsive to environmental and biological changes, so discoverable shifts in its composition and/or function might offer new biomarkers useful for the diagnosis of oral cancers (OCs) and oropharyngeal cancers (OPCs) (Aas et al., 2005). While host biomarkers are subject to individual biological variations (Lim et al., 2017), there are indications that the core oral microbiome is consistently conserved among unrelated subjects (Lim et al., 2017; Shaw

et al., 2017). The incorporation of the oral microbiome panel in other tumor biomarkers may therefore help reduce human biological variations, which prevented, so far, the utilization of molecular diagnosis and stratification in OCs and OPCs (Zaura et al., 2014; Lim et al., 2017). Moreover, salivary diagnostics is a rapidly developing field, and combined with biomarker identification and validation, it may provide a platform for the development of a noninvasive, salivary-based tool for the stratification of OSCC patients and for individualized treatments.

This study aimed to investigate the salivary mycobiome in a cohort of OSCC patients and in non-OSCC controls from Sudan and its possible impact on clinical variables, including overall survival (OS). We employed the NGS methodology to explore fungal diversities and communities in saliva and describe the salivary fungal compositional patterns in OSCC patients compared to individuals without OSCC. The fungal genus *Malassezia* was identified as an independent prognostic biomarker for OS of OSCC patients.

MATERIALS AND METHODS

Ethical Considerations

This is a prospective study involving OSCC patients (n = 59) and healthy non-cancer controls (n = 13) recruited between 2012 and 2015 at Khartoum Dental Teaching Hospital, Sudan. The National Health Research Ethics Committee, Federal Ministry of Health, Sudan, approved the research in Sudan (fmoh/rd/SEC/09). Written informed consent was obtained from both patients and controls. The Regional Ethical Committee in Norway approved the project (REKVest 3.2006.2620 REKVest 3.2006.1341).

Study Participants

The inclusion criteria were as follows: age older than 18 years, with histologically confirmed primary OSCC, did not receive any previous surgical and chemo- or radiotherapy, and consented to participate in the study. Critically ill patients, patients under medication, and those positive for human immunodeficiency virus (HIV) and hepatitis B surface antigen (HBs Ag) were excluded from the study. Human papilloma virus (HPV)-positive cases were also excluded from the study. Detailed clinical information (age, gender, tobacco habits, and alcohol use) was obtained through interviews. A routine dental examination was performed on participating individuals, which included registration of the periodontal status, plaque, gingival index, community periodontal index (CPI), simplified oral hygiene, fillings and missing teeth, and carious teeth by a team of trained and calibrated dentists specifically for this project. Non-cancer controls were included after informed consent and consecutively recruited from patients attending the outpatient clinic for trauma and benign conditions. The tumor localization, tumor size, TNM stage, comorbid conditions, last date of follow-up, and survival data were obtained from patients' hospital records. TNM stage was noted according to the guidelines of the American Joint Committee on Cancer, version 7.0. Information on current smoking habits and history of smoking was reported in pack-years (PY) (Masters, 2018), with calculations for consumption of the smokeless tobacco "*toombak*" adjusted according to the average of manually prepared portions in Sudan (Idris et al., 1995).

Saliva Sample Collection

Unstimulated saliva samples were collected. Briefly, the donor was asked not to eat and not to use oral hygiene products 1 h before saliva collection. At least 2 ml of unstimulated saliva was collected on ice and then kept in a portable liquid nitrogen container until further storage at −80°C at the end of the collection day. The sample collection time did not exceed 20 min.

Fungal DNA Extraction and Control Sample Setting

The recommendation for standardized DNA extraction for microbiome studies was followed (Leigh Greathouse et al., 2019). A combined enzymatic–mechanical extraction method was chosen and modified, when needed, for fungi (Huseyin et al., 2017; Rosenbaum et al., 2019). Of the saliva, 300 µl was used for DNA extraction. Sputasol® (300 µl, Oxoid Ltd., Basingstoke, UK) was added and incubated, with shaking, at 37°C for 15 min. Following centrifugation, pellets were reconstituted in 250 µl of phosphate-buffered saline (PBS). For enzymatic digestion, an enzyme cocktail of lysostaphin (4,000 U/ml), mutanolysin (25,000 U/ml), and lysozyme (10 mg/ml) was diluted in TE5 buffer (10 mM Tris-HCl and 5 mM EDTA, pH 8.0) (all from Sigma-Aldrich, Saint-Louis, MO, USA). Fifty microliters of the enzyme cocktail was added to each reconstituted pellet and mixed well, then incubated at 37°C with slight shaking at 350 rpm for 1 h. The FastDNA™ Kit (MP Biomedicals, Irvine, CA, USA) was used after enzymatic digestion. The samples were centrifuged and the pellets were lysed with 800 µl CLS-Y buffer (FastDNA™ Kit, MP Biomedicals, Irvine, CA, USA). The bead-based protocol for isolation was followed according to the manufacturer's instructions.

Two biological fungal mock communities (M1 and M2) were included in the study. Both were constituted from environmental fungi: M1 was composed of wood-decomposing polypore fungi (*Mycena galopus*, *Mycena galericulata*, *Mycena leptocephala*, *Mycena epipterygia*, *Serpula lacrymans*, and *Amanita muscaria*), and M2 was constructed from eight fungi isolated from air (*Boeremia exigua* var. *exigua*, *Cladosporium*, *Penicillium chloroleucon*, *Aspergillus fumigatus*, *Discostroma fuscellum*, *Paraphaeosphaeria michotii*, *Mucor hiemalis*, and *Leptosphaerulina chartarum*).

Three single-species positive controls were also prepared from three *Candida* reference strains (*C. albicans* ATCC 10231, *Candida parapsilosis* ATCC 22019, and *Candida glabrata* ATCC MYA-2955).

Serially diluted samples of fungal species isolated from a healthy volunteer and grown on Sabouraud dextrose agar (SDA; Sigma-Aldrich, St. Louis, MO, USA) at 37°C for 48 h were also included as controls. Dilutions (from 1:10 up to $1:10^6$) were done in both artificial saliva (Saliva Orthana®, NycoDent, Asker, Norway) and human saliva from a volunteer that did not

grow fungi when cultured on SDA. The experimental setup also included three negative controls, two of which were negative extraction controls and the third one just nuclease-free water added before library normalization.

ITS Amplicon PCR

PCR amplification was performed in a 25-µl reaction volume using 12.5 µl of KAPA HiFi HotStart® ReadyMix PCR Master Mix (Kappa Biosystems, Sigma-Aldrich) and 1 µl of the DNA template, in addition to 0.5 µl of each reverse and forward primers and nuclease-free water. The internal transcribed spacer 2 (ITS2) subregion was targeted for amplification, as recommended (Knot et al., 2009; Nilsson et al., 2019a). ITS2 universal primer 5, 8S ITS2-F GTGAATCATCGARTCT TTGAA, and 28S1 ITS2-R TATGCTTAAGTTCAGCGGGTA (TIB, MOLBIOL, Berlin, Germany) were used to amplify the region of interest. The Veriti Thermal Cycler® (Applied Biosystems, Foster City, CA, USA) was used for amplification. Thermal cycling was done as follows: 3 min at 95°C, initial denaturation followed by 45 cycles of 30 s at 95°C: denaturation, 60 s at 58°C as annealing, 30 s at 72°C for the extension, and a final extension at 72°C for 5 min. The PCR products were examined by electrophoresis in a 1% (*w*/*v*) agarose gel in 1× TAE buffer.

PCR Clean-up and Library Preparation

Two rounds of clean-up, one after amplicon PCR and the other after index PCR, were performed using a bead-based method (Agencourt AMPure XP, Beckman Coulter, Brea, CA, USA). After the first round, 5 µl from each cleaned up sample was transferred to a 96-well PCR plate for indexing. The indices were arranged according to the manufacturer's protocols.

Index PCR and Library Normalization and Denaturation

Nextera XT index primers (Illumina, San Diego, CA, USA) were used for indexing. Index PCR was carried out on the Veriti Thermal Cycler® (Applied Biosystems) with parameters recommended by Illumina (San Diego, CA, USA).

One microliter of a 1:50 dilution of each sample was used for library validation using a Bioanalyzer® DNA 1000 Chip (DNA LabChip® using 2100 Bioanalyzer, Agilent Technologies, Santa Clara, CA, USA). The DNA concentrations of the index PCR products were measured with the Qubit 3.0 Fluorometer® (Invitrogen, Carlsbad, CA, USA), and the DNA concentration was calculated in nanomolars based on the size of the DNA amplicons determined using Bioanalyzer®. The normalized library was combined with HT1 and PhiX, as recommended by Illumina.

The MiSeq Reagent Kit v.3 (600 cycles; Illumina, San Diego, CA, USA) was used for library denaturation and MiSeq sample loading. Sequencing was performed on the Illumina MiSeq platform using a 2 × 300-bp paired-end protocol.

Bioinformatics Processing

Demultiplexed Illumina-generated paired-end sequences were processed using QIIME 2 (version QIIME2-2020.8) (Bolyen et al., 2019). The ITSxpress QIIME 2 plugin (v.1.3) (Rivers et al., 2018) was used to extract the ITS2 region. The sequences were then passed through the DADA2 pipeline (Callahan et al., 2016) for filtration, dereplication, chimera detection, and the merging of paired-end reads to create the so-called amplicon sequence variants (ASVs). The resultant ASVs were included for further analysis. The UNITE database (version 8) (Nilsson et al., 2019b) was trained to create a naive Bayes classifier in order to classify the sequences obtained from the DADA2-generated ASV table. Post-clustering curation using LULU (Frøslev et al., 2017) was performed to avoid diversity overestimation. Unidentified ASVs in the UNITE database were blasted to NCBI and the taxonomy for each was reassigned (considering an *e*-value and similarity or coverage ≥99% of the best hit). Various taxonomic levels were used to classify the sequence data. Species with low abundance (20 reads in less than five samples) were discarded. Three OSCC saliva samples and one non-OSCC control were excluded due to the exclusion criteria for low-abundance samples.

Statistical Analyses

Differences in the composition of the mycobiome between the OSCC and healthy control groups, and within samples, were tested for significance using relevant statistical tests in MicrobiomeR (Lahti et al., 2017), Phyloseq (McMurdie and Holmes, 2013), and MicrobiomeAnalystR (Dhariwal et al., 2017). Alpha diversity was calculated and plotted in Phyloseq, R version 4.0.3. QIIME2 ANCOM parameters (Bandara et al., 2019) and ALDEx2 (Fernandes et al., 2013) plugins were used for the analysis of the composition of microbiomes. The Kaplan–Meier survival estimator and Cox proportional hazards models (with "enter" method) were used for survival analysis, with OS of 2 years after diagnosis as the end point; all patients who were alive or lost to follow-up at the end of data collection were censored. Survival analysis was performed using Statistical Package for Social Sciences (SPSS), version 25 (IBM, Armonk, NY, USA). For all analyses, *p*-values ≤0.05 were considered to be significant.

RESULTS

Cohort Description

The prospective cohort included 59 patients (age range = 25–87 years, mean = 50.6 years, median = 60 years) with histologically proven OSCC and 13 non-OSCC controls (age range = 30–70 years, mean = 46.5 years, median = 45 years). Patients in the OSCC group presented more tobacco consumption (expressed in pack-years for both smoking and smokeless tobacco taken together) than did the controls, although the difference was not statistically significant ($p = 0.06$) (**Table 1**). The average number of decayed teeth (DT) was similar to the general Sudanese population, as previously evaluated (Khalifa et al., 2012), except for the age groups 25–44 and >65 years in our cohort, which showed a higher number of decayed teeth compared to the general Sudanese population. The same was found for missing teeth (MT) (**Table 1**). The mean plaque index of OSCC patients

TABLE 1 | Demographics, oral health findings, and clinicopathological findings of the cohort.

Cohort demographics

		Non-OSCC controls	*Patients*
No. of individuals		13 (7 males, 6 females)	59 (42 males, 17 females)
Age (years), mean	Males	45.4 (30–60)	60 (25–87)
	Females	47.7 (39–70)	60.5 (40–80)
No. of users		2 (15%, all males)	33 (56%, all males)
Pack-years (PY), mean (p = 0.06)**		4.5	51.1

Oral findings

	Non-OSCC controls	*OSCC patients*
DT (p = 0.630)	2.3	3.9
MT (p = 0.287)	6.9	8.8
Community periodontal index, mean ± SD (CPI: p = 0.013*)	1.59 ± 0.67	1.79 ± 0.64
Gingival index, mean ± SD (p = 0.014*)	1.19 ± 0.31	1.67 ± 0.56

Missing teeth and decayed teeth in the OSCC cohort[a]

Age groups (years)	*DT*			*MT*		
	Non-OSCC	*OSCC*	*General population**	*Non-OSCC*	*OSCC*	*General population**
25–34	0	10	3.3	0	0	1.9
35–44	1.4	4.8	4.1	3.8	3.6	4.2
45–54	3.2	3	4	3.8	6.3	5.5
55–64	5	3.1	3.9	20	8.2	8
65–74	–	4.7	3	32	9.2	11.3
75+	–	3.5	3.3	–	14	11.8

Tobacco and alcohol consumption

History	*Toombak, N (%)*		*Smoking, N (%)*		*Alcohol, N (%)*	
	Patients	*Non-OSCC controls*	*Patients*	*Non-OSCC controls*	*Patients*	*Non-OSCC controls*
Yes—current user	5 (8.5)	0 (0)	7 (11.9)	2 (15.4)	1 (1.7)	0 (0)
No	30 (50.8)	11 (84.6)	38	10 (77)	39 (66.1)	11 (86.6)
Past user	22 (37.3)	1 (7.7)	12	0 (0)	13 (22)	1 (1.7)
Unknown	2 (3.4)	1 (7.7)	2 (3.4)	1 (7.7)	6 (10.2)	1 (1.7)

OSCC patients: clinical findings

Tumor location	*N (%)*	*Tumor stage*			*T stage*	*N stage*		*M stage*	
Buccolabial–sulcus	32 (54.2)		*N (%)*		*N (%)*		*N (%)*		*N (%)*
				T1	2 (3.4)	N0	4 (6.8)	M0	39 (66.1)
Tongue	5 (8.5)	I	0 (0)	T2	11 (18. 6)	N1	19 (32.2)	M1	1 (1.7)
Retromolar–palatal–alveolar	15 (25.4)	II	2 (3.4)	T3	15 (25.4)	N2	27 (45.8)	Mx	12 (20.3)
		III	13 (22)	T4	20 (33.9)	N3	1 (1.7)		
		IV	37 (62.7)	Tx	4 (6.8)	Nx	1 (1.7)		
Missing, *N* (%)					7 (11.9)				

OSCC, oral squamous cell carcinoma; DT, mean number of decayed teeth; MT, mean number of missing teeth.
*Significantly different at $p < 0.05$ (*Kruskal–Wallis and **Mann–Whitney U test).*
[a]*Based on Khalifa et al. (2012).*

was comparable to that of the control group (p = 0.59), while the gingival index (mean ± SD = 1.67 ± 0.56) was significantly higher for the OSCC group (p = 0.014) than that for the control group (mean ± SD = 1.19 ± 0.31).

The localization of OSCC lesions was predominantly lower buccal or labial (40.4%); only five cases (6.9%) were localized on the tongue. Of all OSCC patients, 47 (79.6%) presented with locoregional lymph node metastases at the time of diagnosis. Nearly all OSCC patients (96.8%) presented at a late stage (**Table 1**).

Method Performance

A total of 21,698,808 Illumina-generated demultiplexed fungal ITS raw paired-end sequences were imported into QIIME2. The extracted ITS2 region was merged and temporarily clustered into 6,699,920 amplicon reads. After DADA2 filtration, dereplication, chimera detection, and merging of paired-end reads, a total of 3,514,250 reads were retained for further analysis. The quality-filtered, denoised, chimera-removed sequence reads were clustered into 514 ASVs. Post-clustering curation using LULU (Frøslev et al., 2017) and removal of contaminants using the Decontam algorithm (Davis et al., 2018) retained 340 ASVs. The rarefaction curves curves are presented in **Supplementary Figure 1**.

The positive and negative controls showed the expected reference strains (for positive ones) and negative outputs (negative ones). The M2 mock community showed good distribution, while M1 showed a generally quite good coverage,

although taxonomic assignment was obtained correctly only to the class level (*Amanita* and *Mycena* in M1 were identified at the order level, i.e., Agaricomycetes). The composition of our mock communities is reflected in the analysis results, indicating minimal cross-contamination and tag switching. The distribution pattern of the total reads followed the serial dilutions we made (**Supplementary Figure 2**).

Abundance analysis of the serially diluted samples showed a pattern corresponding with the inputs of the diluted samples (**Supplementary Figure 3**).

Candida, *Saccharomyces*, *Malassezia*, *Aspergillus*, and *Cyberlindnera* Were Identified to Be the Most Common Fungi Present in the Salivary Mycobiome

Processed, quality-filtered ASVs were assigned to 36 different fungal genera. Relative abundance analysis showed that the salivary mycobiome was dominated by five genera, namely, *Candida*, *Saccharomyces*, *Malassezia*, *Aspergillus*, and *Cyberlindnera* (**Figure 1A** and **Supplementary Figure 4**). *Agaricus*, *Alternaria*, *Cladosporium*, *Clavispora*, *Naganishia*, *Nakaseomyces*, *Penicillium*, *Rhizopus*, *Vishniacozyma*, and *Sarocladium* were the second most commonly identified fungal genera (**Figure 1B**). *Candida* was found to have a higher relative abundance in the saliva of females than that of males and accounted for more than half of the genera present in females (**Figure 2A**). There was no difference in the diversity of the salivary mycobiome between females and males (**Figure 2B**).

Eight genera were detected exclusively in the saliva of tobacco users (when analyzing together *toombak* dippers and smokers), of which seven were shared by smokers and users of *toombak* (**Figures 2C–E**): *Macrophomina*, *Schizophyllum*, *Cinereomyces*, *Leucosporidium*, *Rhodosporidiobolus*, *Cutaneotrichosporon*, and an unidentified one belonging to the family Ustilaginaceae. *Lodderomyces* was detected only in the saliva of smokers. *Phlebiopsis* and *Filobasidium* were detected only in the saliva of non-tobacco users. No statistically significant differences in the overall oral mycobiome diversity were observed between non-tobacco users and smokers or *toombak* users, even when considering only the OSCC cases (**Figures 2F–H**), although a

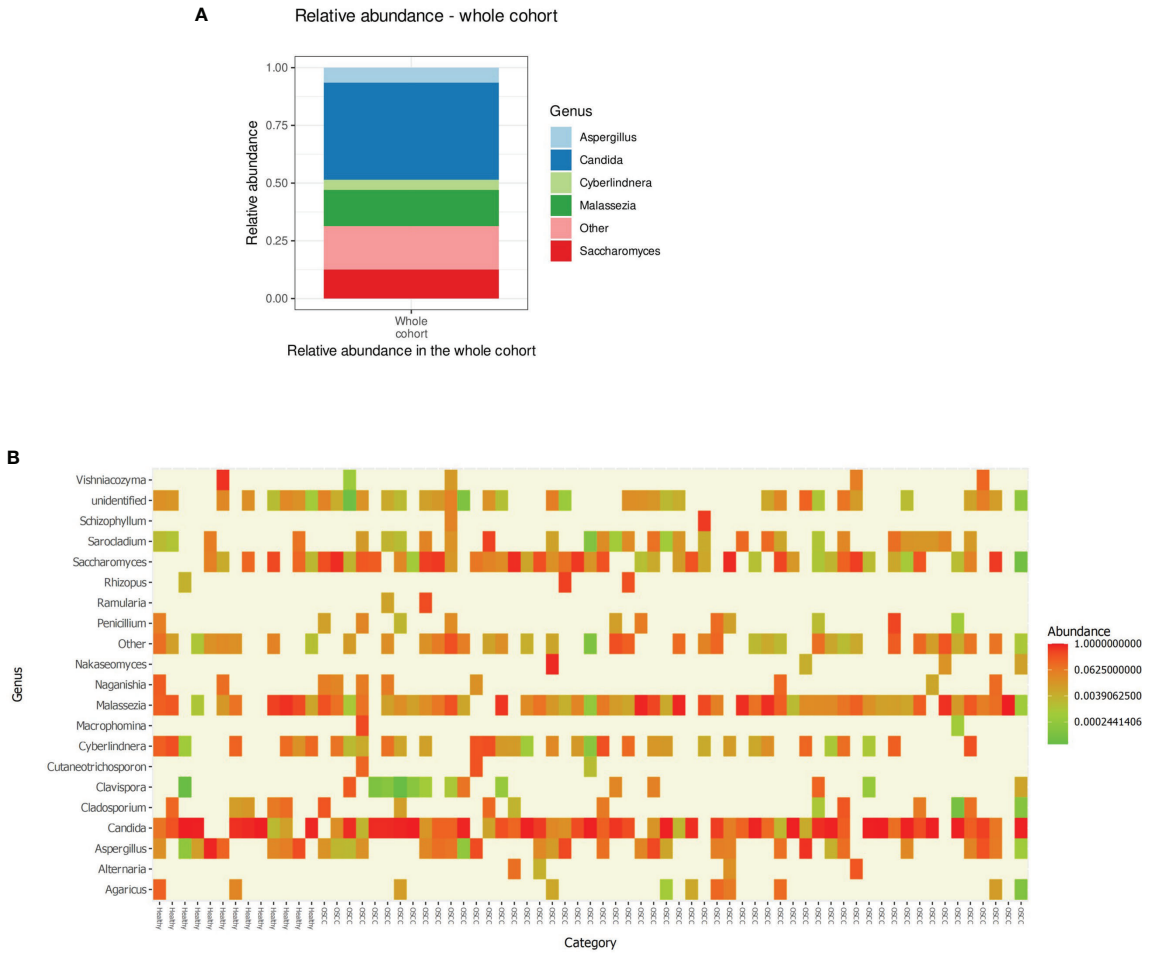


FIGURE 1 | **(A)** Relative abundance of the top five genera in the saliva of the individuals investigated in our cohort. **(B)** Heat map showing the relative abundance of the top 20 genera (*X*-axis sorted non-OSCC controls to the left and OSCC to the right) in each of the investigated sample. OSCC, oral squamous cell carcinoma.

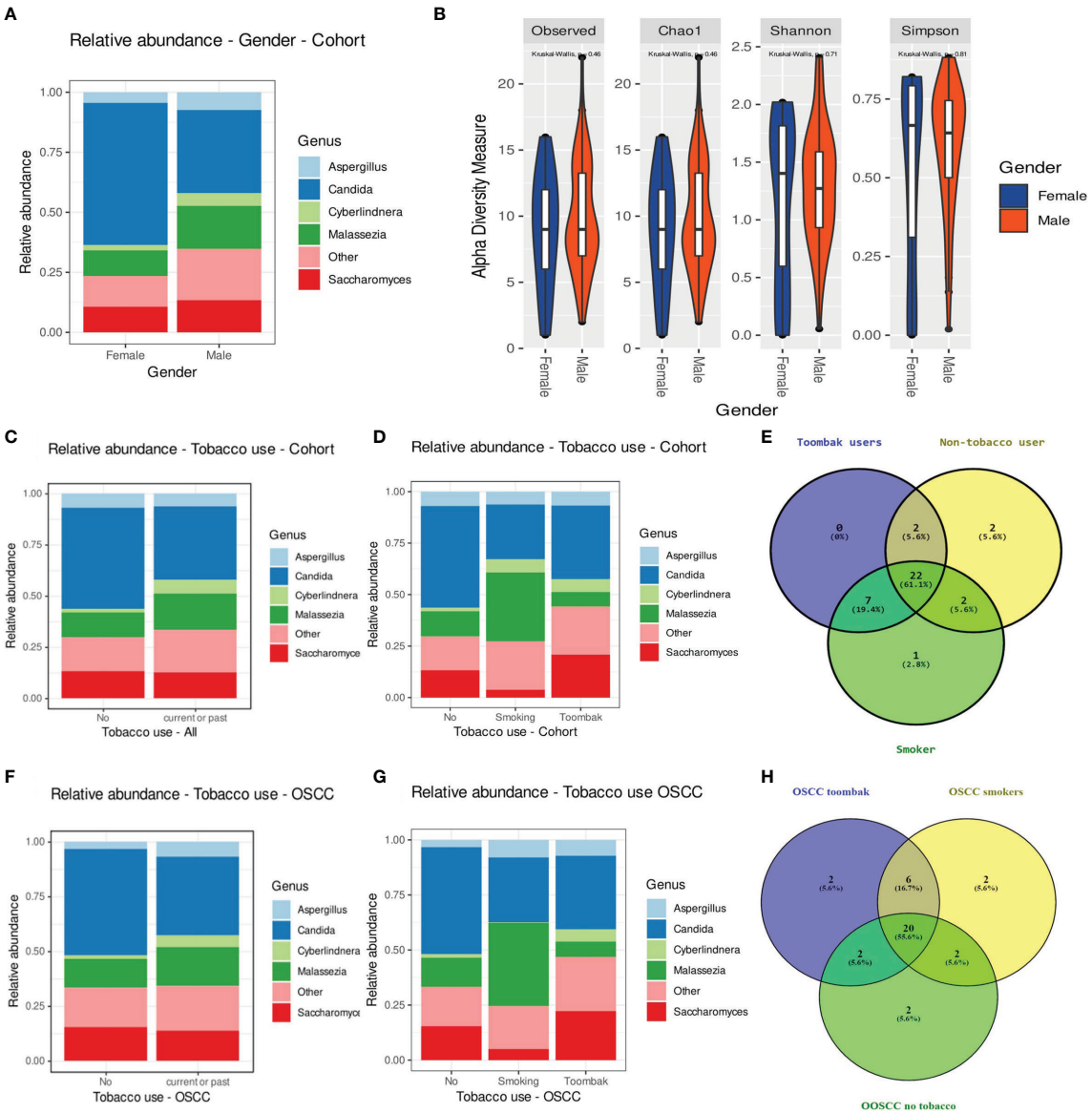


FIGURE 2 | **(A, B)** Relative abundance of the top five genera **(A)** and diversity of the salivary mycobiome **(B)** of the individuals investigated in our cohort grouped by gender. **(C)** Relative abundance of the top five genera in tobacco users *versus* non-smokers. **(D)** Relative abundance of the top five genera in smokers, *toombak* users, and non-smokers. **(E)** Venn diagram showing the distribution of genera in smokers, *toombak* users, and non-smokers. **(F)** Relative abundance of the top five genera in tobacco users *versus* non-smokers in the oral squamous cell carcinoma (OSCC) group. **(G)** Relative abundance of the top five genera in smokers, *toombak* users, and non-smokers in the OSCC group. **(H)** Venn diagram showing the distribution of genera in smokers, *toombak* users, and non-smokers in the OSCC group.

trend toward somehow restricted diversities in smokers or *toombak* users was observed (**Supplementary Figure 6**).

Individuals aged 55–64 years showed the least relative abundance of *Candida* and the highest abundance of *Aspergillus* in their saliva (**Supplementary Figure 5A**). Individuals with severe gingivitis showed a predominance of species other than the identified top five genera in their saliva (**Supplementary Figure 7A**), along with a gradually reduced diversity, compared to the other two groups. Individuals who needed complex periodontal treatments such as root planing or periodontal surgical procedures (CPI higher than 3) showed a higher relative abundance of *Aspergillus* and a lower relative abundance of *Malassezia* than did those in the other two groups (**Supplementary Figure 7B**). Individuals with intermediate CPI (1.1–3), who needed to undergo plaque control procedures, showed the lowest diversity of fungi compared to other subjects with clinically higher or lower CPIs. Individuals with poor oral hygiene, quantified by the use of a simplified oral

hygiene index, showed higher relative abundance of *Candida*, *Aspergillus*, and *Saccharomyces* and a trend toward a lower diversity of fungi (**Supplementary Figure 7C**).

Individuals with the number of decayed teeth higher than that of the mean value for the Sudanese population had a lower relative abundance of *Candida* but a higher relative abundance of *Saccharomyces* than the rest of the participants (**Supplementary Figure 7D**). The opposite was observed for individuals with the number of missing teeth higher than that of the mean value for the Sudanese population (**Supplementary Figure 7E**). The alpha diversity median was also higher, although statistically not significant, for the salivary mycobiome of individuals with more decayed and missing teeth.

Sixteen Genera Were Identified Exclusively in the Salivary Mycobiome of OSCC Patients

The extracted DNA content in the samples from OSCC patients was significantly higher than that in the samples from non-OSCC controls, as evaluated using two different approaches (Qubit® and Bioanalyzer®) ($p < 0.05$). Twenty genera were found in the saliva of both the OSCC and non-OSCC groups. Sixteen genera were found exclusively in the saliva of OSCC patients (**Figure 3A**). Univariate statistical comparison of the relative abundance of the top five genera showed no statistically significant differences between the two groups; the same top five most abundant genera were found in both groups (**Figures 3B** and **4**). Alpha diversity analysis, considering richness and evenness, did not show statistically significant differences between OSCC patients and non-OSCC controls (**Figure 3C**). Non-metric multidimensional scaling (NMDS) and analysis of similarities (ANOSIM)/permutational multivariate analysis of variance (PERMANOVA) were applied in order to test for dissimilarities in the mycobiome composition between OSCC patients and non-OSCC controls. There was no shift observed between the study groups (NMDS stress > 0.2); statistical significance was marginal with ANOSIM ($p_{ANOSIM} = 0.056$) and non-significant with PERMANOVA ($p = 0.265$). Differential abundance analysis using ANCOM and ALDex2 did not show any differentially abundant genera when comparing the OSCC group and the non-OSCC control group.

Although not statistically significant, the salivary carriage of *Candida* was higher in the saliva of OSCC cases than that in non-OSCC controls (relative abundance and log-transformed count in each case shown in **Figures 3B** and **4A**, respectively). The *Candida* species identified in the saliva of those in the OSCC group were *C. albicans* (78.8% of all OSCC cases), *Candida tropicalis* (32.1%), *C. parapsilosis* (37.5%), *C. glabrata* (16.1%), *Candida orthopsilosis* (3.6%), and *Candida sake* (9%). *C. orthopsilosis* and *C. sake* were among the fungi identified exclusively in the saliva of OSCC patients.

In the saliva of OSCC cases, *Saccharomyces* also had a higher abundance than that in the saliva of non-OSCC controls (relative abundance and log-transformed count in each case shown in **Figures 3B** and **4C**, respectively). *Saccharomyces cerevisiae* was second to *C. albicans* in the saliva of OSCC cases (76.8% of OSCC cases), while *Malassezia arunalokei* was the second most predominant species in the saliva of non-OSCC controls (66.70% of controls, $n = 12$; 64.3% of OSCC patients, $n = 56$). Additionally, different species of *Malassezia* were identified in the saliva of OSCC patients and non-OSCC controls. *Malassezia globosa* (64.3%), *Malassezia restricta* (16%), *Malassezia dermatis* (5.4%), *Malassezia furfur* (3.5%), and *Malassezia slooffiae* (1.8%) were identified in the saliva of OSCC patients. In the saliva of non-OSCC controls, only *M. restricta* (33.3%) and *M. globosa* (58.3%) were identified.

Cyberlindnera had lower abundance in the saliva of OSCC cases than that of non-OSCC controls (relative abundance and log-transformed count in each case shown in **Figures 3B** and **4E**, respectively). *Cyberlindnera jadinii* (synonym: *Pichia jadinii*) was detected in the saliva of 50% of the OSCC patients, while it was present in 58.3% of the non-OSCC controls, showing an inverse relation to *C. albicans* in more than half of the whole group (56% of the whole cohort), although the bivariate correlation was statistically not significant (correlation = −0.257, $p = 0.1$).

Malassezia Was Identified as an Independent Predictor of OS for OSCC Patients

The saliva of OSCC patients with tumors located in labial, buccal, or alveolar areas (*toombak* dipping areas) showed a lower relative abundance of *Candida* but a higher relative abundance of *Cyberlindnera* compared to patients with OSCC located in other sites (**Supplementary Figure 5B**). OSCC patients with locoregional lymph node involvement showed higher relative abundance of *Candida* and *Aspergillus* and a lower relative abundance of *Malassezia* compared to the group with no lymph node involvement (**Supplementary Figure 5C**). The same trend was observed for the OSCC patients who died during the follow-up period compared to those still alive at the end of the study (relative abundance in and log-transformed count in each case shown in **Figures 3D** and **4F–J**, respectively). Alpha diversity analysis revealed that lower diversity index values were more commonly found in OSCC patients with locoregional lymph node involvement and those with poorer survival (**Figure 3E**), although not statistically significant. A trend toward a lower relative abundance of *Saccharomyces* and a higher relative abundance of *Aspergillus* with stage has also been observed (**Supplementary Figure 5D**). Alpha diversity analysis showed no statistically significant differences between stages.

Kaplan–Meier analysis revealed that a high relative abundance of *Candida* was associated with poor OS in OSCC patients (Breslow test: $p = 0.043$) (**Figure 5A**). On the contrary, a high relative abundance of *Malassezia* showed association with favorable survival in OSCC patients (Breslow test: $p = 0.039$) (**Figure 5B**). The Cox proportional hazards multiple regression model was applied to adjust the salivary carriage of both *Candida* and *Malassezia* for age ($p = 0.029$) and identified *Malassezia* as an independent predictor of OS (hazard ratio = 0.383, 95% CI = 0.16–0.89, $p = 0.03$).

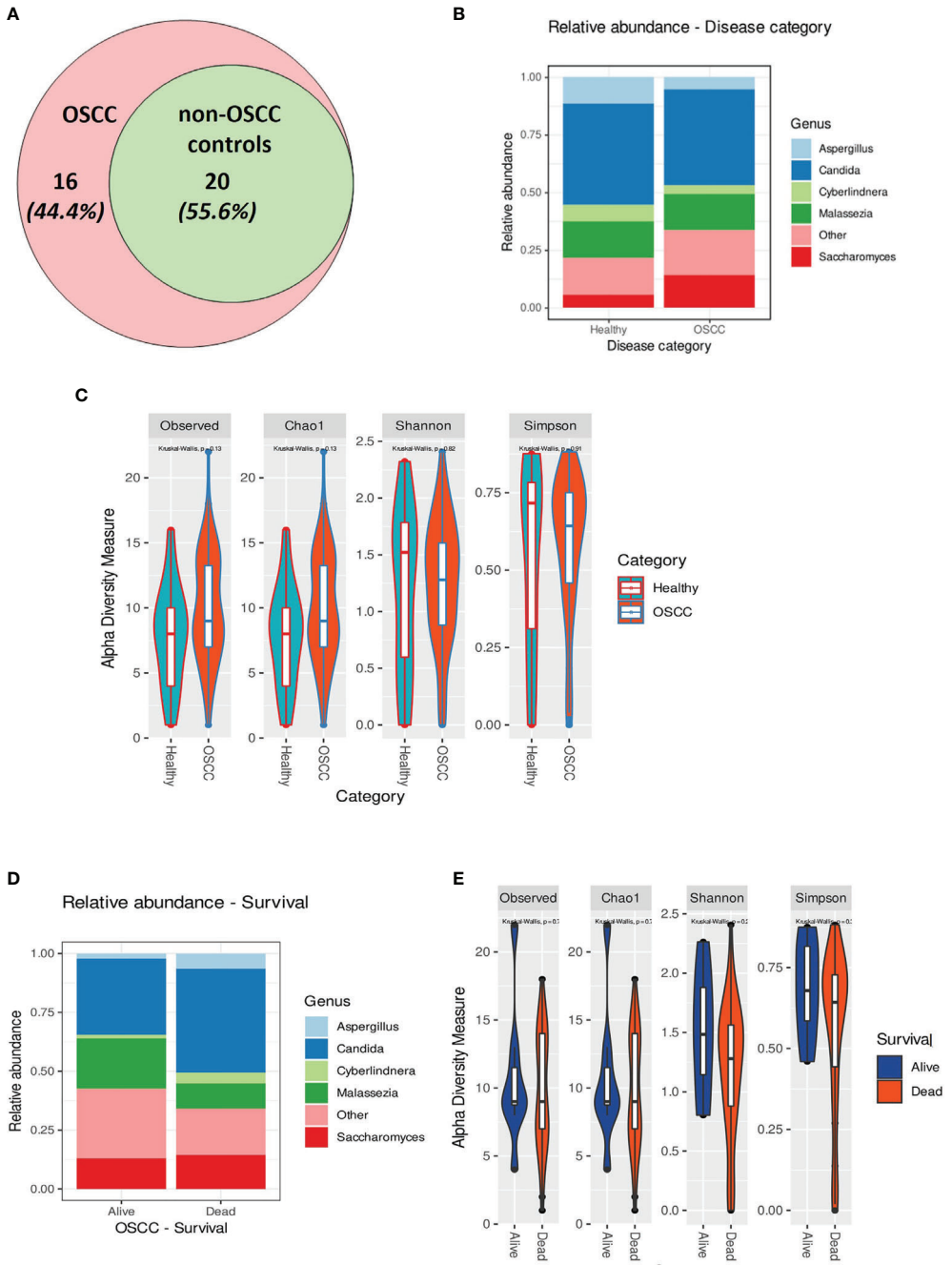


FIGURE 3 | **(A)** Venn diagram showing the number of genera identified in the oral squamous cell carcinoma (OSCC) and non-OSCC groups. Those that were found exclusively in the OSCC group were: *Macrophomina*, *Ramularia*, *Aureobasidium*, *Alternaria*, *Ulocladium*, *Lodderomyces*, *Meyerozyma*, *Schizophyllum*, *Cinereomyces*, *Phlebiopsis*, *Rhodosporidiobolus*, *Rhodotorula glutinis*, *Filobasidium*, *Cutaneotrichosporon*, *unidentified1*, and *unidentified2*. **(B)** Relative abundance in the OSCC and non-OSCC groups showing the top five most predominant genera. **(C)** Alpha rarefaction curve showing the observed features (richness) at different sequencing depths. **(D)** Relative abundance of individuals (alive and dead) in the OSCC groups showing the top five most predominant genera. **(E)** Alpha rarefaction curve showing the observed features (richness) at different sequencing depths for OSCC patients stratified by overall survival (OS).

DISCUSSION AND CONCLUSION

Although the baseline mycobiome profiles utilizing NGS have been established for some time (Ghannoum et al., 2010; Mukherjee et al., 2014; Chandra et al., 2016), studies on the mycobiome in disease and health are scarce, and the actual contribution of the mycobiota in carcinogenesis has only recently been explored (Perera et al., 2017; Al-Hebshi et al., 2019;

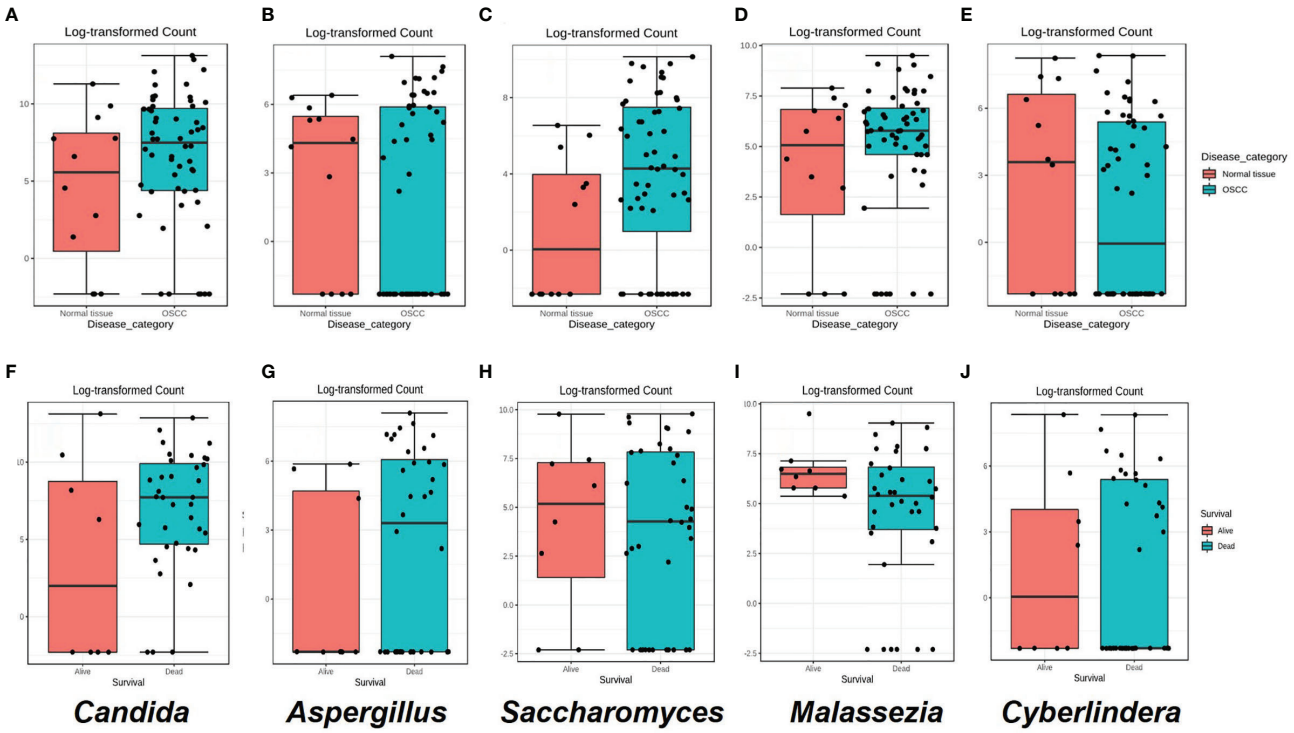


FIGURE 4 | **(A–E)** Box plots for the abundance of the top five genera in the the oral squamous cell carcinoma (OSCC) group *versus* non-OSCC controls. **(F–J)** Box plots for the relative abundance of the top five genera in OSCC patients who were dead or alive at the end of the study period. Classical univariate statistical comparison of the relative abundance showed no statistically significant differences.

Aykut et al., 2019). This study is one of the first characterizing the salivary mycobiome in OSCC and provides significant information despite the fact that it has been run on a relatively smaller number of cases, particularly of non-OSCC controls. Another limitation of this study is that data on antibiotic use were missing. In Sudan, despite instructions, the misuse of antibiotics is a common problem, and the use of antibiotics is known to affect the results of mycobiome analysis (Awad et al., 2007; Oleim et al., 2019).

Most mycobiome studies focused on either the ITS1 or the ITS2 subregion of typically 250–400 bases. Targeting the ITS2 subregion has the additional advantage of including lower length variations and more universal primer sites, resulting in less taxonomic bias than when targeting ITS1 (Nilsson et al., 2019a). In our study, by utilizing 2 × 300-bp sequencing and by merging paired reads, we obtained better taxonomic resolution since the full ITS2 length was covered. We used a relative abundance cutoff of 1%, as used by other studies (Ghannoum et al., 2010; Perera et al., 2017).

The inclusion of negative controls (no saliva sample), positive controls (known species most likely to be found in the samples), and of mock communities was done as a standard for proper assessment and quantification of tag switching, chimera formation, ASV inference stringency, and abundance shifts (Bakker, 2018; Nilsson et al., 2019a). After evaluating the controls, the overall methods used here for DNA extraction, sequencing, and bioinformatics analysis were considered to be sensitive for salivary mycobiome identification, under the conditions and aims of our study.

Although the concept of a healthy core oral mycobiome (Ghannoum et al., 2010) was redefined (Dupuy et al., 2014), with 14 core genera detected in healthy individuals, the overall abundance and diversity of fungal taxa may also be somewhat individualized (Witherden et al., 2017). It is considered that the vast majority of the mycobiome consists of a few genera, with *C. albicans* and *C. parapsilosis* as the major species of the human oral mycobiome (Naglik et al., 2013). The most abundant genera found in our study are in line with these previous baseline findings and with other OSCC-associated salivary mycobiomes reported in previous studies (Ghannoum et al., 2010; Dupuy et al., 2014; Mukherjee et al., 2017; Perera et al., 2017).

Previous studies on the dynamics of the oral bacterial community showed enrichment in both abundance and function with OSCC staging (Yang et al., 2018). We found an enriched but somehow less diverse fungal mycobiome in the most advanced stage group. This might be related to the limited number of cases in the early stages in our cohort. Late tumor stage presentation is typical for OSCC in Sudan (Osman et al., 2010), and as mentioned, this is limiting the conclusions we could draw on the differences between stages in the salivary mycobiome of this cohort.

The salivary microbiome was previously found to be related to dental findings. Gazdeck et al. found a lower bacterial diversity

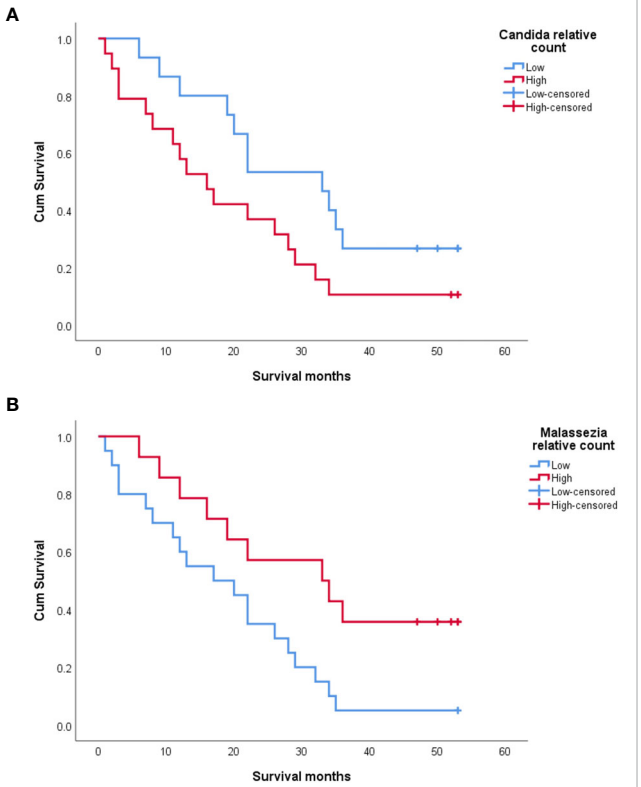


FIGURE 5 | Kaplan–Meier survival curves showing the impact of salivary *Candida* **(A)** and *Malassezia* **(B)** on the overall survival of oral squamous cell carcinoma (OSCC) patients.

in edentulous patients (Gazdeck et al., 2019). We report here higher diversity median values associated with more missing and decayed teeth. This might indicate relevant fungal–bacterial interactions (Deveau et al., 2018) that need further longitudinal studies for final elucidation. For a long time, our understanding of periodontal disease has been based on its bacterial origin (Hajishengallis and Lamont, 2012). However, the crosstalk between fungi and bacteria seems to result in different outcomes for the host. This relationship ranges from synergism to antagonism for different specific microbial interactions (Krüger et al., 2019). Peters et al. showed *Candida* species to be more represented in subjects with periodontal disease and more missing teeth count (Peters et al., 2017), and, in accordance with this, *C. albicans* was shown to enhance *Porphyromonas gingivalis* invasion *in vitro* (Tamai et al., 2011). We observed the same trend for individuals with higher number of missing teeth.

When it comes to its role in carcinogenesis, in addition to the direct role of *Candida* by producing nitrosamines, it was shown that it also affects the metabolism of procarcinogens and influences other bacteria, which may play a role in carcinogenesis (Hooper et al., 2009). The combinatorial effect of carcinogens and *C. albicans* was shown to promote OC in a murine model (Dwivedi et al., 2009). *C. albicans* was also shown to enhance the invasion of OSCC cells by producing specific proteinases capable of degrading the basement membrane and the extracellular matrix (Bakri et al., 2010). The inflammatory response to *C. albicans* is mediated by NF-κB (Müller et al., 2007), which is frequently involved in carcinogenesis where cancer-related inflammation is a feature (Mantovani et al., 2008). Taking this into consideration, our finding of the association between *Candida* and poor prognosis might rely on a biological explanation.

On the other hand, we found the salivary carriage of *Malassezia* as an independent predictor of better prognosis. *Malassezia* has a unique pattern of interaction with pattern recognition receptors compared to *C. albicans* (Goyal et al., 2018). Additionally, *Malassezia* has large intraspecies diversity. The exact composition of different *Malassezia* species at a time point may contribute to different outcomes in the interaction between the fungus and the host (Sparber and LeibundGut-Landmann, 2017). *Malassezia* might have been overrepresented in our study, although it was described as part of the redefined core oral mycobiome in humans (Dupuy et al., 2014). Since *Malassezia* is a normal skin commensal with population densities peaking between 20 and 45 years (Ashbee, 2007), the sample collection method we used might have included some contamination from the lips, in addition to the age-related differences in *Malassezia* enrichment. However, the interest in discovering microbiome profiles associated with survival is growing (Plantinga et al., 2017; Koh et al., 2018). Different methods are used to associate the microbiota at the community level and censored survival time, such as MiRKAT-S (Plantinga et al., 2017) and its follower OMiSA (Koh et al., 2018). We chose to test the categorization of microbial proportions into high and low and run conventional survival analysis. This seems an attractive way to incorporate microbiome signatures in clinically applicable diagnostic tools.

Of interest is that we did not identify *Hannaella* and *Gibberella*, which were found enriched in a cohort of OSCC patients in a recent study and considered to be contaminants (Perera et al., 2017). This might be an indication of the effects of dietary habits, among others, and population-related differences in the mycobiome profiles. These two species are known plant fungi. Nevertheless, the contribution of environment-related fungi to the carcinogenic process cannot be ignored.

The sample type, the method of collection, and, very importantly, the methods for DNA extraction and bioinformatics processing could explain the observed differences between species reported in different studies (Brooks et al., 2015). Curation of the databases used for taxonomical assignment could also affect the findings (Seed, 2015), in addition to the more classical factors such as ethnic differences and diet (Deschasaux et al., 2018). Ethnic differences could be related to different single nucleotide polymorphisms associated with susceptibility to fungi (Romani, 2011). The role of genetic host susceptibility should not be ignored when considering the diversity changes or the associations of the mycobiota with diseases. In any case, a further, more thorough investigation of mycobiome meta-transcriptomes and metaproteomes is needed to answer such questions related to the epidemiological patterns of mycobiomes (Huttenhower et al., 2012).

Also worth mentioning is the fact that the cohort analyzed in this study included patients consuming a special type of smokeless tobacco, the "*toombak*" (the local form of smokeless tobacco used in Sudan). Regarding our findings on tobacco consumption (both smoking and the smokeless tobacco *toombak*), it is worth noting that there may be bias related to self-reporting. However, the predominant site for tumor localization in our OSCC cohort was lower buccal or labial and sulcular, consistent with a *toombak*-related OSCC etiology, and this also correlated with the self-reported habit of packing *toombak* in the mouth in our cohort. Self-reporting of alcohol consumption should be considered with caution as well since it is illegal in Sudan and may carry a social stigma (Gadelkarim Ahmed and Ahmed, 2013). Previous studies have shown that tobacco exposure was associated with a shift of the oral bacteriome at the population level (Beghini et al., 2019). Our study showed the same trend for the oral mycobiome. Some of the genera were identified exclusively in tobacco users, including *toombak* users, since many consume *toombak* besides smoking, and some of the genera, such as *Schizophyllum*, are known plant pathogens; thus, they may be related to the processed tobacco product.

In conclusion, the present study reveals that *Candida*, *Malassezia*, *Saccharomyces*, *Aspergillus*, and *Cyberlindnera* are the most relatively abundant fungal genera in the salivary microbiome of this cohort of Sudanese individuals. *Candida* and *Malassezia* were shown to have an impact on the survival of OSCC patients: a higher salivary carriage of the genus *Candida* was found to be associated with poor prognosis, while *Malassezia* was enriched in patients with favorable prognosis, although only the salivary carriage of *Malassezia* emerged as an independent prognostic biomarker for the survival of OSCC patients. This can serve as groundwork for performing mycobiome-based biomarker studies in larger cohorts of OSCC patients.

DATA AVAILABILITY STATEMENT

The raw data represented in this study has been deposited and is publicly available from SRA (PRJNA722859); NCBI PRJNA722859.

ETHICS STATEMENT

The National Health Research Ethics Committee, Federal Ministry of Health, Sudan, approved the research in Sudan (fmoh/rd/SEC/09). Also, the Regional Ethical Committee in Norway approved this project (REKVest 3.2006.2620 REKVest 3.2006.1341). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

NM and DC conceptualized the study. NM, JL, IA, EM, JF, RJ-L, LM, SM, TO, and EN helped with the methodology. NM, EM, JF, RJ-L, LM, SM, TO, and EN did the formal analysis. NM, JL, IA, ME, TO, EN, AS, AJ, and DC contributed to the investigation. NM prepared the original draft. All co-authors reviewed and edited the manuscript. NM, EM, JF, RJ-L, SM, TO, and DC contributed to visualization. AJ, TO, EN, AS, and DC supervised the study. EN, AS, AJ, and DC administered the project. AJ and DC helped with funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Research Council of Norway through its Centers of Excellence funding scheme (grant no. 22325), Helse Vest (grant no. 911902/2013 and 912260/2019), and the Norwegian Centre for International Cooperation in Education (project no. NORPART-2018/10277).

ACKNOWLEDGMENTS

The authors would like to thank to Prof. Audun Nerland, Dr. Øyvind Kommedal, Dr. Christine Drengnes, Mrs. Tharmini Kalananthan, and Mrs. Sonja Ljostveit for help with optimization of DNA extraction and sequencing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2021.673465/full#supplementary-material

Supplementary Figure 1 | Sequencing depth curves for OSCC and non-OSCC controls.

Supplementary Figure 2 | Pie charts depicting expected and detected distribution of mock communities. Stars refers to identical genera detected.

Supplementary Figure 3 | Bar-graph representing the number of reads according to the serial dilution of samples. Y axis represents log scale of proportions. AS - sample diluted in Artificial Saliva, HS - sample diluted in Human Saliva that did not grow fungi when cultured on Sabraud's Dextrose agar medium.

Supplementary Figure 4 | Bar plots showing the relative abundance in saliva of OSCC cases and non-OSCC controls. Five most dominant genera were shown.

Supplementary Figure 5 | (A) Distribution and differences in abundance of 5 topmost salivary fungi according to age. **(B)** Distribution and differences in abundance of 5 topmost salivary fungi according to tumour localization. **(C)** Distribution and differences in abundance of 5 topmost salivary fungi according to lymph node metastasis status. **(D)** Distribution and differences in abundance of 5 topmost salivary fungi according to tumour stage.

Supplementary Figure 6 | Diversity of the overall oral mycobiome of the individuals of our cohort grouped by tobacco use (no tobacco users, smokers and *Toombak* users).

Supplementary Figure 7 | (A) Distribution and differences in abundance of 5 topmost salivary fungi according to Gingival Index. **(B)** Distribution and differences in abundance of 5 topmost salivary fungi according to Community Periodontal Index.

(C) Distribution and differences in abundance of 5 topmost salivary fungi according to Oral Hygiene Index (Simplified). **(D)** Distribution and differences in abundance of 5 topmost salivary fungi according to mean number of decayed teeth. **(E)** Distribution and differences in abundance of 5 topmost salivary fungi according to mean number of missing teeth. **(F)** Distribution and differences in abundance of 5 topmost salivary fungi according to diabetes status.

REFERENCES

Aas, J. A., Paster, B. J., Stokes, L. N., Olsen, I., and Dewhirst, F. E. (2005). Defining the Normal Bacterial Flora of the Oral Cavity. *J. Clin. Microbiol.* 43, 5721–5732. doi: 10.1128/JCM.43.11.5721-5732.2005

Al-Hebshi, N. N., Borgnakke, W. S., and Johnson, N. W. (2019). The Microbiome of Oral Squamous Cell Carcinomas: A Functional Perspective. *Curr. Oral. Heal. Rep.* 6, 145–160. doi: 10.1007/s40496-019-0215-5

Arweiler, N. B., Netuschil, L., Arweiler, N. B., and Netuschil, L. (2016). "The Oral Microbiota," in *Microbiota of the Human Body Implications in Health and Disease*. Ed. A. W. Walker (Cham: Springer), 45–60. doi: 10.1007/978-3-319-31248-4_4

Ashbee, H. R. (2007). Update on the Genus Malassezia. *Med. Mycol.* 45, 287–303. doi: 10.1080/13693780701191373

Avila, M., Ojcius, D. M., and Yilmaz, Ö. (2009). The Oral Microbiota: Living With a Permanent Guest. *DNA Cell Biol.* 28, 405–411. doi: 10.1089/dna.2009.0874

Awad, A. I., Ball, D. E., and Eltayeb, I. B. (2007). Improving Rational Drug Use in Africa: The Example of Sudan. *East. Mediterr. Heal. J.* 13, 1202–1211. doi: 10.26719/2007.13.5.1202

Aykut, B., Pushalkar, S., Chen, R., Li, Q., Abengozar, R., Kim, J. I., et al. (2019). The Fungal Mycobiome Promotes Pancreatic Oncogenesis *via* Activation of MBL. *Nature* 74 (7777), 264–267. doi: 10.1038/s41586-019-1608-2

Baker, J. L., Bor, B., Agnello, M., Shi, W., and He, X. (2017). Ecology of the Oral Microbiome: Beyond Bacteria. *Trends Microbiol.* 25, 362–374. doi: 10.1016/j.tim.2016.12.012

Bakker, M. G. (2018). A Fungal Mock Community Control for Amplicon Sequencing Experiments. *Mol. Ecol. Resour.* 18, 541–556. doi: 10.1111/1755-0998.12760

Bakri, M. M., Hussaini, H. M., Holmes, A., Cannon, R. D., and Rich, A. M. (2010). Revisiting the Association Between Candidal Infection and Carcinoma, Particularly Oral Squamous Cell Carcinoma. *J. Oral. Microbiol.* 2 (1), 5780. doi: 10.3402/jom.v2i0.5780

Bandara, H. M. H. N., Panduwawala, C. P., and Samaranayake, L. P. (2019). Biodiversity of the Human Oral Mycobiome in Health and Disease. *Oral. Dis.* 25, 363–371. doi: 10.1111/odi.12899

Barrett, A., Kingsmill, V., and Speight, P. (2008). The Frequency of Fungal Infection in Biopsies of Oral Mucosal Lesions. *Oral. Dis.* 4, 26–31. doi: 10.1111/j.1601-0825.1998.tb00251.x

Beghini, F., Renson, A., Zolnik, C. P., Geistlinger, L., Usyk, M., Moody, T. U., et al. (2019). Tobacco Exposure Associated With Oral Microbiota Oxygen Utilization in the New York City Health and Nutrition Examination Study. *Ann. Epidemiol.* 34, 18–25.e3. doi: 10.1016/j.annepidem.2019.03.005

Berkovits, C., Tóth, A., Szenzenstein, J., Deák, T., Urbán, E., Gácser, A., et al. (2016). Analysis of Oral Yeast Microflora in Patients With Oral Squamous Cell Carcinoma. *Springerplus* 5, 1257. doi: 10.1186/s40064-016-2926-6

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Chase, J., Cope, E. K., et al. (2019). Reproducible, Interactive, Scalable and Extensible Microbiome Data Science Using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi: 10.1038/s41587-019-0209-9

Brooks, J. P., Edwards, D. J., Harwich, M. D.Jr, Rivera, M. C., Fettweis, J. M., Serrano, M. G., et al. (2015). The Truth About Metagenomics: Quantifying and Counteracting Bias in 16S rRNA Studies. *BMC Microbiol.* 15, 66. doi: 10.1186/s12866-015-0351-6

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: High-Resolution Sample Inference From Illumina Amplicon Data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869

Chandra, J., Retuerto, M., Mukherjee, P. K., and Ghannoum, M. (2016). "The Fungal Biome of the Oral Cavity", in *Candida Species. Methods Mol. Biol.* Eds. R. Calderone and R. Cihlar (New York, NY: Humana Press), 1356. doi: 10.1007/978-1-4939-3052-4_9

Conche, C., and Greten, F. R. (2018). Fungi Enter the Stage of Colon Carcinogenesis. *Immunity* 49, 384–386. doi: 10.1016/j.immuni.2018.09.002

Cui, L., Morris, A., and Ghedin, E. (2013). The Human Mycobiome in Health and Disease. *Genome Med.* 5, 63. doi: 10.1186/gm467

Davis, N. M., Di, M., Holmes, S. P., Relman, D. A., and Callahan, B. J. (2018). Simple Statistical Identification and Removal of Contaminant Sequences in Marker-Gene and Metagenomics Data. *Microbiome* 6, 226. doi: 10.1186/s40168-018-0605-2

Deschasaux, M., Bouter, K. E., Prodan, A., Levin, E., Groen, A. K., Herrema, H., et al. (2018). Depicting the Composition of Gut Microbiota in a Population With Varied Ethnic Origins But Shared Geography. *Nat. Med.* 24, 1526–1531. doi: 10.1038/s41591-018-0160-1

Deveau, A., Bonito, G., Uehling, J., Paoletti, M., Becker, M., Bindschedler, S., et al. (2018). Bacterial-Fungal Interactions: Ecology, Mechanisms and Challenges. *FEMS Microbiol. Rev.* 42, 335–352. doi: 10.1093/femsre/fuy008

Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C. R., Yu, W. H., et al. (2010). The Human Oral Microbiome. *J. Bacteriol.* 192, 5002–5017. doi: 10.1128/JB.00542-10

Dhariwal, A., Chong, J., Habib, S., King, I. L., Agellon, L. B., and Xia, J. (2017). MicrobiomeAnalyst: A Web-Based Tool for Comprehensive Statistical, Visual and Meta-Analysis of Microbiome Data. *Nucleic Acids Res.* 45, W180–W188. doi: 10.1093/nar/gkx295

Dupuy, A. K., David, M. S., Li, L., Heider, T. N., Peterson, J. D., Montano, E. A., et al. (2014). Redefining the Human Oral Mycobiome With Improved Practices in Amplicon-Based Taxonomy: Discovery of Malassezia as a Prominent Commensal. *PloS One* 9 (3), e90899. doi: 10.1371/journal.pone.0090899

Dwivedi, P. P., Mallya, S., and Dongari-Bagtzoglou, A. (2009). A Novel Immunocompetent Murine Model for Candida Albicans-Promoted Oral Epithelial Dysplasia. *Med. Mycol.* 47, 157–167. doi: 10.1080/13693780802165797

Fernandes, A. D., Macklaim, J. M., Linn, T. G., Reid, G., and Gloor, G. B. (2013). ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. *PLoS One* 8:67019. doi: 10.1371/journal.pone.0067019

Field, E. A., Field, J. K., and Martin, M. V. (1989). Does Candida Have a Role in Oral Epithelial Neoplasia? *Med. Mycol.* 27, 277–294. doi: 10.1080/02681218980000391

Frøslev, T. G., Kjøller, R., Bruun, H. H., Ejrnæs, R., Brunbjerg, A. K., Pietroni, C., et al. (2017). Algorithm for Post-Clustering Curation of DNA Amplicon Data Yields Reliable Biodiversity Estimates. *Nat. Commun.* 8, 1–11. doi: 10.1038/s41467-017-01312-x

Gadelkarim Ahmed, H., and Ahmed, H. G. (2013). Aetiology of Oral Cancer in the Sudan. *J. Oral. Maxillofac. Res.* 4, e3. doi: 10.5037/jomr.2013.4203

Gazdeck, R. K., Fruscione, S. R., Adami, G. R., Zhou, Y., Cooper, L. F., and Schwartz, J. L. (2019). Diversity of the Oral Microbiome Between Dentate and Edentulous Individuals. *Oral. Dis.* 25, 911–918. doi: 10.1111/odi.13039

Ghannoum, M. A., Jurevic, R. J., Mukherjee, P. K., Cui, F., Sikaroodi, M., Naqvi, A., et al. (2010). Characterization of the Oral Fungal Microbiome (Mycobiome) in Healthy Individuals. *PLoS Pathog.* 6, 1000713. doi: 10.1371/journal.ppat.1000713

Goyal, S., Castrillón-Betancur, J. C., Klaile, E., and Slevogt, H. (2018). The Interaction of Human Pathogenic Fungi With C-Type Lectin Receptors. *Front. Immunol.* 9, 1261. doi: 10.3389/fimmu.2018.01261

Hajishengallis, G., and Lamont, R. J. (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. doi: 10.1111/j.2041-1014.2012.00663.x

Hebbar, P. B., Pai, A., and Sujatha, D. (2013). Mycological and Histological Associations of Candida in Oral Mucosal Lesions. *J. Oral. Sci.* 55, 157–160. doi: 10.2334/josnusd.55.157

Hooks, K. B., and O'Malley, M. A. (2017). Dysbiosis and Its Discontents. *MBio* 8, e01492-17 doi: 10.1128/mBio.01492-17

Hooper, S. J., Wilson, M. J., and Crean, S. J. (2009). Exploring the Link Between Microorganisms and Oral Cancer: A Systematic Review of the Literature. *Head Neck* 31, 1228–1239. doi: 10.1002/hed.21140

Huseyin, C. E., Rubio, R. C., O'Sullivan, O., Cotter, P. D., and Scanlan, P. D. (2017). The Fungal Frontier: A Comparative Analysis of Methods Used in the Study of the Human Gut Mycobiome. *Front. Microbiol.* 8, 1432. doi: 10.3389/fmicb.2017.01432

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., et al. (2012). Structure, Function and Diversity of the Healthy Human Microbiome. *Nature* 486, 207–214. doi: 10.1038/nature11234

Idris, A. M., Ahmed, H. M., Mukhtar, B. I., Gadir, A. F., and El-Beshir, E. I. (1995). Descriptive Epidemiology of Oral Neoplasms in Sudan 1970-1985 and the Role of Toombak. *Int. J. Cancer* 61, 155–158. doi: 10.1002/ijc.2910610202

Iliev, I. D., and Leonardi, I. (2017). Fungal Dysbiosis: Immunity and Interactions at Mucosal Barriers. *Nat. Rev. Immunol.* 17, 635–646. doi: 10.1038/nri.2017.55

Jenkinson, H. F., and Lamont, R. J. (2005). Oral Microbial Communities in Sickness and in Health. *Trends Microbiol.* 13, 589–595. doi: 10.1016/j.tim.2005.09.006

Khalifa, N., Allen, P. F., Abu-bakr, N. H., Abdel-Rahman, M. E., and Abdelghafar, K. O. (2012). A Survey of Oral Health in a Sudanese Population. *BMC Oral. Health* 12, 5. doi: 10.1186/1472-6831-12-5

Knot, P. D., Ko, D. L., and Fredricks, D. N. (2009). Sequencing and Analysis of Fungal rRNA Operons for Development of Broad-Range Fungal PCR Assays. *Appl. Environ. Microbiol.* 75, 1559–1565. doi: 10.1128/AEM.02383-08

Koh, H., Livanos, A. E., Blaser, M. J., and Li, H. (2018). A Highly Adaptive Microbiome-Based Association Test for Survival Traits. *BMC Genomics* 19, 210. doi: 10.1186/s12864-018-4599-8

Krogh, P. (1990). The Role of Yeasts in Oral Cancer by Means of Endogenous Nitrosation. *Acta Odontol. Scand.* 48, 85–88. doi: 10.3109/00016359009012738

Krogh, P., Hald, B., and Holmstrup, P. (1987). Possible Mycological Etiology of Oral Mucosal Cancer: Catalytic Potential of Infecting Candida Aibicans and Other Yeasts in Production of N-Nitrosobenzylmethylamine. *Carcinogenesis* 8, 1543–1548. doi: 10.1093/carcin/8.10.1543

Krüger, W., Vielreicher, S., Kapitan, M., Jacobsen, I. D., and Niemiec, M. J. (2019). Fungal-Bacterial Interactions in Health and Disease. *Pathogens* 8, 70. doi: 10.3390/pathogens8020070

Lahti, L., Shetty, S., and Al, E. (2017) *Tools for Microbiome Analysis in R. Microbiome Package Version 1.13.6.* Available at: http://microbiome.github.com/microbiome.

Leigh Greathouse, K., Sinha, R., and Vogtmann, E. (2019). DNA Extraction for Human Microbiome Studies: The Issue of Standardization. *Genome Biol.* 20, 212. doi: 10.1186/s13059-019-1843-8

Lim, Y., Totsika, M., Morrison, M., and Punyadeera, C. (2017). Oral Microbiome: A New Biomarker Reservoir for Oral and Oropharyngeal Cancers. *Theranostics* 7, 4313–4321. doi: 10.7150/thno.21804

Mantovani, A., Allavena, P., Sica, A., and Balkwill, F. (2008). Cancer-Related Inflammation. *Nature* 454, 436–444. doi: 10.1038/nature07205

Masters, N. (2018). *Home | Smoking Pack Years*. Available at: https://www.smokingpackyears.com (Accessed March 6, 2019).

McCullough, M., Jaber, M., Barrett, A. W., Bain, L., Speight, P. M., and Porter, S. R. (2002). Oral Yeast Carriage Correlates With Presence of Oral Epithelial Dysplasia. *Oral. Oncol.* 38, 391–393. doi: 10.1016/S1368-8375(01)00079-3

McMurdie, P. J., and Holmes, S. (2013). Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* 8, e61217. doi: 10.1371/journal.pone.0061217

Mukherjee, P. K., Chandra, J., Retuerto, M., Sikaroodi, M., Brown, R. E., Jurevic, R., et al. (2014). Oral Mycobiome Analysis of HIV-Infected Patients: Identification of Pichia as an Antagonist of Opportunistic Fungi. *PLoS Pathog.* 10, 1003996. doi: 10.1371/journal.ppat.1003996

Mukherjee, P. K., Wang, H., Retuerto, M., Zhang, H., Burkey, B., Ghannoum, M. A., et al. (2017). Bacteriome and Mycobiome Associations in Oral Tongue Cancer. *Oncotarget* 8, 97273–97289. doi: 10.18632/oncotarget.21921

Müller, V., Viemann, D., Schmidt, M., Endres, N., Ludwig, S., Leverkus, M., et al. (2007). Candida Albicans Triggers Activation of Distinct Signaling Pathways to Establish a Proinflammatory Gene Expression Program in Primary Human Endothelial Cells. *J. Immunol.* 179, 8435–8445. doi: 10.4049/jimmunol.179.12.8435

Naglik, J. R., Tang, S. X., and Moyes, D. L. (2013). Oral Colonization of Fungi. *Curr. Fungal Infect. Rep.* 7, 152–159. doi: 10.1007/s12281-013-0129-y

Nasidze, I., Li, J., Quinque, D., Tang, K., and Stoneking, M. (2009). Global Diversity in the Human Salivary Microbiome. *Genome Res.* 19, 636–643. doi: 10.1101/gr.084616.108

Nilsson, R. H., Anslan, S., Bahram, M., Wurzbacher, C., Baldrian, P., and Tedersoo, L. (2019a). Mycobiome Diversity: High-Throughput Sequencing and Identification of Fungi. *Nat. Rev. Microbiol.* 17, 95–109. doi: 10.1038/s41579-018-0116-y

Nilsson, R. H., Larsson, K. H., Taylor, A. F. S., Bengtsson-Palme, J., Jeppesen, T. S., Schigel, D., et al. (2019b). The UNITE Database for Molecular Identification of Fungi: Handling Dark Taxa and Parallel Taxonomic Classifications. *Nucleic Acids Res.* 47, D259–D264. doi: 10.1093/nar/gky1022

Oleim, S. H., Noor, S. K., Bushara, S. O., Ahmed, M. H., and Elmadhoun, W. (2019). The Irrational Use of Antibiotics Among Doctors, Pharmacists and the Public in River Nile State, Sudan. *Sudan. J. Med. Sci.* 14 (4), 276–288. doi: 10.18502/sjms.v14i4.5909

Osman, T. A., Satti, A. A., Bøe, O. E., Yang, Y. H., Ibrahim, S. O., and Suleiman, A. M. (2010). Pattern of Malignant Tumors Registered at a Referral Oral and Maxillofacial Hospital in Sudan During 2006 and 2007. *J. Cancer Res. Ther.* 6, 473–477. doi: 10.4103/0973-1482.77112

Perera, M., Al-hebshi, N. N., Perera, I., Ipe, D., Ulett, G. C., Speicher, D. J., et al. (2017). A Dysbiotic Mycobiome Dominated by Candida Albicans Is Identified Within Oral Squamous-Cell Carcinomas. *J. Oral. Microbiol.* 9, 1385369. doi: 10.1080/20002297.2017.1385369

Peters, B. A., Wu, J., Hayes, R. B., and Ahn, J. (2017). The Oral Fungal Mycobiome: Characteristics and Relation to Periodontitis in a Pilot Study. *BMC Microbiol.* 17, 157. doi: 10.1186/s12866-017-1064-9

Plantinga, A., Zhan, X., Zhao, N., Chen, J., Jenq, R. R., and Wu, M. C. (2017). MiRKAT-S: A Community-Level Test of Association Between the Microbiota and Survival Times. *Microbiome* 5, 17. doi: 10.1186/s40168-017-0239-9

Proctor, L. M., Creasy, H. H., Fettweis, J. M., Lloyd-Price, J., Mahurkar, A., Zhou, W., et al. (2019). The Integrative Human Microbiome Project. *Nature* 569, 641–648. doi: 10.1038/s41586-019-1238-8

Rindum, J. L., Stenderup, A., and Holmstrup, P. (1994). Identification of Candida Albicans Types Related to Healthy and Pathological Oral Mucosa. *J. Oral. Pathol. Med.* 23, 406–412. doi: 10.1111/j.1600-0714.1994.tb00086.x

Rivers, A. R., Weber, K. C., Gardner, T. G., Liu, S., and Armstrong, S. D. (2018). ITSxpress: Software to Rapidly Trim Internally Transcribed Spacer Sequences With Quality Scores for Marker Gene Analysis. *F1000Research* 7, 1418. doi: 10.12688/f1000research.15704.1

Romani, L. (2011). Immunity to Fungal Infections. *Nat. Rev. Immunol.* 11, 275–288. doi: 10.1038/nri2939

Rosenbaum, J., Usyk, M., Chen, Z., Zolnik, C. P., Jones, H. E., Waldron, L., et al. (2019). Evaluation of Oral Cavity DNA Extraction Methods on Bacterial and Fungal Microbiota. *Sci. Rep.* 9, 1531. doi: 10.1038/s41598-018-38049-6

Rosier, B. T., Marsh, P. D., and Mira, A. (2018). Resilience of the Oral Microbiota in Health: Mechanisms That Prevent Dysbiosis. *J. Dent. Res.* 97, 371–380. doi: 10.1177/0022034517742139

Sanjaya, P. R., Gokul, S., Gururaj Patil, B., and Raju, R. (2011). Candida in Oral Pre-Cancer and Oral Cancer. *Med. Hypotheses* 77, 1125–1128. doi: 10.1016/j.mehy.2011.09.018

Seed, P. C. (2015). The Human Mycobiome. *Cold Spring Harb. Perspect. Med.* 5: a019810. doi: 10.1101/cshperspect.a019810

Shaw, L., Ribeiro, A. L. R., Levine, A. P., Pontikos, N., Balloux, F., Segal, A. W., et al. (2017). The Human Salivary Microbiome Is Shaped by Shared Environment Rather Than Genetics: Evidence From a Large Family of Closely Related Individuals. *MBio* 8, e01237-17. doi: 10.1128/mBio.01237-17

Sparber, F., and LeibundGut-Landmann, S. (2017). Host Responses to Malassezia Spp. In the Mammalian Skin. *Front. Immunol.* 8, 1614. doi: 10.3389/fimmu.2017.01614

Tamai, R., Sugamata, M., and Kiyoura, Y. (2011). Candida Albicans Enhances Invasion of Human Gingival Epithelial Cells and Gingival Fibroblasts by Porphyromonas Gingivalis. *Microb. Pathog.* 51, 250–254. doi: 10.1016/j.micpath.2011.06.009

Vasquez, A. A., Ram, J. L., Qazazi, M. S., Sun, J., and Kato, I. (2018). "Oral Microbiome: Potential Link to Systemic Diseases and Oral Cancer," in *Mechanisms Underlying Host-Microbiome Interactions in Pathophysiology of Human Diseases* (Boston, MA: Springer US), 195–246. doi: 10.1007/978-1-4939-7534-1_9

Witherden, E. A., Shoaie, S., Hall, R. A., and Moyes, D. L. (2017). The Human Mucosal Mycobiome and Fungal Community Interactions. *J. Fungi.* 3, 56. doi: 10.3390/jof3040056

Yang, F., Ning, K., Zeng, X., Zhou, Q., Su, X., and Yuan, X. (2016). Characterization of Saliva Microbiota's Functional Feature Based on Metagenomic Sequencing. *Springerplus* 5, 2098. doi: 10.1186/s40064-016-3728-6

Yang, C. Y., Yeh, Y. M., Yu, H. Y., Chin, C. Y., Hsu, C. W., Liu, H., et al. (2018). Oral Microbiota Community Dynamics Associated With Oral Squamous Cell Carcinoma Staging. *Front. Microbiol.* 9, 862. doi: 10.3389/fmicb.2018.00862

Zaura, E., Nicu, E. A., Krom, B. P., and Keijser, B. J. F. (2014). Acquiring and Maintaining a Normal Oral Microbiome: Current Perspective. *Front. Cell. Infect. Microbiol.* 4, 85. doi: 10.3389/fcimb.2014.00085

Zhu, F., Willette-Brown, J., Song, N. Y., Lomada, D., Song, Y., Xue, L., et al. (2017). Autoreactive T Cells and Chronic Fungal Infection Drive Esophageal Carcinogenesis. *Cell Host Microbe* 21, 478–493.e7. doi: 10.1016/j.chom.2017.03.006

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Cellular and Infection Microbiology

BRIEF RESEARCH REPORT
published: 21 October 2021
doi: 10.3389/fcimb.2021.720637

Reprocessing 16S rRNA Gene Amplicon Sequencing Studies: (Meta)Data Issues, Robustness, and Reproducibility

Xiongbin Kang[1,2]*, *Dong Mei Deng*[1], *Wim Crielaard*[1] and *Bernd W. Brandt*[1]*

[1] Department of Preventive Dentistry, Academic Centre for Dentistry Amsterdam (ACTA), University of Amsterdam and Vrije Universiteit Amsterdam, Amsterdam, Netherlands, [2] Genome Data Science, Center for Biotechnology, Faculty of Technology, Bielefeld University, Bielefeld, Germany

OPEN ACCESS

Edited by:
Dominik Heider,
University of Marburg, Germany

Reviewed by:
Daniela Beisser,
University of Duisburg-Essen, Germany
Alfonso Benítez-Páez,
Principe Felipe Research Center (CIPF), Spain

***Correspondence:**
Xiongbin Kang
kang@cebitec.uni-bielefeld.de
Bernd W. Brandt
b.brandt@acta.nl

Specialty section:
This article was submitted to Microbiome in Health and Disease, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: 04 June 2021
Accepted: 20 September 2021
Published: 21 October 2021

Citation:
Kang X, Deng DM, Crielaard W and Brandt BW (2021) Reprocessing 16S rRNA Gene Amplicon Sequencing Studies: (Meta)Data Issues, Robustness, and Reproducibility. Front. Cell. Infect. Microbiol. 11:720637. doi: 10.3389/fcimb.2021.720637

High-throughput sequencing technology provides an efficient method for evaluating microbial ecology. Different bioinformatics pipelines can be used to convert 16S ribosomal RNA gene amplicon sequencing data into an operational taxonomic unit (OTU) table that is used to analyze microbial communities. It is important to assess the robustness of these pipelines, each with specific algorithms and/or parameters, and their influence on the outcome of statistical tests. Articles with publicly available datasets on the oral microbiome were searched for, and five datasets were retrieved. These were from studies on changes in microbiota related to smoking, oral cancer, caries, diabetes, or periodontitis. Next, the data was processed with four pipelines based on VSEARCH, USEARCH, mothur, and UNOISE3. OTU tables were rarefied, and differences in α-diversity and β-diversity were tested for different groups in a dataset. Finally, these results were checked for consistency among these example pipelines. Of articles that deposited data, only 57% made all sequencing and metadata available. When processing the datasets, issues were encountered, caused by read characteristics and differences between tools and their defaults in combination with a lack of detail in the methodology of the articles. In general, the four mainstream pipelines provided similar results, but importantly, P-values sometimes differed between pipelines beyond the significance threshold. Our results indicated that for published articles, the description of bioinformatics methods and data deposition should be improved, and regarding reproducibility, that analysis of multiple subsamples is required when using rarefying as library-size normalization method.

Keywords: 16S rRNA gene sequencing, reprocessing, rarefying, robustness, reproducibility, microbiome

INTRODUCTION

The development of massively parallel sequencing technologies made rapid sequencing of hundreds of samples at unprecedented depth possible (Schuster, 2008; Caporaso et al., 2011). This enabled researchers to apply 16S rRNA gene amplicon sequencing to analyze the composition and dynamics of complex microbial communities in depth (Woo et al., 2008). In the past decade, this has provided insights into diverse microbial communities, ranging from the ocean microbiome (Moran, 2015; Sunagawa et al., 2015; Mestre et al., 2018) or the soil microbiome (Fierer, 2017; Bahram et al., 2018; Delgado-Baquerizo et al., 2021; Xun et al., 2021) to the human microbiome (Turnbaugh et al., 2007; NIH HMP Working Group et al., 2009; Crielaard et al., 2011; Cho and Blaser, 2012; Gilbert et al., 2018).

To date, multiple approaches have been developed to process 16S rRNA gene amplicon sequencing data (Lemos et al., 2017). The most widely used software tools are USEARCH (Edgar, 2010), VSEARCH (Rognes et al., 2016), QIIME (Caporaso et al., 2010) [succeeded by QIIME 2 (Bolyen et al., 2019)], and mothur (Schloss et al., 2009). In addition, interest has grown in high-resolution clustering and error-correction of the sequences provided by tools, such as DADA2 (Callahan et al., 2016) and UNOISE (Edgar, 2016b). During the last years, many other pipelines combining different tools have been developed, such as OCToPUS (Mysara et al., 2017), FROGS (Escudié et al., 2018), PEMA (Zafeiropoulos et al., 2020), AmpliconTagger (Tremblay and Yergeau, 2019), Natrix (Welzel et al., 2020), and the MicrobiomeAnalyst platform (Chong et al., 2020). Conceptually, the processing pipelines are similar and can be divided into several steps: (1) paired-read merging; (2) quality filtering; (3) chimera removal; (4) clustering into operational taxonomic units (OTUs); and (5) taxonomic classification. After construction of the OTU table, researchers proceed to analyze the microbial composition and diversity of the microbial communities and to further interpret biological phenomena, for example, the relationship between obesity and gut microbiota (Komaroff, 2017).

However, algorithms and/or parameters in different processing pipelines often differ. So far, there is no single gold-standard pipeline to produce an OTU table (or higher-resolution count table), which means that both different tools and different parameters for the same step are being used in different pipelines.

Many existing processing steps have been evaluated, such as the influence of chimera checking methods (Edgar, 2016a; Mysara et al., 2017), denoising methods (Bonder et al., 2012; May et al., 2014), and clustering methods on the OTU table (Bonder et al., 2012; May et al., 2014; Westcott and Schloss, 2015; Mysara et al., 2017; Westcott and Schloss, 2017). Another study has assessed robustness and reproducibility of clustering methods on OTUs, while varying clustering thresholds (Schmidt et al., 2015). In addition, entire clustering or denoising pipelines have also been compared (Westcott and Schloss, 2015; Mysara et al., 2017; Nearing et al., 2018; Tremblay and Yergeau, 2019; Prodan et al., 2020). Several of these studies have shown in detail that both the number and composition of OTUs, from the same dataset, depend on the selected methods.

Here, we focused on the robustness of "final" results, which means a conclusion drawn from the same sequencing data is concordant among different processing pipelines [*cf.* (Schloss, 2018)]. We aimed to evaluate this using several published 16S rRNA gene amplicon studies and different mainstream pipelines. We are specifically not evaluating differences in the OTU tables themselves. We, and several others, have done that in the past and refer the reader interested in that to the articles cited above. While different pipelines likely result in different OTU tables due to their distinct algorithms and parameters, (biological) conclusions should rather not change. For example, if microbial profiles differ (significantly) between cases and controls, this should rather not depend on the pipeline. Thus, the aim is to look into statistical conclusions based on the analyses of the microbial profiles originating from several pipelines run on the same dataset.

To this end, four different pipelines based on VSEARCH, USEARCH, mothur, and UNOISE3, which are extensively used for 16S rRNA gene sequence data processing, were implemented; and publicly available datasets were retrieved and processed with these pipelines. Our aim is not to perform an exhaustive comparison of available pipelines. VSEARCH (Rognes et al., 2016) can be seen as an open-source reimplementation of USEARCH (Edgar, 2010). Since VSEARCH is used as a replacement for USEARCH, both tools were included as to see how their differences affect the final outcome. In addition, mothur (Schloss et al., 2009) was chosen as an often-used pipeline with an excellent SOP. Finally, UNOISE3 was selected as an example of a denoising method. It was found that UNOISE3 "showed the best balance between resolution and specificity" (Prodan et al., 2020).

Using the resulting OTU tables, differences in microbial α-diversity and β-diversity between groups within a study were evaluated and the results (P-values) compared among the pipelines, using *exactly* the same dataset. Since random subsampling is often used, we also evaluated reproducibility of results: a collection of subsampled OTU tables was generated as to compare the distribution of P-values within and between the pipelines. P-values are used here to illustrate differences among pipelines and should not be (mis)used to conclude about scientific importance (Baker, 2016; Wasserstein et al., 2019).

MATERIALS AND METHODS

Dataset Search

Articles on the oral microbiome were searched for, and their respective datasets were retrieved. To limit the influence of the 16S rRNA region, this study only searched for datasets using the V4 16S rRNA region, published during the past 5 years (Illumina MiSeq sequencing). Both sequencing and metadata had to be publicly available. Initially, articles with deposited datasets were searched for using the NCBI website as this hosts both PubMed and the Sequence Read Archive (SRA). PubMed search results

were linked to SRA using LinkOut (not possible anymore in the new PubMed). However, many articles that deposited data in the SRA, with article title and DOI, were lost in this process due to incomplete linking between these databases. Therefore, studies were searched for using Google Scholar with the following query (February 9, 2019): intitle:oral 16S "V4 region" OR "V4 variable region" OR "V4 hypervariable region" "accession OR SRA." The final papers were screened on reported P-values for comparisons: at least one test on the microbiome data had to report a P-value between 0.0001 and 0.05. Finally, sequencing data and metadata were downloaded from the NCBI.

Pipelines

Four different processing pipelines were built to produce OTUs tables: a mothur pipeline [version 1.41.3], a VSEARCH [version 2.11.0-linux-x86_64], a USEARCH [version 11.0.667_i86linux32], and a UNOISE3 [version 11.0.667_i86linux32] pipeline. **Figure 1** presents an overview of the four pipelines, and **Supplementary Table 1** lists their details. In general, each pipeline used the standard commands with either default or otherwise well-accepted parameters. For mothur, we followed the MiSeq Standard Operation Procedure (https://www.mothur.org/wiki/MiSeq_SOP, d.d. 2019-01-24). We only changed the value of maxlength in screen.seqs from 275 to 258 as the V4 region has a small length variation and as to use the same value in all four pipelines. In the VSEARCH, USEARCH, and UNOISE3 pipelines, the reads were merged and quality-filtered per sample and then combined into one file. In the (32-bit) USEARCH/UNOISE3 pipelines, (64-bit) VSEARCH was used to dereplicate these quality-filtered sequences. Since the read lengths in the different studies differed (250 nt, but 150 nt in dataset 4 only), during merging a maximum of 10% mismatches in the overlap region was used.

Analysis of OTU Tables

Statistical analyses were conducted with R [version 3.5.1, (R Core Team, 2018)] and the R packages microbiome [version 1.4.2, (Lahti and Shetty, 2017)], phyloseq [version 1.26.0, (McMurdie and Holmes, 2013)], and vegan [version 2.5-4, (Oksanen et al., 2019)]. The Mann-Whitney test was applied to test for differences in α-diversity (Shannon diversity index) between two different sample types, while differences in β-diversity were assessed using PERMANOVA (adonis, Bray-Curtis distance, 9999 permutations). Spearman's rank correlation coefficient was used to correlate the Shannon diversity index between different pipelines. To evaluate the similarity between OTU tables (mothur only), a Procrustes Analysis and Mantel test were conducted with QIIME v1.9.1 (Caporaso et al., 2010) using the Bray-Curtis distance and 999 permutations.

Random subsampling was used to normalize unequal sample depth (library size). The subsampling depth for each dataset was determined such that most samples remained in the analysis, while adhering to minimum of around 2,000 reads/sample. In addition, as sample depths depend slightly on the pipeline, the subsampling depth was chosen such that the OTU tables from the different pipelines contained the same samples. To assess the

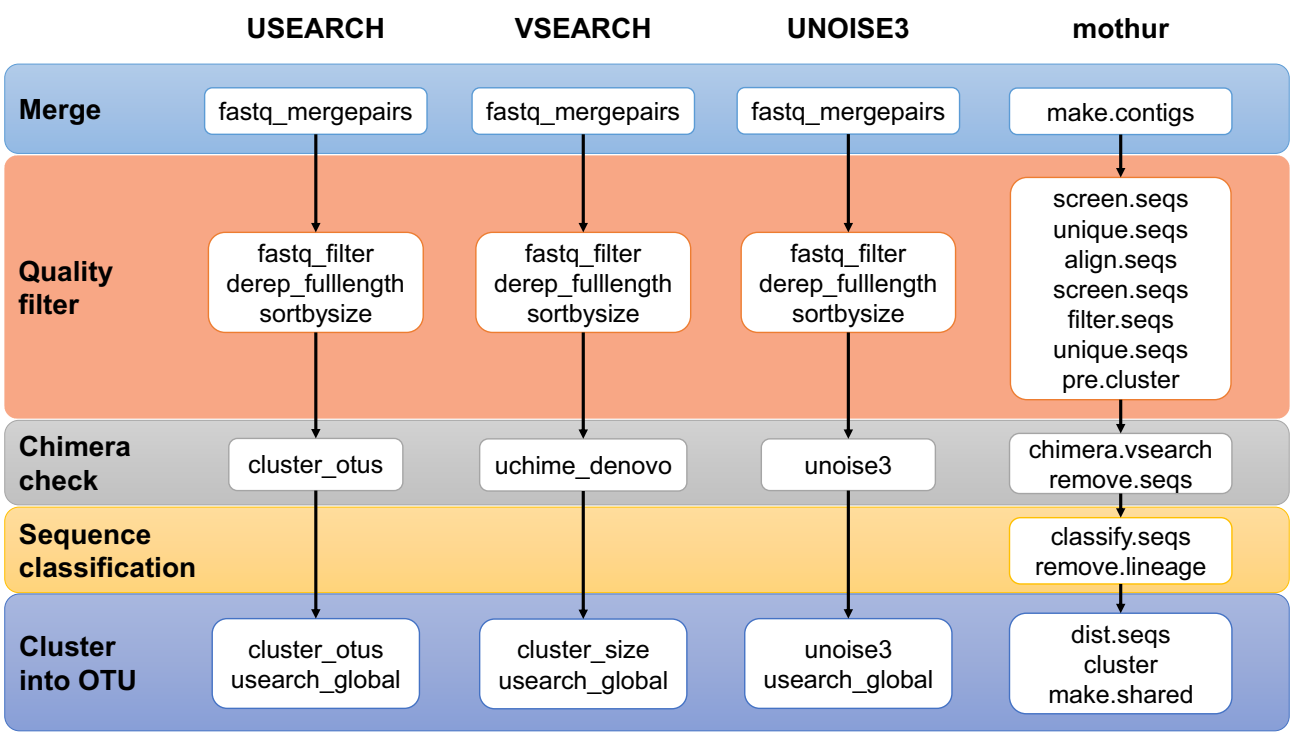


FIGURE 1 | Overview of the four pipelines compared in this study. See **Supplementary Table 1** for details.

reproducibility of statistical tests, 1,000 random subsamples of the same OTU table were analyzed.

RESULTS

Different publicly available datasets on the oral niche were searched for and processed with the VSEARCH, USEARCH, mothur, and UNOISE3 pipelines. During processing, different issues were encountered with specific datasets and pipelines that had to be addressed first. Next, the influence of the pipelines on diversity comparisons and reproducibility of results were evaluated.

Dataset Search

The literature search returned 60 articles of which, upon inspection, many did not satisfy our criteria (see *Dataset Search* in *Materials and Methods*). Out of 53 articles that included an accession number to, for example, NCBI's SRA or the European Nucleotide Archive, 45 studies actually deposited the raw data, while only 30 included the metadata in the database or in the article. Finally, 11 studies remained that used the V4 16S rRNA region and were related to the oral niche (19 studies were excluded: 14 studies used the V3-V4 region, 1 study used the V1-V2 region, 1 study the V1-V3 region, 3 studies on gut only). Based on screening with the P-value criterion, five oral microbiome studies were selected from these 11 studies. This criterion was used to restrict our analyses, since results would unlikely differ for more extreme P-values.

These datasets passing all criteria were the following. Dataset 1 (Stewart et al., 2018) was a study on the effects of tobacco smoke and electronic cigarette vapor exposure on the oral and gut microbiota. Dataset 2 (Schmidt et al., 2014) was on the relation between oral cancer and oral microbiota, and dataset 3 (Gomez et al., 2017) on the influence of host genetics on caries using monozygotic and dizygotic twins. Dataset 4 (Xiao et al., 2017) studied the impact of diabetes on the oral microbiota using mice, while dataset 5 (Chen et al., 2018) investigated the effects of periodontitis and its treatment on oral microbiota. **Table 1** shows an overview of these datasets. The raw read lengths were 250 nt, and, for dataset 4 only, 150 nt.

Data Processing

Although it seemed straightforward to process the retrieved sequence data with one of the pipelines, several unexpected issues were encountered that were related to the sequence data in combination with a specific pipeline. The read pairs of dataset 2 could not be merged by VSEARCH and dataset 4 lost 57% sequences in the mothur pipeline. In addition, mothur could not cluster the sequences of datasets 3 and 4 into OTUs on our compute nodes (64 Gb RAM, 16 core CPU: E5-2650 v2 2.60 GHz) within the imposed time limit of 120 h. Finally, the deposited data of dataset 5 consisted of already merged read pairs.

Therefore, the processing of these datasets was slightly altered to address these issues. In dataset 2, the tail of the reverse reads contained approximately 100 bp low-quality bases (Q <= 2) preventing the read pairs to be merged by VSEARCH. However, USEARCH merged these reads, because the used version automatically trims these low-quality tails (Q <= 2) before merging (Q <= 2, min. length 64 nt). Therefore, we pre-filtered dataset 2 using Trimmomatic v.0.38 (Bolger et al., 2014) with "TAIL:3 MINLEN:64" and used this filtered data as input for all pipelines.

TABLE 1 | Overview of the five used datasets.

Dataset	Accession	Study size	Selected depth	Species	Sample type	Sample types
Dataset 1 (Stewart et al., 2018)	PRJNA413706	90	9,500	Human	Saliva, buccal swabs, feces	Electronic cigarette users, tobacco smokers, and matched controls
Dataset 2 (Schmidt et al., 2014)	PRJEB4953	83	22,000	Human	Buccal swabs	Oral cancer, precancer, and healthy controls
Dataset 3 (Gomez et al., 2017)	PRJNA383868*	484	2,800	Human	Plaque	Twins, healthy or with enamel or dentin caries
Dataset 4 (Xiao et al., 2017)	SRP108800	81	1,900	Mouse	Saliva, feces	Normoglycemic, diabetic, and diabetic IL-17A antibody-treated mice
Dataset 5 (Chen et al., 2018)	SRP075100	238	7,900	Human	Saliva, plaque	Chronic periodontitis patients and periodontally healthy adults

*Dataset 1 consisted of 90 samples from 30 participants: 10 tobacco smokers (TS), 10 electronic cigarette (EC) users, and 10 non-smoking controls. Fecal, saliva, and buccal swab samples were collected from each individual. Dataset 2 contained 83 samples divided over three groups: oral cancer (Cancer, n=21; Contralateral normal, n=19), precancer (Precancer, n=13; Contralateral normal, n=11), and healthy (lateral tongue, n=9; Floor of mouth, n=10) persons. For dataset 3, metadata included 485 dental plaque samples (484 twins and 1 singleton; dizygotic n=280; monozygotic n=205), while this singleton (1061.1_RD1) was not present in the SRA (*accessions: SRR5467515–SRR5467785 and SRR5467788–SRR5468062). In addition, eight samples did not include zygotic information. Finally, 271 dizygotic (DZ) and 205 monozygotic (MZ) samples remain. Samples from MZ/DZ twins were compared according to caries status: without dental caries (Health) or enamel/dentin caries, or treated caries. Dataset 4 contained 81 samples (45 oral swab samples and 36 fecal samples, which should be oral swab samples). Normoglycemic (Pre–Diab NG) and diabetes-prone mice (Pre–Diab DB) before and after (Diab NG; Diab DB) the development of hyperglycemia were sampled. In the oral swab samples, Normal refers to mice that received oral bacterial from normoglycemic mice, Diabetic to mice that received oral bacterial from diabetic mice, and Diabetic + IL17 to the mice treated with IL-17A antibodies and oral bacterial from diabetic mice. Dataset 5 comprised 238 samples collected from periodontal healthy individuals and chronic periodontitis patients: D1P (diseased/pre-treatment plaque n=96), D2P (diseased/post-treatment plaque n=19), HP (healthy plaque n=42), D1S (diseased/pre-treatment saliva n=45), D2S (diseased/post-treatment saliva n=18), and HS (healthy saliva n=18).*

From dataset 4, many sequences were removed after the mothur screen.seqs command on the aligned sequences, in which the sequences are required to span at least the V4 region (from 1968 to 11550) in the alignment. Manual inspection showed that many sequences ended one position early and that the first base call after the V4 (806R) reverse primer was absent. Therefore, for dataset 4 only, the value of the end parameter in this screen.seqs command was changed from 11,550 to 11,549 to avoid losing 57% of the sequence data.

In addition, both datasets 3 and 4 contained many singletons. This caused the OTU clustering to fail in mothur. Therefore, singletons were removed from datasets 3 and 4 in the mothur pipeline (split.abund, cutoff=1). For dataset 3, it was also possible to generate an OTU table with cluster.split (taxlevel=2, cutoff=0.03). To evaluate the difference between these two OTU tables (i.e., from cluster.split or singletons removed), they were compared. Spearman's correlation of the Shannon diversities (R = 0.9895, P-value < 2.2e-16), Procrustes Analysis (M^2 = 0.01; p < 0.001), and the Mantel test (r = 0.98837, P-value = 0.001) showed that the OTU tables were very similar. Since dataset 4 could not be processed with cluster.split within the wall-time limit of 120 h, no comparison could be made and the dataset with singletons removed was used.

After the modifications described above, all five datasets were processed with all four pipelines. The total numbers of raw, merged, quality-filtered reads, and reads mapped to the OTU table are summarized in **Supplementary Table 2**. For dataset 4, the OTU table from mothur contained only 69% of sequences of the table from the other pipelines. This turned out to be caused by the removal of non-bacterial sequences (chloroplast, mitochondria, unknown, archaea, eukaryota) in the SOP mothur pipeline. Since only dataset 4 contained many non-bacterial sequences, for all pipelines applied to dataset 4, OTUs classified as non-bacterial were removed as to make a fair comparison.

Robustness of Results

For each of the datasets, OTU tables were generated by the different pipelines. The fraction of quality-filtered mapped reads represented in the OTU table was similar (datasets 1, 2, 3, 5 combined: average 0.95, standard deviation 0.023; dataset 4: average 0.60, standard deviation 0.0018; **Supplementary Table 2**). However, within a dataset, the number of OTUs depended on the pipeline, where the mothur pipeline generated most OTUs. Next, OTU tables were rarefied to avoid the influence of sample depth differences within one dataset (**Table 1**), and the general similarity among these tables from the pipelines was compared using Spearman's rank correlation of the Shannon diversity index. All correlations were high, ranging from 0.94 to 1.0 (P-values < 2.2e-16).

To evaluate the robustness of α-diversity results, the Shannon diversity indices of two different sample types present in the dataset (the original study) were compared for the different pipelines (Mann-Whitney test; single subsampled OTU table). A heatmap (**Supplementary Figure 1A**) shows the resulting P-values, most of which were similar to the original results. Since conclusions, thus biological inferences, are more likely to depend on data processing details when P-values are closer to the significance threshold, we zoomed in on the eight comparisons that had a P-value below 0.05.

Using the significance threshold of 0.05, five comparisons resulted in identical biological conclusions, while there were three conflicts between the four pipelines (**Figure 2A**). Recently, studies proposed to lower the significance threshold to 0.005, which would "immediately improve the reproducibility of scientific research" (Benjamin et al., 2018; Ioannidis, 2018). When the significance threshold was lowered to 0.005, one conflict remained.

Similarly, as to assess the robustness of between-group differences, the microbial profiles of the two groups of sample types were subjected to PERMANOVA (Bray-Curtis distance; **Supplementary Figure 1B**). In most cases, the P-values were similar among the different pipelines and to the original results. Of the 28 comparisons (**Supplementary Figure 1B**), 17 groups had P-values below 0.05 (**Figure 2B**). Similar to the α-diversity tests, lowering the significance threshold improved robustness. However, at any significance threshold, differences between pipelines, here on the *same* data, can appear (**Figure 3**).

In some cases, published results differed from ours, which can also be related to a different distance metric used (datasets 1 and 2 did not use Bray-Curtis). As an example, we take dataset 1, which was processed by the authors using USEARCH (Stewart et al., 2018). In our study, the P-values for the fecal microbiota of controls (Con) *versus* electronic cigarettes (EC) users slightly depended on the pipeline (P-value range: 0.03–0.07). However, the much higher P-value reported by the authors was related to the weighted UniFrac distance metric. Indeed, when using the OTU table provided by authors, all PERMANOVA (Bray-Curtis) results became very similar (**Supplementary Figure 1B**).

Reproducibility of Results

This study also evaluated the reproducibility, defined here as "re-analysis with exactly the same pipeline and same dataset supports an identical conclusion." To this end, each OTU table was subsampled 1,000 times, and statistical tests were done as above, for each of the 1,000 tables, thus providing 1,000 P-values (boxplots in **Figure 3**). Since the P-value ranges for a given pipeline can cross a significance threshold (e.g., **Figure 3-1B**) or can be large (**Figure 3-2D**), care should be taken with reporting results (publication bias).

Subsampling datasets with a large standard deviation in sample depths can lead to a larger variation in test results. For example, within dataset 4, the P-value distribution for UNOISE of the first comparison (**Figure 3-2D**; Diab NG *vs* DB; range: 0.0001–0.0012) differed from other three (e.g., Oral swab normal *vs* Diabetic+IL17: 0.001–0.004). In the first comparison, the median depths of the groups differed a lot (7,666 *vs* 52,512); in the latter, they were much closer (11,387 *vs* 11,572). Thus, when subsampling, results can show more variation since random subsamples vary more when subsample depth is low compared to the sample depth and/or when there is a bias in sample depth between groups. However, this argument does not hold for, for example, dataset 5: D2P *vs* HP and D1P *vs* HP. Here, all three sample groups have very similar medians. However, variability

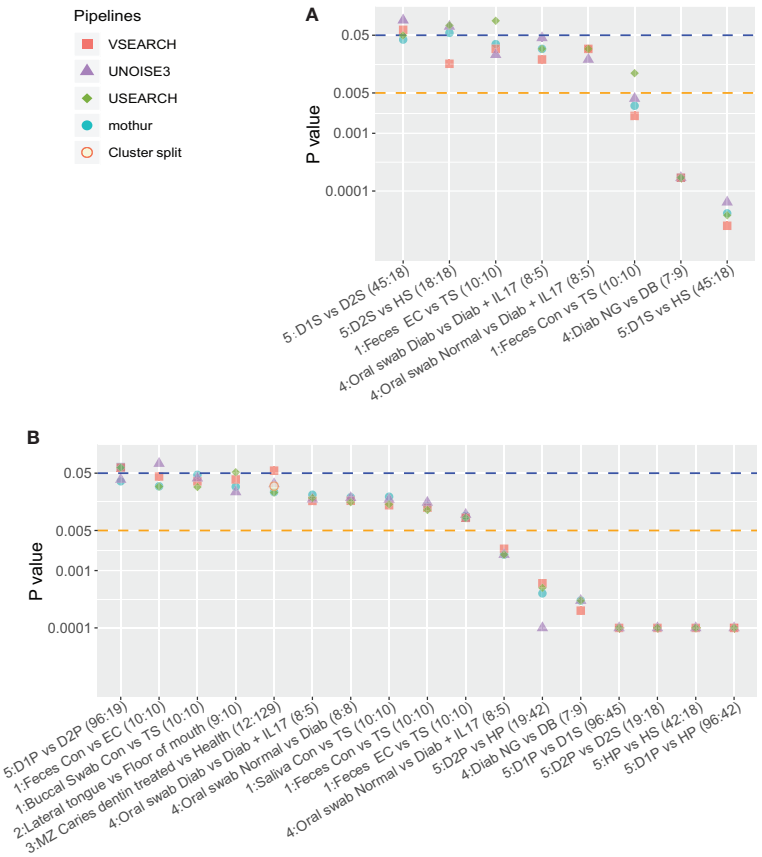


FIGURE 2 | Overview of P-values of the five datasets sorted on decreasing average P-value. The first number before the colon indicates the dataset. Cluster.split is an alternative mothur pipeline used only for dataset 3. **(A)** P-values of Mann-Whitney tests on the Shannon diversity index (α-diversity) and **(B)** P-values of PERMANOVA (Bray-Curtis distance, β-diversity) tests between two sample types. In **(A)**, at a threshold of 0.05, once VSEARCH differed from the other pipelines, once USEARCH, and once mothur and USEARCH differed from VSEARCH and UNOISE3. At a threshold of 0.005, there was one conflict (USEARCH). In **(B)** at 0.05, there were four conflicts: once mothur and UNOISE3 were the same, but differed from USEARCH and VSEARCH, once UNOISE3, once USEARCH, once VSEARCH.

can also be caused by biological differences as well as sample size differences (D2P: 19, D1P: 96, HP: 42 samples). As to exclude biological and other differences between samples, dataset 1 was subsampled at a lower depth to illustrate the increased variability using the same data (**Supplementary Figure 2**). Not surprisingly, a lower subsampling depth results in higher variability of test results.

DISCUSSION

It is difficult to make research sufficiently transparent and reproducible, especially in interdisciplinary fields such as microbiome studies (Schloss, 2018). In this study, we evaluated the robustness and reproducibility of 16S rRNA gene amplicon studies using four mainstream pipelines.

It was not straightforward to reprocess or reproduce results of these studies. During our literature search, we encountered many articles with no or incomplete data availability, even though an accession number was provided: only 57% provided sequencing data and metadata. In addition, while correct and complete descriptions of methods and metadata are crucial, they are often not provided in sufficient detail. Although unclear descriptions of processing methods were not such an issue in this work, since we used our own pipelines, phrases like "reads were quality-filtered" or "clustered using UCLUST" are much too imprecise.

Due to (implicit) differences between tools used for the pipelines, we sometimes had to adapt a pipeline to the data at hand (see *Data Processing* in *Results*). For example, in dataset 4, about 35% of the sequences was taxonomically classified as chloroplast (40% as non-bacterial). However, in the corresponding article (QIIME 1 was used), we did not explicitly find that these sequences were removed, although that seemed to be the case (**Supplementary Figure 3**). Clearly, each dataset requires specific steps, also with respect to quality filtering, and it is important to be aware of differences among tools (even related ones as USEARCH and VSEARCH, or different versions of the same tool).

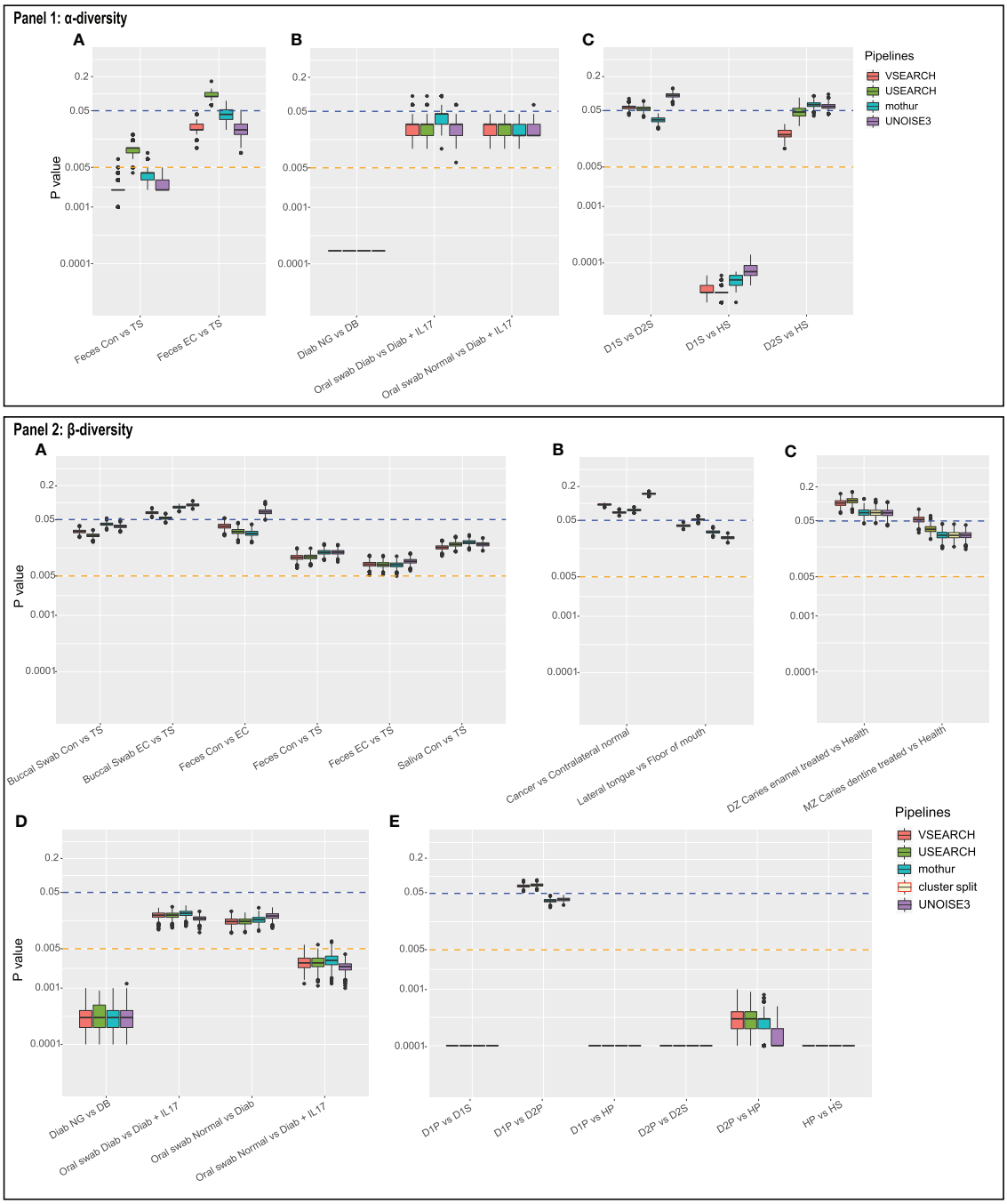


FIGURE 3 | Panel 1 shows the distribution of P-values of Mann-Whitney tests on the Shannon diversity index between the indicated two sample types for 1,000 random subsamples in **(A)** dataset 1, **(B)** dataset 4, and **(C)** dataset 5. Panel 2 shows the distribution of P-values of PERMANOVA (Bray-Curtis distance) tests for 1,000 random subsamples in datasets 1 to 5 **(A–E)**. Cluster.split is an alternative mothur pipeline used only for dataset 3 **(C)**.

The pipelines resulted in a different number of OTUs, which is not surprising. Nearing et al. (2018) reported that several denoising pipelines largely influenced α-diversity (observed OTUs) and possibly impact results based on α-diversity, while the weighted β-diversity metrics (Bray-Curtis, weighted UniFrac) were very similar among different pipelines. When comparing the results of tests on diversity, i.e., the distribution of P-values between pipelines and within a pipeline (**Figures 2**, **3**), tests on α-diversity (Shannon) seem to show a larger variation than on β-diversity (Bray-Curtis, PERMANOVA).

Irrespective of the above, some differences related to tests on α-diversity were initially unexpected, such as between USEARCH and VSEARCH (e.g., Shannon diversity in **Figure 2A**, datasets 1 and 5). Since VSEARCH can be seen as an open-source USEARCH, it was hypothesized that this difference was caused mainly by the different method of chimera checking in these pipelines: USEARCH performs this during clustering, while with VSEARCH this was done before clustering (uchime_denovo). To analyze this, dataset 1 was processed with a VSEARCH pipeline in which the chimera-checking method was replaced by USEARCH (cluster_otus). Indeed, now the test results were more similar to those of USEARCH (**Supplementary Figure 4**). Thus, in this case, the results for Shannon diversity seem to be sensitive to chimera-detection methods. According to a previous study, different chimera-detection methods influenced the accuracy of clustering (May et al., 2014). The result of this study further demonstrated that differences in chimera-checking methods also affected robustness. For these datasets, DADA2, which has a different chimera-checking method, can also show differences due to false positive chimeras (Edgar, 2016b).

The ranges of P-values, based on the 1,000 subsampled OTU tables, sometimes exceeded a significance threshold. This showed that when OTU tables are rarefied, reproducibility can be affected. A P-value of 0.06 does not really differ from 0.04 [*cf.* (Halsey et al., 2015)], and larger differences occur using exactly the same data (**Figure 3**). At lower subsampling depth, with respect to the median sample depth of a group, and/or when depths have large standard deviation, reproducibility can decrease. Especially in such cases, given rarefying is the chosen normalization method, multiple randomly subsampled OTU tables should be evaluated, and the median P-value be used.

Here, rarefying, which is still very often used, was applied to normalize library size. The comparison of normalization methods was beyond the scope of this study, but we note that different methods are available [proportion, CSS, log-ratio, TMM, *cf.* Weiss et al. (2017)]. While McMurdie and Holmes (2014) stated rarefying should not be used to detect differentially abundant species and better be generally avoided, Weiss et al. (2017) later reported that rarefying itself seemed not to increase false discovery rates of many differential abundance-testing methods, and even lowered the false discovery rate when the average library size for groups differed a lot (~10×). While it is not straightforward which normalization method should best be used, even though data normalization methods now receive ample attention, we should not forget "subsampling" occurs several times during experimental procedures, ranging from biological sampling, dilution of DNA for amplicon PCR, to generating the equimolar mix for sequencing.

Irrespective of the normalization technique, care should be taken with PERMANOVA. As stated with its introduction (Anderson, 2001), calculating all possible permutations usually is unrealistic, considering computational time. However, increasing the number of permutations improves the precision of the P-value (Anderson, 2001). With a lower number of permutations (e.g., 999 instead of 9,999), the range of P-values (using same OTU table) increases, which can affect reproducibility. This then shows that the permutation space is too undersampled and the number of permutations should be increased (*cf.* page 37 in Anderson, 2001). Thus, PERMANOVA should be repeated as to check the P-value varies little.

This study did not evaluate the differences caused by the use of different diversity indices (e.g., species richness, Chao1 richness, Shannon index) or distance metrics [(weighted) UniFrac, Bray-Curtis, Jaccard], since these are different downstream choices. When evaluating results from published studies, we should remember that different α-diversity indices or β-diversity metrics can lead to different conclusions. However, here, the focus was on whether different conclusions would result from different amplicon processing pipelines.

Although QIIME 1 also was often used, it has not been supported since 2018, and we did not include it as to keep comparisons concise. In addition, based on previous studies, the default QIIME 1 pipeline has higher error rates due to chimeras and higher amount of spurious OTUs comparing with others (Mysara et al., 2017; Prodan et al., 2020). We, therefore, only compared VSEARCH, USEARCH, mothur, and UNOISE3 in this article as example pipelines, to limit variations and maintain focus, but note that QIIME 2 also supports a VSEARCH pipeline.

In general, results of the four pipelines were robust and reproducible, with some conflicts around the 0.05 threshold (**Figure 2**). The choice of 0.05 as P-value threshold was arbitrary, and it was proposed to lower the P-value threshold to 0.005 to "improve reproducibility of scientific research" among studies (Benjamin et al., 2018; Ioannidis, 2018). However, a different, sometimes related, pipeline for the *same* dataset (study) resulted in different P-values. Although we cannot conclude that a lower threshold should be used, we should keep in mind that P-values just below 0.05 may not be very robust or reproducible, and a lower threshold also comes at a cost (Di Leo and Sardanelli, 2020). Irrespective of the used thresholds, we recommend that real P-values are always reported (not as: "P<0.05").

In our limited exploratory analysis, we did not find that clustering methods consistently differed to the denoising method. With the introduction of UNOISE, Robert Edgar stated, "I suggest you try both. If a biological conclusion is different, then you should worry that neither result is trustworthy" (Edgar, 2019). Yet, it is important to realize that using the same sequencing data, (1) results among pipelines can differ; (2) it will often not be straightforward to uncover why specific differences occur; (3) generally a single pipeline is used, so differences will remain unnoticed. In addition, other measures than P-values can be considered as these show large sample-to-sample variability and have other issues (Halsey et al., 2015; Wasserstein et al., 2019). Nevertheless, a discussion on the use P-values is beyond the scope of this article, and there is no consensus this subject (Halsey et al., 2015; Ioannidis, 2018; Ioannidis, 2019; Wasserstein et al., 2019; Di Leo and Sardanelli, 2020).

In summary, we conclude the following: Sequencing data and metadata should be properly deposited and journals should

check if data have actually been made publicly available. Not surprisingly, different pipelines can lead to different statistical conclusions; thus, methods should be described in detail and include software versions, algorithms, and parameters used. While "the only direct protection [to the threat of selection bias] must come from standards for reproducible research (Ioannidis, 2019)", microbiome research and its data processing highly depend on wet- and dry-lab technology, and even if standards would exist, they would repeatedly (need to) change (Amaral and Neves, 2021). This means that more care should be taken to share methods and (raw) data.

DATA AVAILABILITY STATEMENT

The datasets analyzed in this study can be found in the Sequence Read Archive (NCBI) or the European Nucleotide Archive (EBI) as indicated in **Table 1** or in their cited publications.

AUTHOR CONTRIBUTIONS

BB designed the research and assisted XK in the study design, data collection, and bioinformatics analysis. XK, DD, and BB wrote the draft. DD, WC, and BB reviewed the manuscript. All authors contributed in the preparation and finalization of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

XK is supported by the China Scholarship Council (CSC), grant number 201808440303.

ACKNOWLEDGMENTS

We would like to thank the authors of the papers used in this research for depositing their sequencing and metadata and replying to our queries on their data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2021.720637/full#supplementary-material

REFERENCES

Amaral, O. B., and Neves, K. (2021). Reproducibility: Expect Less of the Scientific Paper. *Nature* 597, 329–331. doi: 10.1038/d41586-021-02486-7

Anderson, M. J. (2001). A New Method for Non-Parametric Multivariate Analysis of Variance. *Austral. Ecol.* 26, 32–46. doi: 10.1111/j.1442-9993.2001.01070.pp.x

Bahram, M., Hildebrand, F., Forslund, S. K., Anderson, J. L., Soudzilovskaia, N. A., Bodegom, P. M., et al. (2018). Structure and Function of the Global Topsoil Microbiome. *Nature* 560, 233–237. doi: 10.1038/s41586-018-0386-6

Baker, M. (2016). Statisticians Issue Warning Over Misuse of P Values. *Nature* 531, 151. doi: 10.1038/nature.2016.19503

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., et al. (2018). Redefine Statistical Significance. *Nat. Hum. Behav.* 2, 6–10. doi: 10.1038/s41562-017-0189-z

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, Interactive, Scalable and Extensible Microbiome Data Science Using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi: 10.1038/s41587-019-0209-9

Bonder, M. J., Abeln, S., Zaura, E., and Brandt, B. W. (2012). Comparing Clustering and Pre-Processing in Taxonomy Analysis. *Bioinformatics* 28, 2891–2897. doi: 10.1093/bioinformatics/bts552

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: High-Resolution Sample Inference From Illumina Amplicon Data. *Nat. Methods* 13, 581–583. doi: 10.1038/nMeth.3869

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME Allows Analysis of High-Throughput Community Sequencing Data. *Nat. Methods* 7, 335–336. doi: 10.1038/NMETH.F.303

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., et al. (2011). Global Patterns of 16S rRNA Diversity at a Depth of Millions of Sequences Per Sample. *Proc. Natl. Acad. Sci. U. S. A.* 108 (Suppl 1), 4516–4522. doi: 10.1073/pnas.1000080107

Chen, C., Hemme, C., Beleno, J., Shi, Z. J., Ning, D., Qin, Y., et al. (2018). Oral Microbiota of Periodontal Health and Disease and Their Changes After Nonsurgical Periodontal Therapy. *ISME J.* 12, 1210–1224. doi: 10.1038/s41396-017-0037-1

Cho, I., and Blaser, M. J. (2012). The Human Microbiome: At the Interface of Health and Disease. *Nat. Rev. Genet.* 13, 260–270. doi: 10.1038/nrg3182

Chong, J., Liu, P., Zhou, G., and Xia, J. (2020). Using MicrobiomeAnalyst for Comprehensive Statistical, Functional, and Meta-Analysis of Microbiome Data. *Nat. Protoc.* 15, 799–821. doi: 10.1038/s41596-019-0264-1

Crielaard, W., Zaura, E., Schuller, A. A., Huse, S. M., Montijn, R. C., and Keijser, B. J. F. (2011). Exploring the Oral Microbiota of Children at Various Developmental Stages of Their Dentition in the Relation to Their Oral Health. *BMC Med. Genomics* 4, 22. doi: 10.1186/1755-8794-4-22

Delgado-Baquerizo, M., Eldridge, D. J., Liu, Y.-R., Sokoya, B., Wang, J.-T., Hu, H.-W., et al. (2021). Global Homogenization of the Structure and Function in the Soil Microbiome of Urban Greenspaces. *Sci. Adv.* 7, eabg5809. doi: 10.1126/sciadv.abg5809

Di Leo, G., and Sardanelli, F. (2020). Statistical Significance: P Value, 0.05 Threshold, and Applications to Radiomics-Reasons for a Conservative Approach. *Eur. Radiol. Exp.* 4, 18. doi: 10.1186/s41747-020-0145-y

Edgar, R. C. (2010). Search and Clustering Orders of Magnitude Faster Than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Edgar, R. C. (2016a). UCHIME2: Improved Chimera Prediction for Amplicon Sequencing. *BioRxiv* 074252. doi: 10.1101/074252

Edgar, R. C. (2016b). UNOISE2: Improved Error-Correction for Illumina 16S and ITS Amplicon Sequencing. *BioRxiv* 081257. doi: 10.1101/081257

Edgar, R. C. (2019) *FAQ: Should You Use UPARSE or UNOISE*. Available at: https://drive5.com/usearch/manual/faq_uparse_or_unoise.html (Accessed 2021-08-02).

Escudié, F., Auer, L., Bernard, M., Mariadassou, M., Cauquil, L., Vidal, K., et al. (2018). FROGS: Find, Rapidly, OTUs With Galaxy Solution. *Bioinformatics* 34, 1287–1294. doi: 10.1093/bioinformatics/btx791

Fierer, N. (2017). Embracing the Unknown: Disentangling the Complexities of the Soil Microbiome. *Nat. Rev. Microbiol.* 15, 579–590. doi: 10.1038/nrmicro.2017.87

Gilbert, J. A., Blaser, M. J., Caporaso, J. G., Jansson, J. K., Lynch, S. V., and Knight, R. (2018). Current Understanding of the Human Microbiome. *Nat. Med.* 24, 392. doi: 10.1038/nm.4517

Gomez, A., Espinoza, J. L., Harkins, D. M., Leong, P., Saffery, R., Bockmann, M., et al. (2017). Host Genetic Control of the Oral Microbiome in Health and Disease. *Cell Host Microbe* 22, 269–278.e263. doi: 10.1016/j.chom.2017.08.013

Halsey, L. G., Curran-Everett, D., Vowler, S. L., and Drummond, G. B. (2015). The Fickle P Value Generates Irreproducible Results. *Nat. Methods* 12, 179–185. doi: 10.1038/nmeth.3288

Ioannidis, J. P. A. (2018). The Proposal to Lower P Value Thresholds to .005. *JAMA-J. Am. Med. Assoc.* 319, 1429–1430. doi: 10.1001/jama.2018.1536

Ioannidis, J. P. A. (2019). What Have We (Not) Learnt From Millions of Scientific Papers With P Values? *Am. Statistician* 73, 20–25. doi: 10.1080/00031305.2018.1447512

Komaroff, A. L. (2017). The Microbiome and Risk for Obesity and Diabetes. *JAMA-J. Am. Med. Assoc.* 317, 355–356. doi: 10.1001/jama.2016.20099

Lahti, L., and Shetty, S. (2017). *Tools for Microbiome Analysis in R. Version 1.4.2.* Available at: http://microbiome.github.com/microbiome.

Lemos, L. N., Morais, D. K., Tsai, S. M., Roesch, L., and Pylro, V. (2017). "Bioinformatics for Microbiome Research: Concepts, Strategies, and Advances," in *The Brazilian Microbiome*. Eds. V. Pylro and L. Roesch (Cham, Switzerland: Springer), 111–123.

May, A., Abeln, S., Crielaard, W., Heringa, J., and Brandt, B. W. (2014). Unraveling the Outcome of 16S rDNA-Based Taxonomy Analysis Through Mock Data and Simulations. *Bioinformatics* 30, 1530–1538. doi: 10.1093/bioinformatics/btu085

McMurdie, P. J., and Holmes, S. (2013). Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* 8, e61217. doi: 10.1371/journal.pone.0061217

McMurdie, P. J., and Holmes, S. (2014). Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Comput. Biol.* 10, e1003531. doi: 10.1371/journal.pcbi.1003531

Mestre, M., Ruiz-González, C., Logares, R., Duarte, C. M., Gasol, J. M., and Sala, M. M. (2018). Sinking Particles Promote Vertical Connectivity in the Ocean Microbiome. *Proc. Natl. Acad. Sci.* 115, E6799–E6807. doi: 10.1073/pnas.1802470115

Moran, M. A. (2015). The Global Ocean Microbiome. *Science* 350:aac8455. doi: 10.1126/science.aac8455

Mysara, M., Njima, M., Leys, N., Raes, J., and Monsieurs, P. (2017). From Reads to Operational Taxonomic Units: An Ensemble Processing Pipeline for MiSeq Amplicon Sequencing Data. *Gigascience* 6, giw017. doi: 10.1093/gigascience/giw017

Nearing, J. T., Douglas, G. M., Comeau, A. M., and Langille, M. G. (2018). Denoising the Denoisers: An Independent Evaluation of Microbiome Sequence Error-Correction Approaches. *PeerJ* 6, e5364. doi: 10.7717/peerj.5364

NIH HMP Working Group, Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L., et al. (2009). The NIH Human Microbiome Project. *Genome Res.* 19, 2317–2323. doi: 10.1101/gr.096651.109

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2019) *Vegan: Community Ecology Package. R Package Version 2.5-4.* Available at: https://CRAN.R-project.org/package=vegan.

Prodan, A., Tremaroli, V., Brolin, H., Zwinderman, A. H., and Nieuwdorp, M. (2020). Comparing Bioinformatic Pipelines for Microbial 16S rRNA Amplicon Sequencing. *PLoS One* 15, e0227434. doi: 10.1371/journal.pone.0227434

R Core Team (2018). *R: A Language and Environment for Statistical Computing* (Vienna, Austria: R Foundation for Statistical Computing). Available at: https://www.R-project.org/.

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: A Versatile Open Source Tool for Metagenomics. *PeerJ* 4, e2584. doi: 10.7717/peerj.2584

Schloss, P. D. (2018). Identifying and Overcoming Threats to Reproducibility, Replicability, Robustness, and Generalizability in Microbiome Research. *mBio* 9, e00525–e00518. doi: 10.1128/mBio.00525-18

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing Mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09

Schmidt, B. L., Kuczynski, J., Bhattacharya, A., Huey, B., Corby, P. M., Queiroz, E. L., et al. (2014). Changes in Abundance of Oral Microbiota Associated With Oral Cancer. *PLoS One* 9, e98741. doi: 10.1371/journal.pone.0098741

Schmidt, T. S. B., Matias Rodrigues, J. F., and von Mering, C. (2015). Limits to Robustness and Reproducibility in the Demarcation of Operational Taxonomic Units. *Environ. Microbiol.* 17, 1689–1706. doi: 10.1111/1462-2920.12610

Schuster, S. C. (2008). Next-Generation Sequencing Transforms Today's Biology. *Nat. Methods* 5, 16–18. doi: 10.1038/nmeth1156

Stewart, C. J., Auchtung, T. A., Ajami, N. J., Velasquez, K., Smith, D. P., de la Garza, R.2nd, et al. (2018). Effects of Tobacco Smoke and Electronic Cigarette Vapor Exposure on the Oral and Gut Microbiota in Humans: A Pilot Study. *PeerJ* 6, e4693. doi: 10.7717/peerj.4693

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., et al. (2015). Structure and Function of the Global Ocean Microbiome. *Science* 348:1261359. doi: 10.1126/science.1261359

Tremblay, J., and Yergeau, E. (2019). Systematic Processing of Ribosomal RNA Gene Amplicon Sequencing Data. *Gigascience* 8, giz146. doi: 10.1093/gigascience/giz146

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., and Gordon, J. I. (2007). The Human Microbiome Project. *Nature* 449, 804–810. doi: 10.1038/nature06244

Wasserstein, R. L., Schirm, A. L., and Lazar, N. A. (2019). Moving to a World Beyond "P < 0.05". *Am. Statistician* 73 (sup1), 1–19. doi: 10.1080/00031305.2019.1583913

Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., et al. (2017). Normalization and Microbial Differential Abundance Strategies Depend Upon Data Characteristics. *Microbiome* 5, 27. doi: 10.1186/s40168-017-0237-y

Welzel, M., Lange, A., Heider, D., Schwarz, M., Freisleben, B., Jensen, M., et al. (2020). Natrix: A Snakemake-Based Workflow for Processing, Clustering, and Taxonomically Assigning Amplicon Sequencing Reads. *BMC Bioinf.* 21, 526. doi: 10.1186/s12859-020-03852-4

Westcott, S. L., and Schloss, P. D. (2015). De Novo Clustering Methods Outperform Reference-Based Methods for Assigning 16S rRNA Gene Sequences to Operational Taxonomic Units. *PeerJ* 3, e1487. doi: 10.7717/peerj.1487

Westcott, S. L., and Schloss, P. D. (2017). OptiClust, An Improved Method for Assigning Amplicon-Based Sequence Data to Operational Taxonomic Units. *mSphere* 2, e00073–e00017. doi: 10.1128/mSphereDirect.00073-17

Woo, P. C. Y., Lau, S. K. P., Teng, J. L. L., Tse, H., and Yuen, K.-Y. (2008). Then and Now: Use of 16S rDNA Gene Sequencing for Bacterial Identification and Discovery of Novel Bacteria in Clinical Microbiology Laboratories. *Clin. Microbiol. Infect.* 14, 908–934. doi: 10.1111/j.1469-0691.2008.02070.x

Xiao, E., Mattos, M., Vieira, G. H. A., Chen, S., Correa, J. D., Wu, Y., et al. (2017). Diabetes Enhances IL-17 Expression and Alters the Oral Microbiome to Increase Its Pathogenicity. *Cell Host Microbe* 22, 120–128.e124. doi: 10.1016/j.chom.2017.06.014

Xun, W., Liu, Y., Li, W., Ren, Y., Xiong, W., Xu, Z., et al. (2021). Specialized Metabolic Functions of Keystone Taxa Sustain Soil Microbiome Stability. *Microbiome* 9, 35. doi: 10.1186/s40168-020-00985-9

Zafeiropoulos, H., Viet, H. Q., Vasileiadou, K., Potirakis, A., Arvanitidis, C., Topalis, P., et al. (2020). PEMA: A Flexible Pipeline for Environmental DNA Metabarcoding Analysis of the 16S/18S Ribosomal RNA, ITS, and COI Marker Genes. *Gigascience* 9, giaa022. doi: 10.1093/gigascience/giaa022

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Cellular and Infection Microbiology

ORIGINAL RESEARCH
published: 25 October 2021
doi: 10.3389/fcimb.2021.734416

Improved Metabolite Prediction Using Microbiome Data-Based Elastic Net Models

*Jialiu Xie[1], Hunyong Cho[1], Bridget M. Lin[1], Malvika Pillai[2], Lara H. Heimisdottir[3], Dipankar Bandyopadhyay[4], Fei Zou[1,5], Jeffrey Roach[6], Kimon Divaris[3,7] and Di Wu[1,8]**

[1] Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, [2] Carolina Health Informatics Program, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, [3] Division of Pediatric and Public Health, Adams School of Dentistry, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, [4] Department of Biostatistics, School of Medicine, Virginia Commonwealth University, Richmond, VA, United States, [5] Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, [6] Research Computing, University of North Carolina, Chapel Hill, NC, United States, [7] Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, [8] Division of Oral and Craniofacial Health Research, Adams School of Dentistry, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Microbiome data are becoming increasingly available in large health cohorts, yet metabolomics data are still scant. While many studies generate microbiome data, they lack matched metabolomics data or have considerable missing proportions of metabolites. Since metabolomics is key to understanding microbial and general biological activities, the possibility of imputing individual metabolites or inferring metabolomics pathways from microbial taxonomy or metagenomics is intriguing. Importantly, current metabolomics profiling methods such as the HMP Unified Metabolic Analysis Network (HUMAnN) have unknown accuracy and are limited in their ability to predict individual metabolites. To address this gap, we developed a novel metabolite prediction method, and we present its application and evaluation in an oral microbiome study. The new method for predicting metabolites using microbiome data (ENVIM) is based on the elastic net model (ENM). ENVIM introduces an extra step to ENM to consider variable importance (VI) scores, and thus, achieves better prediction power. We investigate the metabolite prediction performance of ENVIM using metagenomic and metatranscriptomic data in a supragingival biofilm multi-omics dataset of 289 children ages 3–5 who were participants of a community-based study of early childhood oral health (ZOE 2.0) in North Carolina, United States. We further validate ENVIM in two additional publicly available multi-omics datasets generated from studies of gut health. We select gene family sets based on variable importance scores and modify the existing ENM strategy used in the MelonnPan prediction software to accommodate the unique features of microbiome and metabolome data. We evaluate metagenomic and metatranscriptomic predictors and compare the prediction performance of ENVIM to the standard ENM employed in MelonnPan. The newly developed ENVIM method showed superior metabolite predictive accuracy than MelonnPan when trained with metatranscriptomics data only, metagenomics data only, or both. Better metabolite prediction is achieved in the gut microbiome compared with the oral microbiome setting. We report the best-predictable compounds in all these three datasets from two different body sites.

OPEN ACCESS

Edited by:
Thuy Do,
University of Leeds, United Kingdom

Reviewed by:
Luiz Felipe Valter De Oliveira,
BiomeHub, Brazil
Jun Chen,
Mayo Clinic, United States

***Correspondence:**
Di Wu
dwu@unc.edu

Specialty section:
This article was submitted to Microbiome in Health and Disease, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: *01 July 2021*
Accepted: *22 September 2021*
Published: *25 October 2021*

Citation:
Xie J, Cho H, Lin BM, Pillai M, Heimisdottir LH, Bandyopadhyay D, Zou F, Roach J, Divaris K and Wu D (2021) Improved Metabolite Prediction Using Microbiome Data-Based Elastic Net Models. Front. Cell. Infect. Microbiol. 11:734416. doi: 10.3389/fcimb.2021.734416

For example, the metabolites trehalose, maltose, stachyose, and ribose are all well predicted by the supragingival microbiome.

Keywords: microbiome, metatranscriptome, metabolome, prediction, elastic net, random forest, metagenomics

INTRODUCTION

The importance of the human microbiome in health and disease is undeniable; site-specific microbial communities interact both with the environment and the host and influence numerous biological processes (Wilkinson et al., 2021). Aside from the logical interest in understanding the composition of the microbiome (Tsilimigras and Fodor, 2016) (i.e., relative abundance of identified taxa), measuring and understanding its associated metabolic activities are arguably of utmost biological relevance. Recent studies have linked the metabolome with several important health conditions including inflammatory bowel disease (IBD) (Lloyd-Price et al., 2019), obesity and type II diabetes (Canfora et al., 2019), cholesterol levels (Kenny et al., 2020), and early childhood dental caries (ECC) (Heimisdottir et al., 2021). Despite the rapidly increasing availability of microbiome data in large health cohorts, metabolomics data are still scant. This is an important limitation because the lack of, or considerable missingness of, metabolite information in microbiome studies can diminish their potential in inferring functions and important biological targets.

It follows that methods that help fill in the functional information gaps in microbiome studies are valuable and necessary. Because "matched" microbiome and metabolome datasets are extremely scant, most current methods rely on metabolic pathway inferences from taxonomic and metagenomic data, such as in the HMP Unified Metabolic Analysis Network (HUMAnN) (Franzosa et al., 2018). While the value of this approach is well-documented for the analysis of some microbial consortia (e.g., the human gut) (Lloyd-Price et al., 2019; Thomas et al., 2019), HUMAnN cannot make predictions for individual metabolites. Moreover, its accuracy has not been benchmarked and its performance in other microbial communities with distinct ecology and function (e.g., the oral cavity) remains unknown. This is important because metabolomes measured at different body sites may include, besides the products of microbial metabolism, biochemical contributions from the host and the environment [e.g., dietary sugars in the study of dental biofilm (Heimisdottir et al., 2021)]. Although an accurate determination of metabolite sources may not always be possible, predictions of these biofilm metabolites using microbiome information are highly desirable.

Along these lines, in 2016, Noecker and colleagues (Noecker et al., 2016) added to the available analytical toolbox by leveraging 16S rRNA data. Their method enabled model-based integration of metabolite observations and species abundances using taxonomy and paired metabolomics data from ~70 vaginal samples. More recently, MelonnPan (Mallick et al., 2019) was developed to obtain metabolomic profiling of microbial communities using amplicon or metagenomic sequences. This new method was motivated by and applied in the context of paired microbiome and metabolome data in the context of an IBD cohort. The motivation for the present new method development is to improve existing analytical approaches available for metabolite prediction and functions using microbiome data (Sanna et al., 2019). To this end, we leverage existing microbiome and metabolome data from a study of early childhood oral health (ECC study) and two IBD studies of the human gut microbiome. The elastic net model (ENM, used in MelonnPan), compared to LASSO or ridge regression, benefits from keeping both the singularities at the vertices, which is necessary to accommodate data sparsity, and the strict convex edges for grouping among correlated variables.

Inspired by MelonnPan and MIMOSA, we propose an improved prediction method for individual metabolites using microbiome information in the same (i.e., matched or paired) biological samples, called "elastic net variable importance model (ENVIM)". ENVIM improves upon ENM algorithms by weighting microbial gene family features using random forest variable importance (VI) to enhance the contribution of most prediction-informative genes. ENVIM outputs predicted metabolites from matched microbiome samples, as well as gene families and their weights informing metabolite prediction.

In this paper, we present the development, application, and evaluation of ENVIM. We compare it against MelonnPan in three datasets generated from oral and gut samples, so that we can also compare metabolite predictive performance between different body sites. The predictors can be three different gene family data types: metagenome only, metatranscriptome only, and the combination of both metagenome and metatranscriptome data. The top predictable compounds have been reported in these three datasets from two different body sites. To quantify the taxonomic and functional relationship of the most prediction-contributing microbial gene families in ENVIM, an enrichment analysis is performed and several predictive gene families are detected in species of the oral biofilm.

MATERIAL AND METHODS

Cohorts and Data Description

In the following section, we describe the microbiome and metabolome data used for the new method development and application, alongside the three contributing studies.

ZOE 2.0 Study Data

ZOE 2.0 is a community-based molecular epidemiologic study of early childhood oral health in North Carolina (Divaris et al., 2020; Divaris and Joshi, 2020). The study collected clinical information on preschool-age children's (ages 3–5) dental cavities (referred to as

early childhood caries or ECC) (Ginnis et al., 2019) and supragingival biofilm samples from a sample of over 6,000 children (Divaris et al., 2019). A subset of participants' biofilm samples underwent metagenomics, metatranscriptomics, and metabolomics analyses, under the umbrella Trans-Omics for Precision Dentistry and Early Childhood Caries or TOPDECC (accession: phs002232.v1.p1) (Divaris et al., 2020). As such, metagenomics (i.e., shotgun whole-genome sequencing or WGS), metatranscriptomics (i.e., RNA-seq), and global metabolomics data (i.e., ultra-performance liquid chromatography-tandem mass spectrometry) (Evans et al., 2009; Evans et al., 2014; Heimisdottir et al., 2021) from supragingival biofilm samples of ~300 children, paired with clinical information on ECC, are available. After exclusions due to phenotype and metabolite missingness described in a previous publication (Heimisdottir et al., 2021), the joint microbiome–metabolome data include 289 participants. There are 503 known metabolites included in the ZOE 2.0 dataset. Metagenomics and metatranscriptomics data in reads per kilobase (RPK) were generated using HUMAnN 2.0. Here, we use species-level (205 species), gene family (403k gene families), pathway (397 pathways), and metabolome (503 metabolites) data.

Lloyd-Price Study Data

The Lloyd-Price dataset (Lloyd-Price et al., 2019) was obtained from the IBD multi-omics database (https://ibdmdb.org). It is derived from a longitudinal study that sought to generate profiles of multiple types of omics data among 132 participants for 1 year and up to 24 time points. Several different types of omics data of the study include WGS shotgun metagenomics, RNA-seq metatranscriptomics, and metabolomics. The corresponding metadata include demographic information such as occupation, education level, and age. These gut microbiome data are in counts per million (CPM) and were derived using functional profiles 3.0 in HUMAnN 3.0. For this study, we merged data of individual gene families for 1,638 samples for 130 subjects and individual metatranscriptomics gene families for 817 samples for 109 subjects, respectively. The merged metagenomics gene family data include about 2,741k gene families and 1,580 samples. Merged metatranscriptomics gene family data include about 1,079k gene families and 795 samples. The metabolomics data were generated using four liquid chromatography tandem mass spectrometry (LC-MS) methods including polar compounds in the positive and negative ion modes, lipids, and free fatty acids and bile acids and include 81,867 metabolites in 546 samples for 106 subjects. Most metabolites have not been annotated into known biochemicals and, thus, were excluded from prediction. After limiting the dataset to known metabolites and removing "redundant ions" in "HMDB" ID, there remained 526 metabolites to be predicted.

Mallick Study Data

The Mallick data comprised the main real-life dataset used in the development of the MelonnPan method (Mallick et al., 2019). These gut microbial data (WGS shotgun sequencing and metabolomics) were collected from two cross-sectional IBD cohort studies, namely, the Prospective Registry cohort for IBD Studies at the Massachusetts General Hospital (PRISM, with 155 subjects) and the Netherlands IBD cohort (NLIBD, with 65 subjects). Therefore, they comprise two independent cohorts of subjects. The raw data were obtained through a combination of shotgun metagenomic sequencing and the same four LC-MS methods (Franzosa et al., 2019) as in the Lloyd-Price study. Gene family data in RPK units were derived using HUMAnN 2.0 and normalized to reads per kilobase per million sample reads (RPKM). The raw metagenomics gene family dataset includes one million gene families. The investigators (Mallick et al., 2019) filtered out genes with low abundance and prevalence resulting in a processed dataset of 811 gene families available in the R package *MelonnPan* (melonnpan.training.data and melonnpan.test.data) for 222 total subjects. The microbiome data have been preprocessed and normalized into relative abundance. The metabolite abundance data (8,848 metabolites and 220 subjects) have been made available by Franzosa et al. (2019). Those authors used 466 metabolites for analyses, a subset that was confirmed experimentally against laboratory standards prior to application in *MelonnPan*. In the present study, we use information from these 466 metabolites to compare the power of the new ENVIM method against *MelonnPan*. To accomplish this, we normalized the metabolite abundance data for all 8,848 metabolites into relative abundance (compositional format, obtained *via* dividing the normalized abundance by the sample-level total normalized abundance). Among them, we used the same 466 metabolites with laboratory standards as selected in the paper of *MelonnPan* (Mallick et al., 2019). Data missingness is not an issue in the Mallick metabolome data.

Metabolomics Data Preprocessing and Normalization

An overview of the approach for metabolome data is presented in **Figure 1** and elaborated in detail below.

Metabolomics Missing Data Imputation: ZOE 2.0 and Lloyd-Price Studies

In ZOE 2.0, 87% of metabolites have some missing data, whereas 58% have missing values in Lloyd-Price. To address missingness in these two cohorts, we applied a rigorous feature-wise quantile regression imputation of left-censored data (QRILC) (Wei et al., 2018) to impute missing metabolite values and avoid underestimated metabolite-level variance, as in a previous publication (Heimisdottir et al., 2021). Each of the 289 included participants has <90% missing data across the 503 metabolites in ZOE 2.0. We applied a similar preprocessing filter for the Lloyd-Price data (i.e., removing outlier subjects, **Supplemental Figure 1**), resulting in the exclusion of 15 outlier subjects with the largest numbers of missing metabolite values, as well as outlier metabolites with >90% missing values. Consequently, we proceeded to analyze 522 metabolites in 531 samples from the Lloyd-Price data.

The QRILC imputation method was applied after a natural log data transformation, and the imputed data were exponentiated to back transform the data to RPK (in ZOE 2.0) or CPM (in Lloyd-Price) scales. Because MelonnPan requires metabolite data to be inputted as compositional, we converted RPK and CPM imputed data to a compositional format before predictive modeling.

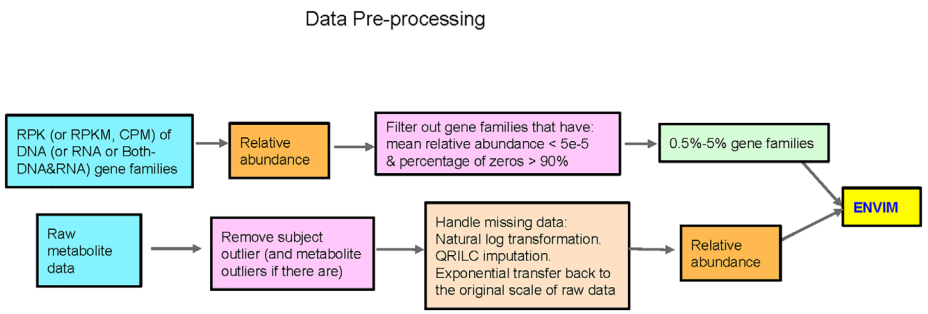


FIGURE 1 | Flowchart of data preprocessing in microbiome and metabolome. QRILC was not used for the Mallick data, but was used for the ZOE 2.0 and Lloyd-Price data. Metabolites that have percentage of NA > 90% will also be removed before handling missing data.

Metabolites Filtered by Metabolic Pathways (ZOE 2.0, Lloyd-Price, and Mallick)

We used the MetaCyc database to retain only "reactive" metabolites (Caspi et al., 2014). To achieve this, we considered the membership of the metabolites in any MetaCyc metabolic pathway, reflecting reactions between bacteria and metabolites, and carried out the following steps:

1. In the MetaCyc database, we identify metabolites in each of the pathways predicted by both metagenomics and metatranscriptomics data in Functional Profile 2.0 generated by HUMAnN 2.0 (ZOE 2.0) and Functional Profile 3.0 generated by HUMAnN 3.0 (Lloyd-Price data). Of note, no pathway information exists in the available Mallick metagenomics and metatranscriptomics data.
2. We used metabolite labels (KEGG ID, HMDB, PubChem, and metabolite name, provided in Metabolome data annotation, provided by the manufacturer) in each of the three datasets, as the mapping IDs for each metabolite.
3. In MetaCyc, regardless of the metabolite label, only one unique MetaCyc "weblink" or universal mapping id is returned if the metabolite is in the database. This way, reactive metabolites identified in step 1 can be matched with metabolites identified in step 2. 3) Therefore, we identify metabolites that are in the observed pathways. Finally, we filter out metabolites with low abundance (metabolites with mean relative abundance $<10^{-4}$) and low prevalence (metabolites with percentage of zeros in >90% of the samples). Consequently, there were 149 metabolites in pathways in ZOE 2.0, 125 in Lloyd-Price, and 251 in the Mallick data. Metabolites in Mallick data only have been filtered by the abundance, without being filtered by metabolic pathways. To compare the prediction of metabolites in pathways with the prediction of all metabolites, we considered both sets of metabolites in our analyses.

Microbiome Data Preprocessing and Normalization

An overview of the approach for microbiome data is presented in **Figure 1** and elaborated in detail below. First, we matched gene family-level microbiome data with metabolome data by participant or sample unique identifier. Then, the scaled (RPK, RPKM, or CPM) gene family abundances were converted to compositional data, relative to the total scaled gene family abundances within a sample. Then, we filtered out gene family features with low relative abundance (mean relative abundance $<5 \times 10^{-5}$) and low prevalence (percentage of zeros in >90% of the samples) and thus kept 0.5%–5% of gene family features. The same procedures were performed for both metatranscriptomics (briefly referred to as "RNA" thereafter) and metagenomics data (briefly referred to as "DNA" hereafter). When both DNA and RNA data (briefly as "Both" hereafter) are considered predictors, a gene name may correspond to two "gene features," one for each data type. The same data preprocessing and normalization procedures were followed for the three cohorts, with sample sizes and feature numbers presented in **Table 1**. To prevent overfitting when evaluating ENM and ENVIM, we divided samples into training (75% of subjects) and testing datasets (25% of subjects).

TABLE 1 | Sample size and number of selected gene family features.

		Training genes	Testing genes	Genes in both	Subjects	Metabolites	Metabolites (in pathways)
ZOE 2.0	DNA (total 403k genes)	1,355	1,276	1,214	289	503	149
	RNA (total 403k genes)	1,805	1,826	1,667	287	503	149
	Both (total 806k genes)	3,158	3,183	2,948	287	503	149
Lloyd-Price	DNA (total 2,741k genes)	726	712	633	359	522	125
	RNA (total 1,079k genes)	726	704	600	282	522	125
	Both (total 3,820k genes)	1,424	1,508	1,211	269	522	125
Mallick	DNA (total 1,000k genes)	811	811	811	220	466	251 (filter only)

Testing genes: genes that can be used in the testing set. Training genes: genes that can be used in the training set. Genes in both: genes that are in both training and testing sets.

The Existing ENM Method for Microbiome Data-Based Metabolite Prediction

As mentioned previously, the existing method available for predicting metabolite abundance using metagenomics data is MelonnPan (Mallick et al., 2019) (Model-based Genomically Informed High-dimensional Predictor of Microbial Community Metabolic Profiles). In this study, in MelonnPan, we used all filtered metagenomic gene family features in the 10-fold cross-validated elastic net model (ENM) (Zou and Hastie, 2005) to predict metabolite abundance (**Equation 1**).

However, using all filtered metagenomic gene family features in the model may dilute the effect of some important gene family features contributing to the prediction of metabolite abundance. This limitation can be improved upon, and therefore, in this paper, we set out to improve the ENM and develop a new algorithm.

The MelonnPan software was downloaded from GitHub (https://github.com/biobakery/melonnpan) or in *MelonnPan* package in R, and the CSV output files "Predicted_Metabolites.txt" (Train) and "MelonnPan_Predicted_Metabolites.txt" (Test) are used as the prediction results of MelonnPan.

The ENM assumes the model,

$$y_i = x_i^{'}\beta + \varepsilon_i,$$

where $\beta = (\beta_0, \beta_1, \ldots, \beta_p)'$ and $\hat{\beta}$, the ENM estimator of β, is found by minimizing the objective function of ENM,

$$L_{ENM} = \frac{1}{2N}\sum_{i=1}^{N}(y_i - x_i^{'}\beta)^2 + \lambda\sum_{j=1}^{p}\left\{\frac{1-\alpha}{2}\beta_j^2 + \alpha|\beta_j|\right\}. \quad \textbf{Equation 1}$$

Evaluation Methods

Following Cohen's criterion (Cohen, 1988), by which a correlation coefficient of 0.3 is considered to be the median size, we define well-predicted (WP) metabolites as those with Spearman correlation ≥0.3, and those with correlation <0.3 as poorly predicted. This criterion has also been used in the development of MelonnPan (Noecker et al., 2016). We evaluated the predictive performance of the new method ENVIM by comparing it against MelonnPan. Additionally, we compared Spearman correlations and mean square error (MSE) between the predicted and observed metabolites in both the training stage and the testing stage for all three datasets and both methods.

RESULTS

The Improved ENM Based on Variable Importance Score (ENVIM)

Algorithm and Procedure in ENVIM

The new algorithm ENVIM (**Equation 2**) was developed by extending the existing ENM with the random forest-derived variable importance to enhance the weights of important features in the prediction. ENM was previously used in the MelonnPan framework for microbiome-based metabolome prediction. The procedure in ENVIM and the comparison between ENM and ENVIM are shown in **Figure 2**. Because ENM assumes the normality of the error term, and there are typically excess zeros, skewness, and extreme values in metagenomics and metatranscriptomics data, we rank-transform gene family features in each sample to a normal distribution by using the *rntransform* (Aulchenko et al., 2007) function in the R package *GENABEL* for training data and testing data separately. The training metabolite abundance data are transformed to a normal distribution using a Box–Cox transformation. After fitting the model in the training data, predicted metabolite abundances are transformed back to relative abundances with γ determined by the training metabolite abundance data.

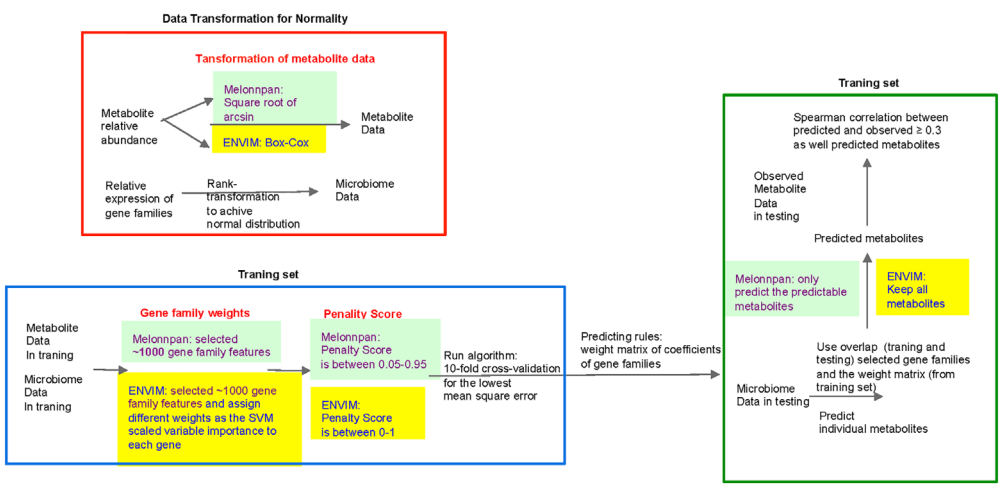


FIGURE 2 | Flowchart of MelonnPan and the elastic net variable importance model (ENVIM). The three differences between them include (red text) 1) transformation of metabolite data, 2) gene family weights, and 3) penalty score. The predictable metabolites are defined as the metabolites that have a significant Spearman correlation with the adjusted *q*-value (testing whether the correlation is zero) below the default threshold in the training set.

Including all gene families into the model could make the cross-validated MSE larger, whereas including only a small part could make the error larger. Therefore, to identify a model with minimum cross-validated error, one needs to iterate different numbers of gene families. Because we prioritize gene families with high importance relative to metabolites, we use a non-linear regression model to determine the importance of gene families for each metabolite. We train a cross-validated random forest model (Breiman, 2001) by using the training data and use the *varImp* function in the *caret* package (Kuhn, 2008) in R to find the scaled importance score (0–100) between each independent feature and the metabolite abundance. We introduce a unique step that uses the scaled variable importance scores to define sets of the top gene families according to a predefined set of thresholds, for example, 90, 80, 70, etc. We use the *glmnet* (Friedman et al., 2010) package in R to run cross-validated ENM and choose penalty parameters for each model.

In the *training stage*, we assign the importance score from 0 to 100 in 10 cumulative intervals (90–100, 80–100,…, 10–100, 0–100) and remove the intervals without gene families. In the ENM, we consider gene families as the independent variables and metabolite abundances as the dependent variables. We consider different sets of gene families with different importance scores. For each set of gene families, we conduct a 10-fold cross-validated ENM and build 10 models with different values of the tuning parameter γ, ranging from 0 to 1. For each model, we measure the MSE between the measured metabolite abundance and the predicted values to determine the best model (i.e., the model with the lowest MSE). To maintain reproducibility, we maintain the same random seed and permute the same fold index number in the ENM. The matrix of regression coefficients of gene families from the best model identified in the training set will be output as a weight matrix.

In the *testing stage*, for the prediction of each metabolite, we use the weight matrix output from the training stage for prediction, if the gene families are also detected in the testing set. Because we have transformed the compositional metabolite abundance to a normal distribution using a Box–Cox transformation in the training stage, we transform the predicted metabolite abundance data back to the original compositional scale based on γ calculated in the training step.

ENVIM assumes the following model:

$$y_i = x_i^{'}\beta + \varepsilon_i,$$

where $\beta = (\beta_0, \beta_1, \ldots, \beta_p)'$, and $\hat{\beta}^{ENVIM} = \arg\min_{\beta} \min_{k \in K} L_{ENVIM}(k)$, the ENVIM estimator of β, is found by minimizing over k and β the objective function,

$$L_{ENVIM}(k) = \frac{1}{2N}\sum_{i=1}^{N}(y_i - x_i^{'}M_k\beta)^2 + \lambda\sum_{j=1}^{p}s_{k,j}\left\{\frac{1-\alpha}{2}\beta_j^2 + \alpha|\beta_j|\right\}. \quad \textbf{Equation 2}$$

Here we define VI_j as the variable importance score for the jth variable given by a random forest; $S_k = \{S_{k,j}\}_j^p = I\{VI_j \geq k\}_{j=1}^p$ is the variable selection indicator vector, which is 1 if the importance score for the jth variable is larger than the threshold k; $M_k = diag\{(1, S_k^{'})\}$ is the corresponding diagonal variable selection matrix that includes the intercept term; and K is a set of the candidate k values. K is defined adaptively so that it covers the range of the variable importance scores reasonably. In our analysis, we set $K = \{0, 10, 20, \ldots, 90\}$.

Three Key Differences Between MelonnPan and ENVIM for Predicting Individual Metabolites

1. Transformation of metabolite abundance data into a normal distribution

MelonnPan transforms relative metabolite abundances with the arcsine square root operator, whereas we use a Box–Cox transformation in ENVIM. To test the normality of the transformed data, we compare the p-values of the Shapiro–Wilk test statistics for both the Box–Cox (**Equation 3**) and the arcsine square root transformations of metabolite relative abundances. The Shapiro–Wilk test is typically used for examining distribution normality for a continuous variable. The smaller the p-values, equivalently, the larger the $-\log_{10}(p\text{-values})$ are, the more evidence the data are not normally distributed. Overall, the $-\log_{10}(p\text{-values})$ from the Box–Cox transformation in ENVIM are smaller than those from the arcsine square root transformation (**Figure 3A**), which indicates that the Box–Cox-transformed data are more normally distributed. In addition, the Box–Cox transformation yields better normal approximation than the arcsine square root transformation for most of the metabolites (**Figure 3B**).

Box–Cox transformation

$$y' = \begin{cases} \frac{y^{\omega}-1}{\omega}, \omega \neq 0 \\ \log(y), \omega = 0 \end{cases}, \qquad \textbf{Equation 3}$$

where y is the relative abundance, and y' is the transformed abundance.

2. Different sets of gene families are carried forward to the prediction model

MelonnPan uses all gene families in the training data in the ENM and ultimately predicts metabolites in the testing stage using the same features. However, regressing against all gene families may dilute the effect of important gene families. Thus, unlike MelonnPan, we use a variable importance criterion to select different sets of gene families and include them in the prediction models.

3. The range of α values in ENM

Alpha (α) is the weight between the L_1 and L_2 penalty terms in the ENM, and in combination with γ values, the set of values that minimizes the 10-fold cross-validated MSE (**Equation 1**) is chosen. When α is 0, the model reduces to a Ridge regression model which has the advantage of dealing with highly correlated independent variables; when α is 1, the model becomes a Lasso regression model which has a variable selection capacity; when α is between 0 and 1, the model includes the advantages of both Ridge regression and Lasso regression. In MelonnPan, the range of α values does not include 0 and 1, which excludes either the

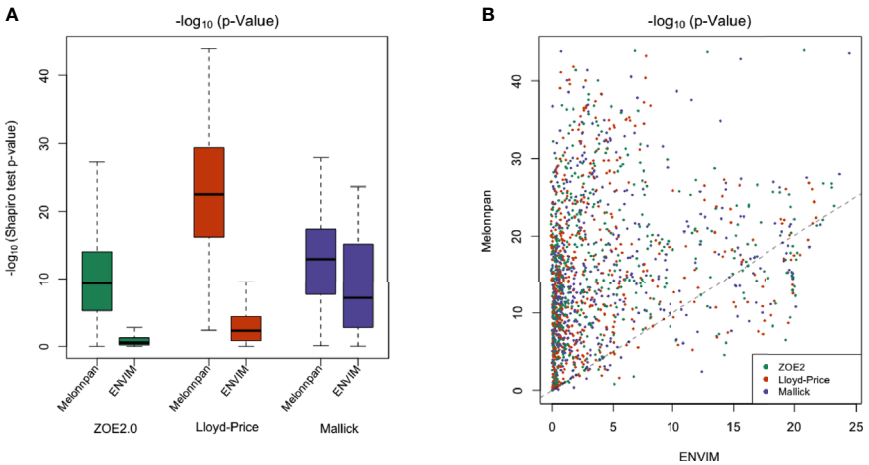


FIGURE 3 | **(A)** Boxplot of $-\log_{10}$ of Shapiro–Wilk test *p*-values to test the normality of transformed relative metabolite abundances in all three data applied with Box–Cox transformation (ENVIM used) and arcsine square root transformation (MelonnPan used). **(B)** Scatter plot for comparing $-\log_{10}$ of *p*-values from the Shapiro–Wilk test (normality) between Box–Cox transformation (*x*-axis) and arcsine sqrt (*y*-axis) transformation. Almost all of the points are above the $y = x$ line, which indicates that the $-\log_{10}$ of *p*-value after Box–Cox transformation is smaller than after arcsine sqrt transformation and normality after Box–Cox transformation is better. Each point is one metabolite.

Ridge or Lasso regression models, and it may not consider variables with high importance. The range of α in our ENVIM includes 0 and 1. By allowing a larger range of α, we can include the Ridge regression model as the potential final model, which does not unduly exclude variables with high importance.

The ENVIM software written in R statistical language is available in GitHub (https://github.com/jialiux22/ENVIM). The *ENVIM_predict* function is for metabolite prediction only, and the *ENVIM* function for both the metabolite prediction and the evaluation of the observed metabolomics dataset in the testing set is also available. Both R functions will output the weight matrix between gene families and metabolites. The weight matrix in testing has the same values as in training if they have the same number of gene families. Some contributing gene families in the weight matrix of the training set may not be measured in the testing set, so the weight matrix used by the testing set includes only the gene families that are shared by both the training and the testing sets.

Method Comparison for Prediction of Individual Metabolites in the Three Datasets

Correlation-Based Method Comparison for All Metabolites

We used microbial gene family data to predict individual metabolites in the matched samples (that are from the same biological sample in that one proportion is for microbiome and the other is for metabolome). We compared the prediction results between ENVIM and MelonnPan, in terms of Spearman correlation and MSE between predicted and observed values of each of the metabolites, in three datasets (ZOE 2.0, Mallick data, and Lloyd-Price data) at each of the three data modalities of microbial gene families, i.e., DNA-seq, RNA-seq, and Both (RNA and DNA). The MSE in the testing set is used for comparison between the methods (**Supplemental Figure 2**).

We have summarized the prediction results (**Table 2** and **Figure 4A**) for all metabolites in terms of Spearman correlation ($r = 0.3$) according to three aspects: method comparison, data modality comparison, and microbial community (i.e., body site) comparison. Overall, in method comparison, ENVIM produces higher percentages of well-predicted metabolites than MelonnPan in all three datasets, in both testing and training sets, and for DNA, RNA, and Both when available (**Table 2**).

Generally, in data modality comparison, RNA gene family data produce higher percentages of well-predicted metabolites than DNA data. In the Lloyd-Price study, RNA-only data typically give higher percentages of well-predicted metabolites. In the ZOE 2.0 and Lloyd-Price data, both DNA and RNA predictors produce similar percentages but are not always superior to the DNA-only or RNA-only data-based predictors. However, results from both DNA and RNA predictors are never the worst. Unsurprisingly, the well-predicted percentage of metabolites in testing sets is lower than in the training set (**Table 2**). The boxplots of Spearman correlations between the predicted and observed metabolites for all metabolites (**Figure 4A**) suggest that the correlations between the ENVIM-predicted and the observed metabolites are higher in RNA than in DNA, but are comparable to correlations in both DNA and RNA. We are aware that in the testing sets, MelonnPan only outputs the predictable metabolites (defined as well-predicted metabolites in the training set, the last columns in **Table 2**), so it is not as appropriate for MelonnPan, as compared with ENVIM, to calculate the Spearman correlation distribution for all

TABLE 2 | Prediction results (first four columns of numbers) in terms of Spearman correlation for all metabolites to be predicted.

	Training (ENVIM)	Training (MelonnPan)	Testing (ENVIM)	Testing (MelonnPan)	Predictable metabolites (defined by MelonnPan)
ZOE 2.0 (NM = 503)					
DNA only	356 (71%)	63 (13%)	**124 (25%)**	47 (9%)	70
RNA only	409 (81%)	157 (31%)	106 (21%)	68 (14%)	163
Both DNA and RNA	423 (84%)	146 (29%)	110 (22%)	73 (15%)	154
Mallick cohort (NM = 466)					
DNA only	408 (88%)	239 (51%)	225 (48%)	178 (38%)	249
Lloyd-Price cohort (NM = 522)					
DNA only	501 (96%)	271 (52%)	322 (62%)	193 (37%)	305
RNA only	521 (100%)	298 (57%)	**393 (75%)**	236 (45%)	318
Both DNA and RNA	518 (99%)	306 (59%)	381 (73%)	232 (44%)	323

Based on the "well-prediction" criterion, defined as Spearman correlation ≥0.3 between the observed and the predicted metabolites, the numbers of well-predicted metabolites with different prediction methods, datasets, and modality levels (DNA, RNA, and Both) are presented for comparing MelonnPan and ENVIM. NM is the number of metabolites to be predicted. Percentages in parentheses (%) represent the number of well-predicted metabolites divided by the total number of metabolites (NM) to be predicted in each study. The Mallick cohort has only metagenomics data available. The last column presents numbers of "predictable metabolites," defined by MelonnPan, also seen in the ***Figure 2*** *legend. Bold in the column of in testing results represents the highest number of well-predicted metabolites among the three modalities (DNA, RNA, both DNA and RNA) in the ZOE2.0 cohort and the Lloyd-Price cohort.*

metabolites in **Figure 4A**. It must also be noted that the highest proportion of well-predicted metabolites is found in the two gut microbiome studies (Lloyd-Price study and Mallick study), and the lowest is in the supragingival dental biofilm (ZOE 2.0 study) (**Table 2**). Since Spearman correlation in both the Lloyd-Price and Mallick datasets is higher than that in ZOE 2.0 (**Figure 4A**), it is reasonable to suggest that metabolite prediction is better in gut microbial communities than in the oral microbial communities.

Besides comparing MelonnPan and ENVIM in terms of percentages of well-predicted metabolites, one can directly compare the Spearman correlations of each predictable metabolite that is predicted by both methods (**Figures 5**, **6**). In the training set (**Figure 5**), for all three gene family data modalities and in all three datasets, we find that the majority of these metabolites have higher correlations in ENVIM compared with MelonnPan. The same holds in the testing set (**Figure 6**). We also observe that most points are along but slightly above the diagonal line in the testing sets (**Figure 6**). This suggests that the metabolites predicted by ENVIM have higher correlations with the observed ones compared with those predicted by MelonnPan. We also find that there are more metabolites in the "ENVIM ≥0.3" category (blue) than in the "MelonnPan ≥0.3" category (red). This is a reflection of more well-predicted metabolites found using ENVIM than using MelonnPan prediction.

To give a more realistic view of the improvement of the ENVIM over MelonnPan, as a tool to predict metabolites in practice, we use one of the two independent cross-sectional cohorts in Mallick data as training to predict the other. The PRISM cohort has 155 subjects and the NLIBD cohort has 65 subjects. For both ENVIM and MelonnPan, we use the microbiome and metabolome data in PRISM as the training set to predict the metabolites in NLIBD (**Table 3**). Among the 466 metabolites, ENVIM has 34% (160/466) in the testing set, while MelonnPan only has 26% (123/466) in the testing set. These percentages are very similar to 37% and 28%, respectively, in ENVIM and MelonnPan from random split of samples in the Mallick study, so that the same conclusion was drawn that better prediction power is in ENVIM than in MelonnPan.

To investigate the sample size effects, we further cut the sample size of the training set by half, or from 155 PRISM subjects to 77 or 78 subjects randomly for 10 times and find that with even half of the samples, ENVIM nearly maintains the well-predicted rates (**Table 4**). ENVIM is less sensitive to the decreased sample size than MelonnPan.

Correlation-Based Method Comparison for Metabolites Within Metabolic Pathways

Metabolites may be associated with the microbiome in the context of metabolic pathways that involve interactions between the host, microbiome, and environment. We further investigate the predictive capability of the two methods for metabolites in MetaCyc metabolic pathways. HUMAnN 2.0 or 3.0 software provides information whether a MetaCyc metabolic pathway has been associated with microbiome data. In the MetaCyc database, we identify metabolites in each of these microbiome-associated pathways. All conclusions regarding method comparison, modality comparison, and body site comparison in the prediction of all metabolites still hold in the context of predicting metabolic-pathway-only metabolites. Additionally, when comparing the percentages of well-predicted metabolites among all metabolites (first four columns of **Table 2**) and those in the metabolic pathways (**Table 5**), we find higher predicted percentages for the latter.

MSE-Based Method Comparison

We use boxplots to compare MSE between measured and predicted metabolite abundances between ENVIM and MelonnPan both for training and testing models, with application to training and testing data for all three studies. We only compare well-predicted metabolites identified by MelonnPan in training, since MelonnPan only generates results for these metabolites. The boxplot demonstrates that the distribution of MSE in the MelonnPan model is approximately the same as the distribution of MSE in ENVIM (**Supplemental Figure 2**). We find no significant difference in MSE between ENVIM and MelonnPan, which suggests that both

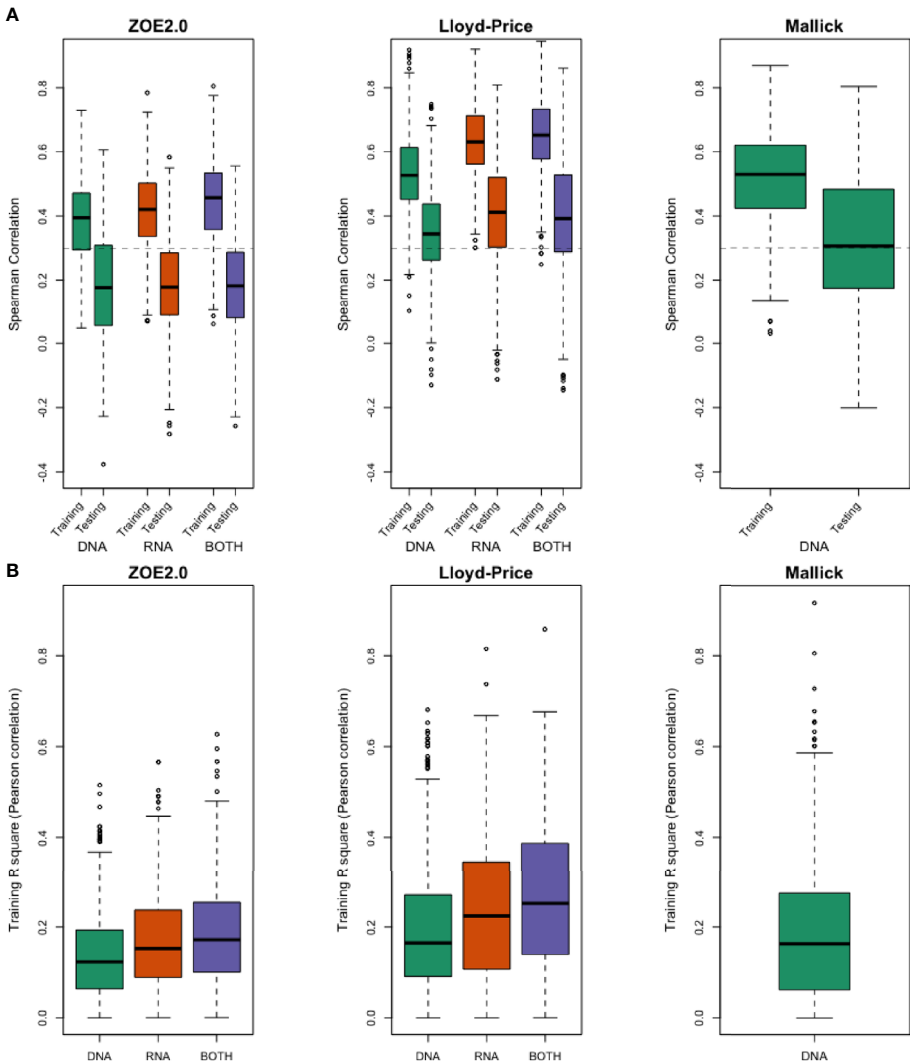


FIGURE 4 | **(A)** Evaluation using Spearman correlation *r* in training stage and testing stage between predicted values and the observed values by using DNA-seq data only, RNA-seq data only, and both for ZOE 2.0 data, Lloyd-Price data, and Mallick data. **(B)** *R*-square in the training stage, as the percentage of variance explained by prediction models to demonstrate the lack of overfitting.

methods predict these metabolites well in terms of MSE. The advantage of ENVIM is that we can predict substantially more well-predicted metabolites than MelonnPan—a consequence of MelonnPan's inability to build a well-performing model in the training step. When using PRISM as the training set and NLIBD as the testing set in the Mallick study, the above conclusion about MSE remains the same (**Supplemental Figure 3**).

ENVIM Outputs Including Predicted Individual Metabolites and Contributing Gene Family Weights

Top Well-Predicted Metabolite Compounds From ENVIM

For simplicity, we present one modality from each of the three studies. For Lloyd-Price and ZOE 2.0, we choose one of the gene family data modalities that has the best ENVIM prediction power to show their top predicted metabolites, that is, the DNA gene family data (124 metabolites as 25% among NM, **Table 2**) in ZOE 2.0 and the RNA gene family data (393 metabolites as 75% among NM, **Table 2**) in Lloyd-Price. Since the Mallick study only has DNA data available, the DNA gene family data are used. To note, both the Lloyd-Price study and the Mallick study have measured metabolites in four metabolome LC-MS platforms (see the *Cohorts and Data Description* section) so that one metabolite may appear multiple times in the top list (for example, urobilin). The top 50 best predicted metabolites for each study are presented in **Figure 7**.

The summarized prediction results are presented in **Supplemental Table 1**. To interpret the results, we take the carbohydrate pathway as an example of a pathway that may provide bacteria with nutrition, which includes a few compounds

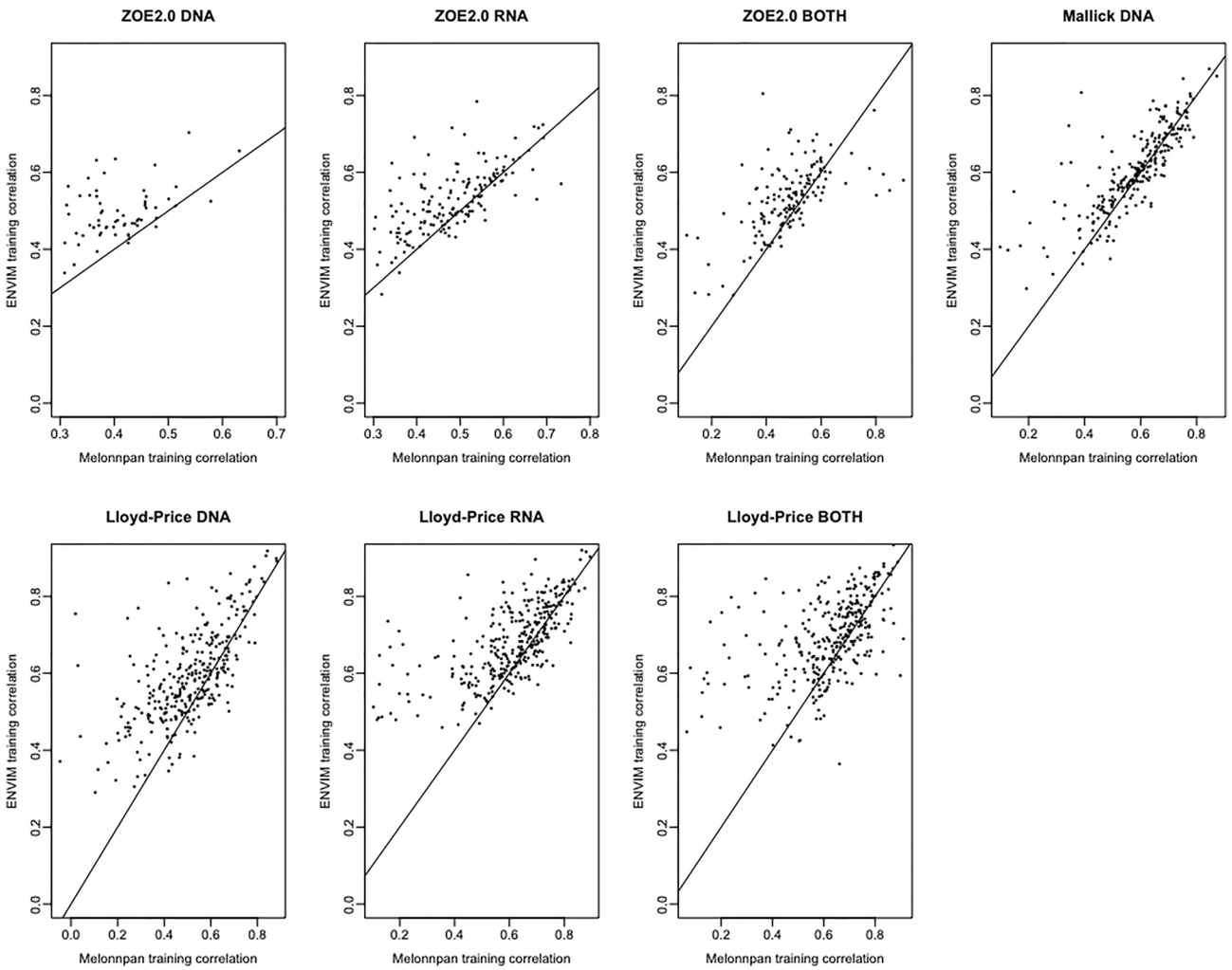


FIGURE 5 | For DNA, RNA, and both in each study and the training set, this shows the scatter plot of Spearman correlation in ENVIM (*y*-axis) and MelonnPan (*x*-axis). Spearman correlation is based on observed metabolite abundance and predicted values. If our calculated correlation is NA, the metabolites will be not included in this figure.

that have been well-predicted by the RNA gene family data. We are aware that the prediction in this paper is not about longitudinal causal relation, but rather, for mathematical prediction. Here, we also show four examples (trehalose, maltose, ribose, and stachyose) that have high Spearman correlation on the $\log_{10}$ scale of the compositional data (**Figure 8A**).

Comparison of Gene Family Lists (With Weight Matrix) Across Three Datasets in ENVIM

We extract gene family names that have non-zero entries in the weight matrix for each metabolite, dataset, and gene family modality (**Supplemental Table 2**) in ZOE 2.0. We compare the contributing gene family names across the ZOE 2.0 and Lloyd-Price to find the number of common contributing genes the different body sites share for predicting metabolites. We find that there are not many overlapping genes ($n < 10$) between ZOE 2.0 data and Lloyd-Price data (data not shown).

Gene Set Enrichment Analysis of Contributing Genes Within Species in ZOE 2.0

We perform gene set enrichment analysis to find the over-represented species of the contributing gene families to predict metabolites in ZOE 2.0. To test that, we start with the weight matrix of gene families and metabolites in the testing set. We identify the contributing gene families that have non-zero values with any well-predicted metabolites.

We obtain the rank of each gene family in the weight matrix based on the absolute value of the regression coefficients ("weights") for each gene family. We use the information of correspondence between gene families and the species level (generated in HUMAnN

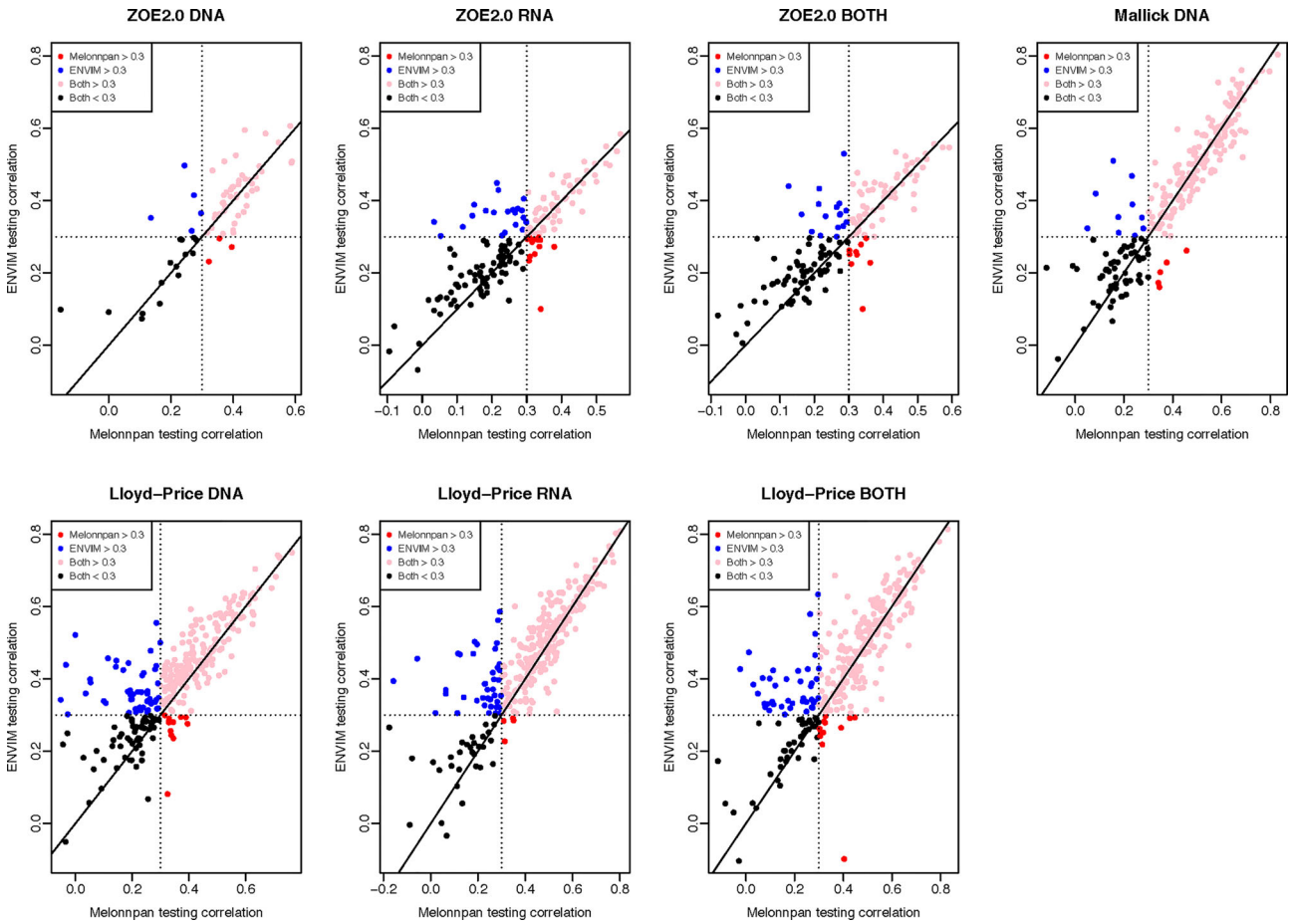


FIGURE 6 | For DNA, RNA, and both in each study and the testing set, this shows the scatter plot of Spearman correlation in ENVIM (*y*-axis) and MelonnPan (*x*-axis). Spearman correlation is based on observed metabolite abundance and predicted values. Here, "Both ≥0.3" refers to the category of metabolites that have Spearman correlation ≥0.3 in both ENVIM- and MelonnPan-predicted results. "ENVIM ≥0.3" refers to the category of metabolites that have Spearman correlation ≥0.3 only between ENVIM-predicted and observed values.

2.0.) to identify the species corresponding to those contributing gene families. For each species, we compare the difference in the cumulative distributions of gene family rank scores between the species and the background species using the Kolmogorov–Smirnov (KS) test that was also used in the original gene set enrichment analysis (GSEA) paper (Subramanian et al., 2005). We use the Benjamini–Hochberg false discovery rate (FDR) approach to correct the KS *p*-values and get *q*-values. There are 36 species in ZOE 2.0 DNA data and 73 species in ZOE 2.0 RNA data found to be significantly ($q < 0.05$) over-represented in the gene set enrichment analysis (**Figure 9**).

Here, we used a different procedure for the gene set enrichment tests compared to what was used in the MelonnPan (Mallick et al., 2019) paper, in that the gene families in genera instead of species were summarized as a gene set, due to the small number of gene families in each species in their prediction procedure. In fact, ENVIM keeps many more genes than MelonnPan (because ENVIM allows larger range of α) so that ENVIM can address the ranks of all contributing gene families instead of the binary prediction power of genes (i.e., whether a gene is used for prediction or not) used in MelonnPan and, furthermore, can perform GSEA at the species level for higher resolution of

TABLE 3 | Prediction results in Mallick data, when using all samples in the PRISM study as the training set and the data in NLIBD study as the testing set.

Mallick cohort (NM = 466) **PRISM (training, *n* = 155)** **NLIBD (testing, *n* = 65)**	**Training (ENVIM)**	**Training (MelonnPan)**	**Testing (ENVIM)**	**Testing (MelonnPan)**
DNA only	387 (83%)	205 (44%)	160 (34%)	123 (26%)

TABLE 4 | Prediction results in Mallick data, when using half of the sample size in the PRISM study as the training set for 10 times and the data in NLIBD study as the testing set.

Mallick cohort (NM = 466) PRISM (training, *n* = 77 or 78) NLIBD (testing, *n* = 65)	Training (ENVIM)	Training (MelonnPan)	Testing (ENVIM)	Testing (MelonnPan)
Seed1	429 (92%)	161 (35%)	147 (32%)	96 (21%)
Seed2	402 (86%)	202 (43%)	162 (35%)	104 (22%)
Seed3	427 (92%)	164 (35%)	140 (30%)	92 (20%)
Seed4	428 (92%)	199 (43%)	148 (32%)	97 (21%)
Seed5	439 (94%)	211 (45%)	160 (34%)	111 (24%)
Seed6	427 (92%)	180 (39%)	157 (34%)	113 (24%)
Seed7	424 (91%)	178 (38%)	143 (31%)	98 (21%)
Seed8	424 (91%)	150 (32%)	142 (30%)	98 (21%)
Seed9	425 (91%)	159 (34%)	150 (32%)	101 (22%)
Seed10	419 (90%)	181 (39%)	152 (33%)	105 (23%)
Mean	424 (91%)	179 (38%)	150 (32%)	102 (22%)

contributing species. Our GSEA strategy also can help avoid the bias of selecting for species that have larger numbers of genes.

Computational Speed

ENVIM was implemented in R statistical language. It can accurately predict metabolites using matched microbiome gene family data. The mean running time in ENVIM of each metabolite using DNA gene family data is 5.2 min for ZOE 2.0 data (6.1 min for Lloyd-Price data, 2 min for Mallick data). The mean running time in ENVIM using RNA gene family data is 4.2 min for ZOE 2.0 data (3.7 min for Lloyd-Price data); the mean running time in ENVIM for both DNA and RNA gene family data is 4.5 min for ZOE 2.0 data (3.6 min for Lloyd-Price data) with MacOS Big Sur Version 11.4 and the desktop iMac Pro 2020.

DISCUSSION

We propose a new computational method for metabolite prediction using microbiome data-based improved elastic net models. We chose different gene family sets based on random-forest-based variable importance scores and modified the existing ENM to accommodate the unique features of microbiome and metabolome data. The newly developed method ENVIM predicts metabolites using metagenomics, metatranscriptomics, or both data types. We apply the algorithm in three datasets, i.e., the ZOE 2.0, Mallick, and Lloyd-Price studies. These three studies have both microbiome and metabolome data in the same matched samples, with reasonably large sample sizes. We are the first to use microbiome data to predict metabolites in more than one study and different body sites. In addition, the ZOE 2.0 and Lloyd-Price studies have both metagenomics and metatranscriptomics, so that we can, for the first time, compare prediction performance using the different gene family modalities (or called data types).

We evaluated metagenomic and metatranscriptomic predictors and compared the prediction performance between the previously developed MelonnPan and ENVIM, among DNA, RNA, and both DNA and RNA gene family data using 1) the proportion of "well-predicted" metabolites defined as those with Spearman correlation between measured and predicted metabolite values ≥0.3, 2) distribution of Spearman correlation, and 3) MSE. The correlation suggests that Both (of DNA and RNA) provides robust prediction

TABLE 5 | Prediction results *via* Spearman correlation for metabolites that are found in metabolic pathways.

	Training (ENVIM)	Training (MelonnPan)	Testing (ENVIM)	Testing (MelonnPan)
ZOE 2.0 (NM = 149)				
DNA only	129 (87%)	35 (23%)	57 (38%)	28 (19%)
RNA only	139 (93%)	73 (49%)	57 (38%)	40 (27%)
Both DNA and RNA	142 (95%)	73 (50%)	60 (40%)	46 (31%)
Mallick cohort (NM = 251)				
DNA only	231 (92%)	132 (53%)	94 (37%)	71 (28%)
Lloyd-Price cohort (NM = 125)				
DNA only	121 (97%)	71 (57%)	70 (56%)	40 (32%)
RNA only	125 (100%)	86 (69%)	103 (82%)	68 (54%)
Both DNA and RNA	124 (99%)	92 (74%)	105 (84%)	79 (63%)

Based on the criterion of Spearman correlation ≥0.3 between observed and predicted metabolites, we present the numbers of well-predicted metabolites with different prediction methods, datasets, and modality levels (DNA, RNA, and Both) and made a comparison between MelonnPan and ENVIM. NM is the number of metabolites to be predicted. Percentages in parentheses (%) represent the numbers of well-predicted metabolites divided by the total number of metabolites (NM) to be predicted in each study. The Mallick cohort has only metagenomics (DNA) data available and no pathway RNA data. The results from the Mallick cohort here are only based on filters (filtering out metabolites with mean relative abundance <10^{-4}) and low prevalence (metabolites with >10% non-zero). In ZOE 2.0 and Lloyd-Price, metabolite data presented in this table have been selected according to membership in pathways and also satisfy the abovementioned filtering criteria.

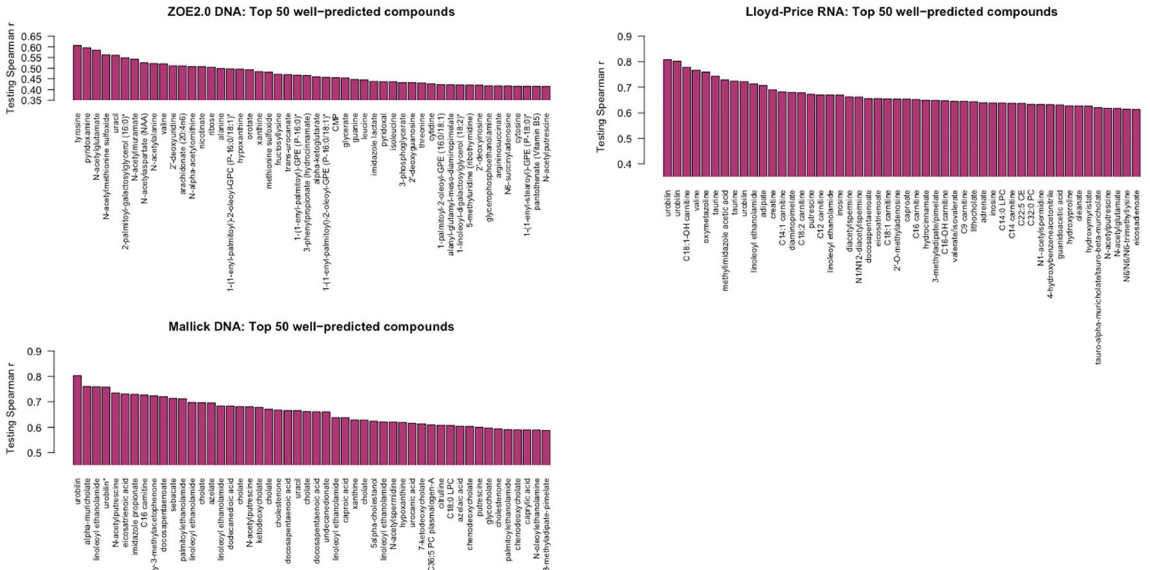


FIGURE 7 | The best predicted 50 metabolite compounds (*x*-axis) in the three studies by ENVIM in the testing set. For Lloyd-Price and ZOE 2.0, we choose the gene family data types that have the best ENVIM prediction power to show their top predicted metabolites, based on **Table 2**.

results that are never the worst among the three data types. Whether DNA or RNA has better prediction performance depends on the study. The percentage of well-predicted metabolites is higher for metabolites that are in a metabolic pathway observed in the microbiome data. Such enrichment of well-predicted metabolites in metabolic pathways supports the strong interaction between microbiome and metabolome. Across all datasets and data types, with or without the pathway filter, we find that ENVIM always outperforms MelonnPan. We also find that prediction performance is better in Lloyd-Price and Mallick than in ZOE 2.0, which suggests that the association between microbiome and metabolites is stronger in the gut than in the oral cavity, since oral metabolites may be more affected by environmental factors like food intake. More microbial omics studies are needed to compare the prediction power across different body sites and to understand how the microbiome interacts with the metabolome differently at different body sites. Acknowledging that the reported findings are not to infer causality but are demonstrative of mathematical prediction, we show four well-predicted metabolites in ZOE 2.0 (**Figure 8**), as examples of compounds that may play roles in bacterial metabolism.

As a result, the numbers of the measured metabolites and the numbers of the to-be predicted metabolites in each of the three studies are very different due to differences in technology platforms, data processing steps, and available data at different body sites. Besides body sites, the data collection and processing steps may have large effects on the prediction performance. The distributional assumption, normalization, transformation, outlier filtering, and missing data handling are important considerations before training the model. We have touched on that, but further exploration may be needed. According to what we observed, the ideal usage of these types of prediction methods is in studies that contain paired microbiome and metabolome data in one batch of samples but lack metabolome data in other batches of samples. In that case, all microbiome samples are sequenced, aligned, and processed comparably, and the uncollected metabolome samples are also assumed to be from the same technical metabolome platform and similar data processing steps (for example, as what we demonstrated in **Tables 2**, **5**). However, the usage of these methods is not limited only to this ideal case. The suitable usage scope has the assumptions of 1) the same population distribution of microbiome data in the training model and in the cohort to be predicted, 2) the same population distributions between the metabolome data in the training model and in the cohort to be predicted, and 3) similar connection between microbiome and metabolome due to, for example, similar ethnicities, clinical characteristics, age groups, and body sites (as shown in **Table 3**). Usage of the IBD Lloyd data to predict metabolites in the IBD Mallick study *via* ENVIM and MelonnPan has been considered. Although these two studies have been generated from the same body sites and similar LC-MS metabolome techniques, their microbiome data have been processed in different versions of HUMAnN software (3.0 vs. 2.0), and in different data scales (CPM vs. RPKM), their metabolome data have been processed using different algorithms in different software, and different filtering criteria have been used in the two studies. These differences suggest that the first and the second assumptions are not held well, and the prediction results are not encouraging (data not shown). Furthermore, the assumptions of similar population distributions depend on the measuring technical platforms, the data processing steps, and proper normalization methods. The questions of what the best normalization method is

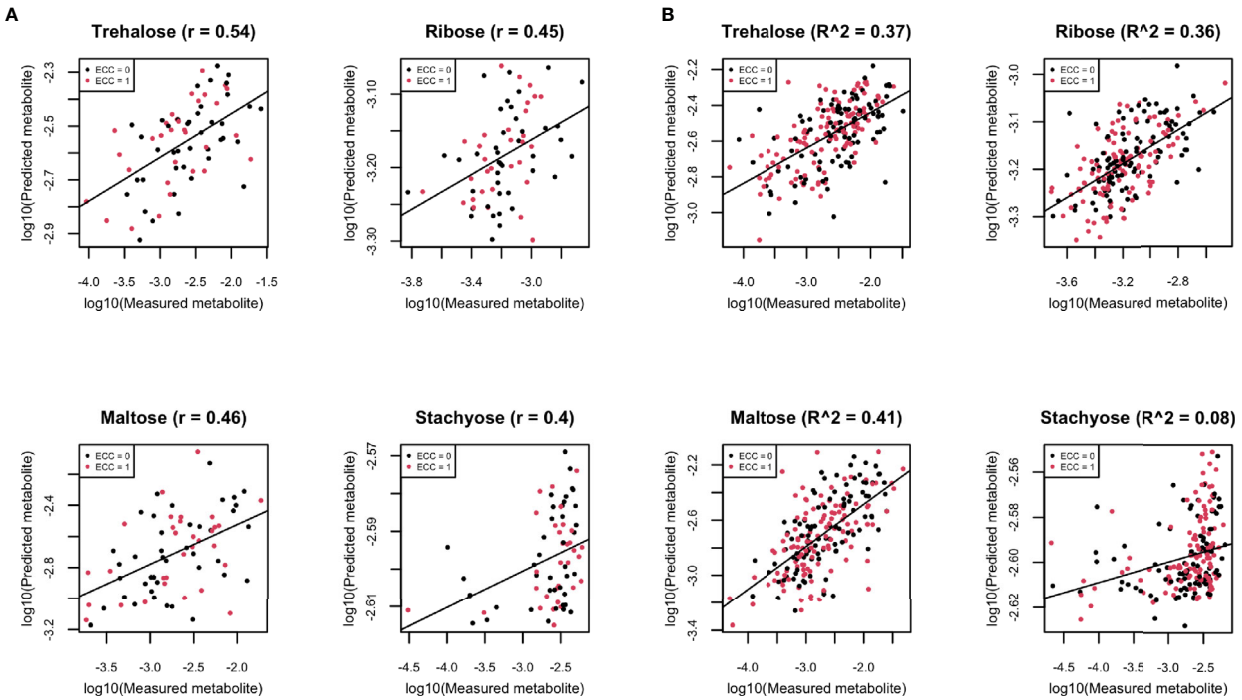


FIGURE 8 | **(A)** Scatter plots of examples of four well-predicted metabolites in ZOE 2.0 by ENVIM, in the testing set. *r* is for Spearman correlation for method evaluation. **(B)** Scatter plots of the same four well-predicted metabolites in ZOE 2.0 by ENVIM, in the training set, where *R*-square (Pearson Correlation) was shown for the percentage of variance explained by prediction models to demonstrate that overfitting is not a big concern. The *x*-axis is the observed metabolites; the *y*-axis is the predicted metabolites. Both *x* and *y* are in $\log_{10}$ scale of the compositional data for normality. ECC stands for early childhood caries, ECC = 0 (about 50% of total samples in ZOE 2.0) is for the healthy group, and ECC = 1 (about 50% of total samples in ZOE 2.0) is for the ECC case group.

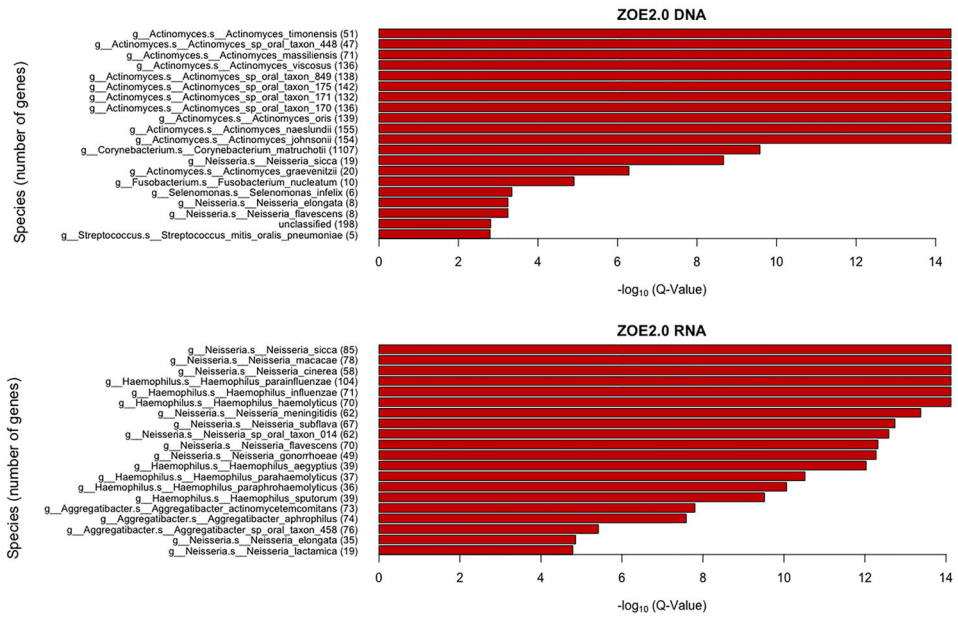


FIGURE 9 | Taxonomic enrichment of metabolite predictive species for the most contributing species to metabolite prediction, based on ZOE 2.0 DNA or RNA by ENVIM. The top 20 significant over-represented bacteria with the smallest *Q*-values ($Q < 0.05$) for ZOE 2.0 data. The *Q*-value is based on the Kolmogorov–Smirnov (KS) test *p*-values after FDR correction. Upper, DNA data; Lower, RNA data.

and which integration procedure best allows assumptions of similar population distributions to overcome the difference across cohorts/technique/data processing are out of the scope of this paper but are very important for further data harmonization of microbiome-related large datasets.

Because higher well-predicted metabolite rates were observe in the training compared to the testing datasets, overfitting of the machine learning model can be a concern; however, overfitting is not a great concern around ENVIM for the following reasons: 1) similar observed mean square error in the training set and the testing set (**Supplemental Figures 2–4**) and 2) small squared Spearman correlation (*R*-square) between fitted and observed metabolites in the training sets (**Figure 4B**). Four well-predicted metabolites in ZOE 2.0 have no large *R*-square in the training set, and similar patterns in the scatter plots between measured and predicted metabolites are observed in both the training set and testing set (**Figure 8B**), and 3) the penalty terms in ENM, cross-validation in tuning the penalty terms, and the use of bootstrapping in random forest relax the potential overfitting problem. Although the overfitting concern is reasonably mitigated, it should be acknowledged that it may not be perfectly avoided. With that in mind, the method performance in the testing set is the most important. We observe that ENVIM has a higher well-predicted metabolite percentage (**Tables 2**, **3**, **5**) and comparable MSE (**Supplemental Figures 2**, **3**) when compared with MelonnPan.

A limitation in the framework for ENVIM, as well as in the framework for MelonnPan, is that the experimental design in studies, including time course or disease statuses, has yet to be considered. However, since the purpose of ENVIM is prediction, the prediction does not need to be conditional on the experimental design. Instead, different disease statuses may have different microbiome profiles and, correspondingly, have different metabolome profiles. Therefore, the non-inclusion of a design matrix in ENVIM is a limitation but not a drawback of the prediction performance.

In summary, we illustrate that the newly developed ENVIM method for microbiome-based metabolite prediction provides good prediction performance and can be used to predict individual metabolites when only microbiome data are available if the same technical microbiome/metabolome platform, similar data processing steps, and the same body site and covariate values can be assumed, or when a proportion of samples in a study have no metabolome data.

DATA AVAILABILITY STATEMENT

The ZOE 2.0 microbiome data have been deposited in dbGaP under the umbrella Trans-Omics for Precision Dentistry and Early Childhood Caries or TOPDECC (accession: phs002232.v1.p1). Metabolomics raw spectral data and associated clinical traits have been deposited in the MetaboLights repository: https://www.ebi.ac.uk/metabolights/MTBLS2215. Descriptive and clinical data for the parent ZOE 2.0 study have been deposited in https://doi.org/10.17615/8yjy-w790.

ETHICS STATEMENT

Ethics approval was received by the Institutional Review Board of UNC-Chapel Hill (#14-1992) for the ZOE 2.0 study; all participants' guardians provided written informed consent for participation in the ZOE 2.0 study. For the Lloyd-Price and Mallick cohorts, the data we have used are publically available on line. Therefore, "ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements."

AUTHOR CONTRIBUTIONS

Conceptualization by DW and JX. Supervision by DW. Investigation by JX and LH. Formal analysis by JX. Writing—original draft preparation by JX, HC, BL, MP, KD, and DW. Reviewed draft also by FZ. Discussion with DB and FZ. All authors contributed to the article and approved the submitted version.

FUNDING

We acknowledge NIH/NIDCR R03-DE02898, NIH/NIDCR U01-DE025046, P30 CA016059 (Massey Cancer Center Support Grant) for funding support, and NLM T15 training grant #T15-LM012500 for funding support.

ACKNOWLEDGMENTS

We also acknowledge Prof. William Valdar for supporting JX.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2021.734416/full#supplementary-material

Supplementary Figure 1 | Diagnosis for outlier samples in metabolome data. The x-axis is the cumulative proportion of samples, and the y-axis is number of non-missing values. The left lower tail dots that are far from the rest may be considered as sample outliers. For ZOE 2.0 data and Lloyd-Price data, we need to remove the 10 outliers subjects from ZOE 2.0 data and 15 outliers from Lloyd-Price data to ensure the distribution of non-missing values is continuous.

Supplementary Figure 2 | Boxplot of -$\log_{10}$ of MSE for DNA, RNA, and BOTH in each of the three studies to compare ENVIM and MelonnPan.

Supplementary Figure 3 | Boxplot of -$\log_{10}$ of MSE for DNA, RNA, and BOTH in Mallick study when PRISM data was used as training to predict metabolites in NLIBD data. This is to compare ENVIM and MelonnPan.

Supplementary Figure 4 | Boxplot of -$\log_{10}$ of MSE for DNA, RNA, and BOTH in each of the three studies, for all metabolites predicted by ENVIM.

Supplementary Table 1 | Overall prediction results, for all gene family data types, all three datasets, and both methods, in Spearman correlation and MSE.

Supplementary Table 2 | The gene lists in DNA or RNA, based on the highest rank or the average rank among metabolites, that contribute to prediction of well-predicted metabolites in ZOE 2.0 by ENVIM. Rank is based on the weight matrix in ENVIM. A larger number of ranks suggests more important gene families.

REFERENCES

Aulchenko, Y. S., Ripke, S., Isaacs, A., and Van Duijn, C. M. (2007). GenABEL: An R Library for Genome-Wide Association Analysis. *Bioinformatics* 23 (10), 1294–1296. doi: 10.1093/bioinformatics/btm108

Breiman, L. (2001). Random Forests. *Mach. Learn.* 45 (1), 5–32. doi: 10.1023/A:1010933404324

Canfora, E. E., Meex, R. C. R., Venema, K., and Blaak, E. E. (2019). Gut Microbial Metabolites in Obesity, NAFLD and T2DM. *Nat. Rev. Endocrinol.* 15 (5), 261–273. doi: 10.1038/s41574-019-0156-z

Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. A., et al. (2014). The MetaCyc Database of Metabolic Pathways and Enzymes and the BioCyc Collection of Pathway/Genome Databases. *Nucleic Acids Res.* 42 (D1), D459–DD71. doi: 10.1093/nar/gkt1103

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (New York: Hillsdle. Erlbaum. Conner, BE). The Box in the Barn. Columbus: Highlights for …; 1988.

Divaris, K., and Joshi, A. (2020). The Building Blocks of Precision Oral Health in Early Childhood: The ZOE 2.0 Study. *J. Public Health Dent* 80 Suppl 1, S31–SS6. doi: 10.1111/jphd.12303

Divaris, K., Shungin, D., Rodríguez-Cortés, A., Basta, P. V., Roach, J., Cho, H., et al. (2019). The Supragingival Biofilm in Early Childhood Caries: Clinical and Laboratory Protocols and Bioinformatics Pipelines Supporting Metagenomics, Metatranscriptomics, and Metabolomics Studies of the Oral Microbiome. *Methods Mol. Biol.* 1922, 525–548. doi: 10.1007/978-1-4939-9012-2_40

Divaris, K., Slade, G. D., Ferreira Zandona, A. G., Preisser, J. S., Ginnis, J., Simancas-Pallares, M. A., et al. (2020). Cohort Profile: ZOE 2.0-A Community-Based Genetic Epidemiologic Study of Early Childhood Oral Health. *Int. J. Environ. Res. Public Health* 17 (21), 8056. doi: 10.3390/ijerph17218056

Evans, A. M., Bridgewater, B., Liu, Q., Mitchell, M., Robinson, R., Dai, H., et al. (2014). High Resolution Mass Spectrometry Improves Data Quantity and Quality as Compared to Unit Mass Resolution Mass Spectrometry in High-Throughput Profiling Metabolomics. *Metabolomics* 4 (2), 1. doi: 10.4172/2153-0769.1000132

Evans, A. M., DeHaven, C. D., Barrett, T., Mitchell, M., and Milgram, E. (2009). Integrated, Nontargeted Ultrahigh Performance Liquid Chromatography/Electrospray Ionization Tandem Mass Spectrometry Platform for the Identification and Relative Quantification of the Small-Molecule Complement of Biological Systems. *Anal Chem.* 81 (16), 6656–6667. doi: 10.1021/ac901536h

Franzosa, E. A., McIver, L. J., Rahnavard, G., Thompson, L. R., Schirmer, M., Weingart, G., et al. (2018). Species-Level Functional Profiling of Metagenomes and Metatranscriptomes. *Nat. Methods* 15 (11), 962–968. doi: 10.1038/s41592-018-0176-y

Franzosa, E. A., Sirota-Madi, A., Avila-Pacheco, J., Fornelos, N., Haiser, H. J., Reinker, S., et al. (2019). Gut Microbiome Structure and Metabolic Activity in Inflammatory Bowel Disease. *Nat. Microbiol.* 4 (2), 293–305. doi: 10.1038/s41564-018-0306-4

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models *via* Coordinate Descent. *J. Stat. Software* 33 (1), 1. doi: 10.18637/jss.v033.i01

Ginnis, J., Ferreira Zandoná, A. G., Slade, G. D., Cantrell, J., Antonio, M. E., Pahel, B. T., et al. (2019). Measurement of Early Childhood Oral Health for Research Purposes: Dental Caries Experience and Developmental Defects of the Enamel in the Primary Dentition. *Methods Mol. Biol.* 1922, 511–523. doi: 10.1007/978-1-4939-9012-2_39

Heimisdottir, L. H., Lin, B. M., Cho, H., Orlenko, A., Ribeiro, A. A., Simon-Soro, A., et al. (2021). Metabolomics Insights in Early Childhood Caries. *J. Dent. Res.* 100 (6), 615–622. doi: 10.1177/0022034520982963

Kenny, D. J., Plichta, D. R., Shungin, D., Koppel, N., Hall, A. B., Fu, B., et al. (2020). Cholesterol Metabolism by Uncultured Human Gut Bacteria Influences Host Cholesterol Level. *Cell Host Microbe* 28 (2), 245–257.e6. doi: 10.1016/j.chom.2020.05.013

Kuhn, M. (2008). Building Predictive Models in R Using the Caret Package. *J. Stat. Softw* 28 (5), 1–26. doi: 10.18637/jss.v028.i05

Lloyd-Price, J., Arze, C., Ananthakrishnan, A. N., Schirmer, M., Avila-Pacheco, J., Poon, T. W., et al. (2019). Multi-Omics of the Gut Microbial Ecosystem in Inflammatory Bowel Diseases. *Nature* 569 (7758), 655–662. doi: 10.1038/s41586-019-1237-9

Mallick, H., Franzosa, E. A., McLver, L. J., Banerjee, S., Sirota-Madi, A., Kostic, A. D., et al. (2019). Predictive Metabolomic Profiling of Microbial Communities Using Amplicon or Metagenomic Sequences. *Nat. Commun.* 10 (1), 3136. doi: 10.1038/s41467-019-10927-1

Noecker, C., Eng, A., Srinivasan, S., Theriot, C. M., Young, V. B., Jansson, J. K., et al. (2016). Metabolic Model-Based Integration of Microbiome Taxonomic and Metabolomic Profiles Elucidates Mechanistic Links Between Ecological and Metabolic Variation. *MSystems* 1 (1), e00013–e00015. doi: 10.1128/mSystems.00013-15

Sanna, S., van Zuydam, N. R., Mahajan, A., Kurilshikov, A., Vila, A. V., Võsa, U., et al. (2019). Causal Relationships Among the Gut Microbiome, Short-Chain Fatty Acids and Metabolic Diseases. *Nat. Genet.* 51 (4), 600–605. doi: 10.1038/s41588-019-0350-x

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proc. Natl. Acad. Sci.* 102 (43), 15545–15550. doi: 10.1073/pnas.0506580102

Thomas, A. M., Manghi, P., Asnicar, F., Pasolli, E., Armanini, F., Zolfo, M., et al. (2019). Metagenomic Analysis of Colorectal Cancer Datasets Identifies Cross-Cohort Microbial Diagnostic Signatures and a Link With Choline Degradation. *Nat. Med.* 25 (4), 667–678. doi: 10.1038/s41591-019-0405-7

Tsilimigras, M. C., and Fodor, A. A. (2016). Compositional Data Analysis of the Microbiome: Fundamentals, Tools, and Challenges. *Ann. Epidemiol.* 26 (5), 330–335. doi: 10.1016/j.annepidem.2016.03.002

Wei, R., Wang, J., Su, M., Jia, E., Chen, S., Chen, T., et al. (2018). Missing Value Imputation Approach for Mass Spectrometry-Based Metabolomics Data. *Sci. Rep.* 8 (1), 663. doi: 10.1038/s41598-017-19120-0

Wilkinson, J. E., Franzosa, E. A., Everett, C., Li, C., Hu, F. B., Wirth, D. F., et al. (2021). A Framework for Microbiome Science in Public Health. *Nat. Med.* 27 (5), 766–774. doi: 10.1038/s41591-021-01258-0

Zou, H., and Hastie, T. (2005). Regularization and Variable Selection *via* the Elastic Net. *J. R. Stat. Soc: Ser. B (Stat Methodol)* 67 (2), 301–320. doi: 10.1111/j.1467-9868.2005.00503.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership