NATIONAL & KAPODISTRIAN UNIVERSITY OF ATHENS

SCHOOL OF NATURAL SCIENCES
DEPARTMENT OF INFORMATICS & TELECOMMUNICATIONS

INTERDEPARTMENTAL MASTER'S PROGRAM
"DATA SCIENCE AND INFORMATION TECHNOLOGIES"

Master Thesis

"Microbial Signatures in Human Periodontal Disease: a Metatranscriptome Meta-Analysis"

Armen Ashot Ovsepian

Assessment committee: **Theodore Dalamagas, PhD, Research Director**
George Georgakilas, PhD, Scientific Associate
Artemis G. Hatzigeorgiou, Professor

ATHENS

APRIL 2024

Master Thesis

“Microbial Signatures in Human Periodontal Disease: a Metatranscriptome Meta-Analysis”

Armen A. Ovsepian
A.M.: 7115152100033

Assessment committee: Theodore Dalamagas, PhD, Research Director
George Georgakilas, PhD, Scientific Associate
Artemis G. Hatzigeorgiou, Professor

APRIL 2024

ΠΕΡΙΛΗΨΗ

Οι μελέτες μεταγονιδιωματικής και μεταγραφωματικής έχουν συμβάλει καταλυτικά στην κατανόηση των μικροβιακών κοινοτήτων του στόματος και του λειτουργικού τους δυναμικού. Παρ' όλα αυτά, οι μελέτες αυτές δεν έχουν παράξει σταθερά και ομοιόμορφα αποτελέσματα, γεγονός που υπαγορεύει την ανάγκη για συγκρίσεις μεταξύ μελετών προκειμένου να προκύψουν πιο ισχυρά ευρήματα. Εδώ, διενεργήθηκε μια μετα-ανάλυση τεσσάρων γεωγραφικά και τεχνικά διαφορετικών μελετών μεταγραφωματικής για την ανθρώπινη περιοδοντίτιδα. Συνολικά, συμπεριλήφθηκαν 54 δείγματα υποουλικής πλάκας, 27 υγιή και 27 περιοδοντίτιδας. Αντιμετωπίζοντας τη μελέτη ως μεταβλητή τυχαίας επίδρασης και με κατώφλι ρυθμού ψευδούς ανακάλυψης (False Discovery Rate - FDR 10^{-3}), εντοπίσαμε 26 είδη με διαφορετική αφθονία, τα περισσότερα από τα οποία έχουν συσχετιστεί στο παρελθόν με το στοματικό μικροβίωμα ή την περιοδοντική νόσο. Ο πυρήνας του μικροβιώματος (που ορίζεται από 80% ομαδική εμφάνιση και κορυφαία σχετική αφθονία 25%) της ομάδας των υγιών και της ομάδας της περιοδοντίτιδας περιελάμβανε 40 και 80 είδη, αντίστοιχα. Αξιοσημείωτο είναι ότι 38 είδη του πυρήνα του υγιούς μικροβιώματος ήταν κοινά με τον πυρήνα της ομάδας περιοδοντίτιδας. Η λειτουργική ανάλυση έδειξε ότι 50 οικογένειες γονιδίων (UniRef-90) που εμπλέκονται στη διαμεμβρανική μεταφορά και έκκριση, στο μεταβολισμό των αμινοξέων, στη σύνθεση επιφανειακών πρωτεϊνών και μαστιγίων, στο μεταβολισμό της ενέργειας και στην υπερελίκωση του DNA, μεταγράφηκαν σημαντικά περισσότερο ($FDR \leq 10^{-2}$) στα δείγματα περιοδοντίτιδας. Επιπλέον, 4 γονίδια βακτηριακών παραγόντων μολυσματικότητας, από τη Virulence Factor DataBase (VFDB), συμπεριλαμβανομένου του εξαρτώμενου υποδοχέα TonB από το *P. gingivalis*, του επιφανειακού αντιγόνου BspA από το *T. forsynthia,* και της αδρεψίνης A (PsaA) και της γλυκεραλδεΰδη-3-φωσφορικής αφυδρογονάσης τύπου I (GAPDH) από το γένος *Streptococcus*, βρέθηκαν επίσης να μεταγράφονται σημαντικά περισσότερο ($FDR \leq 0,05$) στην ομάδα περιοδοντίτιδας. Τα ευρήματα αυτά προωθούν την κατανόηση του στοματικού μικροβιώματος και του μεταγραφώματος στην υγεία και την περιοδοντίτιδα.

Θεματική περιοχή: Μεταγραφωματική ανάλυση

Λέξεις κλειδιά: Περιοδοντίτιδα, μεταγραφωματική, μέτα-ανάλυση, δίκτυο μικροβίων, μικροβίωμα-πυρήνας, στοματικό μικροβίωμα, δυσβίωση

SUMMARY

Metagenomics and metatranscriptomics studies have reshaped our understanding of oral microbial communities and their functional potential. Nevertheless, these studies have not consistently produced uniform results, thus prompting the need for cross-study comparisons to obtain more robust findings. Here, a meta-analysis of four geographically and technically diverse oral shotgun metatranscriptomics studies of human periodontitis was performed. In total, 54 subgingival plaque samples, 27 healthy and 27 periodontitis, were included. By treating the study as a random effect variable and with a false discovery rate (FDR) threshold of 10^{-3}, we identified 26 differentially abundant species, most of which have previously been associated with oral microbiome or periodontal disease. The core microbiota (defined by 80% group occurrence and top 25% relative abundance) of the healthy and periodontitis group comprised 40 and 80 species, respectively. Notably, 38 species of the healthy core microbiota were shared with the periodontitis group's core. Functional analysis showed that 50 gene families (UniRef-90) involved in transmembrane transport and secretion, amino acid metabolism, surface protein and flagella synthesis, energy metabolism, and DNA supercoiling, were significantly more transcribed (FDR $\leq 10^{-2}$) in periodontitis samples. Additionally, 4 bacterial virulence factor genes, from Virulence Factor DataBase (VFDB), including TonB-dependent receptor from *P. gingivalis*, surface antigen BspA from *T. forsynthia*, and adhesin A (PsaA) and Type I glyceraldehyde-3-phosphate dehydrogenase (GAPDH) from the *Streptococcus* genus, were also found to be significantly more transcribed (FDR ≤ 0.05) in periodontitis group. These findings promote our understanding of the oral microbiome and transcriptome in health and periodontitis.

Subject area: Metatranscriptomics analysis

Keywords: Periodontitis, metatranscriptomics, Meta-analysis, microbial network analysis, microbiome, core-microbiome, Oral microbiome, Dysbiosis

ACKNOWLEDGEMENTS

I would like to thank Prof. Artemis Hatzigeorgiou (University of Thessaly) for giving me the opportunity to conduct the study in her group. Her scientific guidance and the assistance of the laboratory members were invaluable for this work. Special thanks to Anargyros Skoulakis, Filippos Kardaras, Marios Miliotis and Spyros Tastsoglou.

I would also like to sincerely thank the assessment committee members Theodore Dalamagas, PhD, Research Director, and George Georgakilas, PhD, Scientific Associate from Athena Research Center for reading and evaluating this project.

CONTENTS

LIST OF FIGURES ... 7
LIST OF TABLES ... 9
PREFACE ... 10
1 Introduction ... 11
1.1 Human oral microbiota ... 11
1.2 Periodontitis ... 12
1.3 Study of the oral microbiome with metagenomics and metatranscriptomics ... 14
1.4 Normalization of metagenomics data ... 15
1.5 Meta-analysis ... 17
2 Materials and methods ... 18
2.1 Study inclusion and data acquisition ... 18
2.2 Data preprocessing ... 19
2.3 Taxonomic profiling, functional analysis and virulence factor genes detection ... 19
2.4 Data handling ... 19
3 Results ... 20
3.1 Studies inclusion ... 20
3.2 Microbial diversity (alpha- and beta-diversity) ... 20
3.3 Differentially abundant species and genera between healthy and periodontitis ... 23
3.4 Core microbiota ... 23
3.5 Differentially transcribed gene families (UniRef90) and virulence factors ... 25
4 Discussion ... 26
5 Code availability ... 30
Appendix 1 ... 31
References ... 42

LIST OF FIGURES

Fig. 1. Major habitats within the mouth. (A) Supragingival plaque, subgingival plaque, keratinized gingiva. (B) Hard palate, buccal mucosa; higher-magnification images show micro-habitats on buccal mucosa. Buccal mucosa smear modified from (Datar et al., 2013). Microplicae on buccal squamous cell modified from (Kullaa et al., 2014). (C) Palatine tonsils, throat, tongue dorsum; higher-magnification images show micro-habitats on tongue dorsum. Tongue diagram modified from Rice University's OpenStax (*https://openstax.org/books/anatomy-and-physiology/pages/23-3-the-mouth-pharynx-and-esophagus*); filiform papillae modified from *https://histology.medicine.umich.edu/*; hairs of filiform papillae modified from (Kullaa-Mikkonen & Sorvari, 1985). Adapted from (Mark Welch et al., 2020). 13

Fig. 2. The stages of periodontal disease. Source: https://www.divadentistry.com/periodontal-disease-stages/ 14

Fig. 3. Scanning electron micrograph of subgingival dental plaque on the surface of a tooth extracted due to periodontal disease. Microbial cells are connected by a meshwork of fibrous material (M), apparently the collapsed remains of a hydrated polymeric matrix. Some microbial cells are associated with small particles or vesicles (V) that may provide a source for matrix polymers. Adapted from (Jakubovics et al., 2021). .. 16

Fig. 4. Illustration of the meta-analysis workflow integrating metatranscriptomics data from four studies. 22

Fig. 5. Microbial species diversity. (**A**) Box plot of α-diversity (Shannon index) in healthy (blue) and periodontitis (red) groups of all studies combined. For statistical analysis, blocked Wilcoxon test was used to treat "study" as a blocking factor. The p-value is shown on the top of the figure. (**B**) Box plot of α-diversity in healthy (blue) and periodontitis (red) groups in each individual study. The p-values of the two-sided Wilcoxon tests are shown at the top of the figures. (**C**) Principal coordinate analysis (PCoA) with Bray-Curtis dissimilarity matrix of the CSS normalized species abundances. 22

Fig. 6. Differentially abundant genera and species between healthy (blue) and periodontitis (red) groups identified by MaAsLin2. (**A**) The 25 differentially abundant genera (FDR $\leq 10^{-3}$). (**B**) The 26 differentially abundant species (FDR $\leq 10^{-3}$). The log of CSS normalized abundance is presented. The FDR values are shown on the right side of the bar plot. The statistical significance in each individual study included in the meta-analysis is also shown. A white fill indicates an FDR > 0.05, a grey fill indicates an FDR ≤ 0.05, and a black fill indicates FDR $\leq 10^{-3}$. 24

Fig. 7. Differentially transcribed gene families (UniRef90, taxonomy) between healthy (blue) and periodontitis (red) health groups identified by MaAsLin2. The normalized abundance of the 52 gene families, with FDR $\leq 10^{-2}$, are presented. The FDR are shown on the right side of the bar plot. 27

Fig. 8. Differentially transcribed virulence factors between healthy (blue) and periodontitis (red) health groups. The log of normalized abundance of the 4 virulence factor genes with FDR ≤ 0.05 are presented. The name of the genes and the NCBI accession (in parenthesis) are shown. The FDR are shown on the right side of the bar plot. The statistical significance in each individual study included in the meta-analysis is also shown. White fill indicates FDR > 0.05, while black fill indicates FDR ≤ 0.05. 27

Fig. S1. The variance explained by the confounding factors (normalized library sizes, smoking status and study origin) against the variance explained by the oral health status, with (**A**) and without (**B**) setting the study factor as a random effect. The analysis of the confounding factors for species abundance was conducted following the methodology described by (Wirbel et al., 2019). Briefly, the total variance within the abundance of a given microbial species was compared to the variance explained by disease status and the variance explained by the confounding factor, with a linear model assumption. Variance calculations were performed with a non-parametric approach to account for non-Gaussian distribution of microbiome abundance data. The library size potential confounder, with continuous values, was transformed into categorical data as quartiles. The size of the spots represents the trimmed mean abundance. The species that have been found to have statistically significant abundance (qval ≤ 10^{-3}), after differential analysis with MaAsLin2, are shown in red. 35

LIST OF TABLES

Table 1. Sampling details of each study included in the meta-analysis. 21

Table 2. Bacterial species that are part of the core microbiota of healthy and/or periodontitis group, which have been found to be overrepresented in periodontitis. Species in bold were part of the core microbiota of both health status groups, while the rest of the species were only part of the periodontitis core microbiota. The mean relative abundance % of the species in each health group, and the FDR is also shown. 25

Table S1. Studies that were considered for our meta-analysis. The study details, the decision on inclusion outcome (YES/NO) and the justification of exclusion are shown. 31

Table S2. The 61 differentially abundant species between healthy and periodontitis groups identified by MaAsLin2 (FDR > 0.05). 36

Table S3. The 82 bacterial species that are part of the core microbiota of healthy and/or periodontitis group. 38

Table S4. The 81 differentially transcribed gene families (UniRef90, taxonomy) between healthy and periodontitis health groups identified by MaAsLin2 (FDR ≤ 10^{-2}). 40

PREFACE

This thesis is part of the interdepartmental Master's study programme "Data Science and Information Technologies" organized by the Department of Informatics & Telecommunications, National & Kapodistrian University of Athens in collaboration with the Biomedical Research Foundation of the Academy of Athens (BRFAA), and the Athena Research Center. This dissertation was submitted to the secretary of postgraduate study programmes, Department of Informatics & Telecommunications, National & Kapodistrian University of Athens, as part of my assessment for the acquisition of MSc degree. The work presented hereby was conducted between December 2022 and December 2023. The assessment committee consisted of Theodore Dalamagas, Research Director, and George Georgakilas, Scientific Associate from Athena Research Center, and Professor Artemis Hatzigeorgiou from the University of Thessaly.

In this document, the summary of the study is initially presented. Thereafter, introductory information on the human oral microbiome and periodontal disease is provided. Special focus is given on metagenomics and metatranscriptomics of the oral microbiome, normalization methods and meta-analysis. Finally, the obtained results and the discussion are presented.

1 Introduction

1.1 Human oral microbiota

More than three centuries ago, Antony van Leeuwenhoek was one of the first to observe microbes, possibly bacteria from his own dental plaque, with the use of a microscope that he had designed (Huang et al., 2021; Radaic & Kapila, 2021).

Humans have co-evolved with trillions of microbes residing within our bodies. These microbes create intricate niche-specific ecosystems that are finely attuned with their environmental conditions and highly adaptable to the dynamic host physiology (Lloyd-Price et al., 2016). The total number of bacteria in the 70 kg "reference man" is estimated to be 3.8 x 10^{13}, accounting for about half of the total cells on a human body (Sender et al., 2016). Our understanding of the human microbiome is largely based on the gut microbiome, which is the most studied case, but the established principles could generally be applied to microbial habitats throughout the body (Lloyd-Price et al., 2016). The microbiome is involved in various important processes in our body. In the gut, commensal microorganisms can regulate the biosynthetic processes of micronutrients and they can modulate their absorption (Barone et al., 2022; Rowland et al., 2018; Valdes et al., 2018). Additionally, they can play a key role in the host defense against pathogens (Iacob et al., 2019), and the development and function of the immune system (Graham & Xavier, 2023; Lambring et al., 2019; Zheng et al., 2020).

The first investigations into the ecology of the microbiome sought to identify a 'core' set of microorganisms or prevalent molecular pathways that were universally present in healthy individuals, and attempted to relate these to the health status (Turnbaugh et al., 2007). However, microbiomes regularly show high levels of interpersonal diversity even in the absence of disease (Huttenhower et al., 2012; Lloyd-Price et al., 2016). This makes it difficult to directly link microbial components or imbalances to disease states (Lloyd-Price et al., 2016). Nevertheless, in recent decades, a number of studies have documented profound changes in the microbial community structures between healthy and diseased study groups in various diseases, such as in inflammatory bowel disease (IBD) (Frank et al., 2007) and diabetes (Karlsson et al., 2013). These alterations of the microbiome are defined as 'dysbiosis' and it is speculated that they could be major contributing factors to the initiation and/or persistence of many of these diseases (Petersen & Round, 2014).

The oral cavity is an open system and microbes are inhaled with every breath, ingested with every meal or drink, or introduced from our immediate physical surroundings (Mark Welch et al., 2020). The warm, moist, and nutrient-rich environment of the mouth makes it a relatively hospitable environment for microbes (Mark Welch et al., 2020). The oral microbiome is one of the most complex and diverse microbial communities in the human body and it has been related to oral and systemic health (Huang et al., 2021). The oral sub-habitats include the buccal mucosa, the tongue dorsum and the hard structures of the teeth, which are comprised by those above (supragingival) and below (subgingival) the gingival margin (Mark Welch et al., 2020; Xu et al., 2015) (**Fig. 1**). Despite the proximity of the oral sub-habitats, the distinct ecological conditions that prevail in each of them create special niches and they contain markedly different

microbial communities (Aas et al., 2005; Mager et al., 2003). The microbiome of the oral cavity comprises bacteria, fungi, viruses and archaea, with bacteria playing a predominant role and being the most studied (Radaic & Kapila, 2021). Cultivation-independent molecular methods have identified more than 700 microbial species in the human oral microbiome, over 60% of which have not been cultivated (Aas et al., 2005; Dewhirst et al., 2010). A total of 141 different bacterial taxa representing six different bacterial phyla have been detected. The six phyla include the *Firmicutes* (e.g. *Streptococcus*, *Gemella*, *Eubacterium*, *Selenomonas* and *Veillonella*), the *Actinobacteria* (*e.g. Actinomyces*, *Atopobium* and *Rothia*), the *Proteobacteria* (*e.g. Neisseria*, *Eikenella* and *Campylobacter*), the *Bacteroidetes* (*e.g. Porphyromonas*, *Prevotella* and *Capnocytophaga*), the *Fusobacteria* (*e.g. Fusobacterium* and *Leptotrichia*), and the TM7 phylum, for which there are no cultivable representatives (Aas et al., 2005).

The microorganisms of the oral microbiome are entangled in a wide range of interspecies interactions, which include synergistic, signaling, or antagonistic interactions and polymicrobial biofilm formation (Diaz & Valm, 2020; Jakubovics et al., 2021; Radaic & Kapila, 2021). Microorganisms within the oral microbiome have been associated with a range of oral infectious diseases, including caries (tooth decay), periodontitis (gum disease), endodontic (root canal) infections, alveolar osteitis (dry socket), and tonsillitis (Dewhirst et al., 2010).

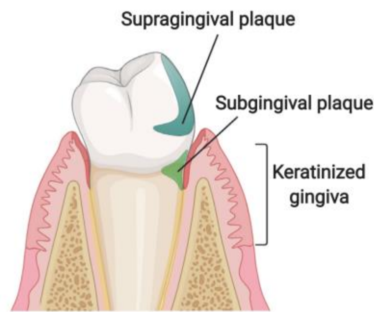
1.2 Periodontitis

Qualitative and/or quantitative shifts of the oral microbiome lead to microecological dysbiosis, an imbalance responsible for the development of periodontitis, which is a chronic oral inflammatory disease that progressively destroys the supporting periodontal tissues (Huang et al., 2021; Jakubovics et al., 2021; Morillo-Lopez et al., 2022). Dental plaque biofilms play an essential role in the initiation and progression of periodontitis (Huang et al., 2021; Morillo-Lopez et al., 2022). Without proper treatment, this oral infection can ultimately result in tooth loss (Huang et al., 2021). During the initial phase of the periodontal disease, known as gingivitis, the gums typically swell, become red, and may bleed easily. In its advanced stage, termed periodontitis, the gums can recede from the teeth, resulting in bone loss and potential loosening or loss of teeth (**Fig. 2**).

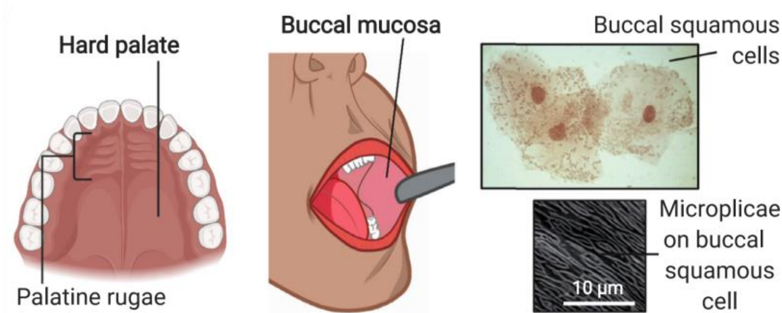
The prevalence of periodontitis is high, with approximately 10% of the global population being affected by severe periodontitis (Frencken et al., 2017). The predicted rise in retention procedures aimed at stabilizing tooth position, coupled with an aging population, may contribute to an increase in the number of patients affected by periodontitis in the future (Duran-Pinedo et al., 2023).

Moreover, a growing body of evidence indicates that periodontal diseases can elevate the risk of various systemic conditions such as diabetes (Genco et al., 2020), cardiovascular disease (Van Dyke et al., 2021), respiratory diseases (Molina et al., 2023), Alzheimer's disease (Dominy et al., 2019), and cancer (Nwizu et al., 2020). Thus, a better understanding of periodontitis progression can enhance both oral and systemic well-being.

A



B



C

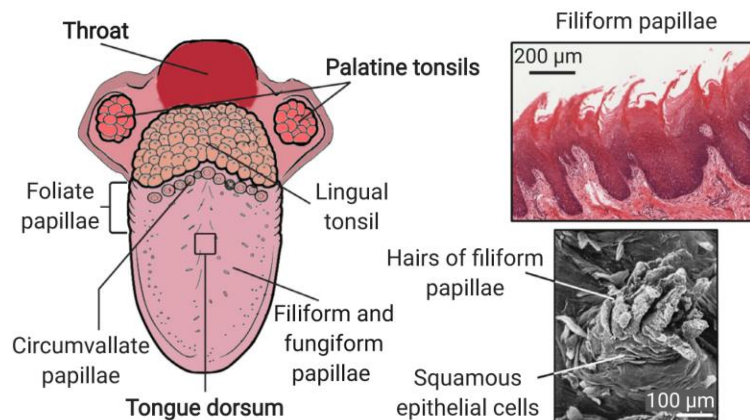


Fig. 1. Major habitats within the mouth. (A) Supragingival plaque, subgingival plaque, keratinized gingiva. (B) Hard palate, buccal mucosa; higher-magnification images show micro-habitats on buccal mucosa. Buccal mucosa smear modified from (Datar et al., 2013). Microplicae on buccal squamous cell modified from (Kullaa et al., 2014). (C) Palatine tonsils, throat, tongue dorsum; higher-magnification images show micro-habitats on tongue dorsum. Tongue diagram modified from Rice University's OpenStax (*https://openstax.org/books/anatomy-and-physiology/pages/23-3-the-mouth-pharynx-and-esophagus*); filiform papillae modified from *https://histology.medicine.umich.edu/*; hairs of filiform papillae modified from (Kullaa-Mikkonen & Sorvari, 1985). Adapted from (Mark Welch et al., 2020).

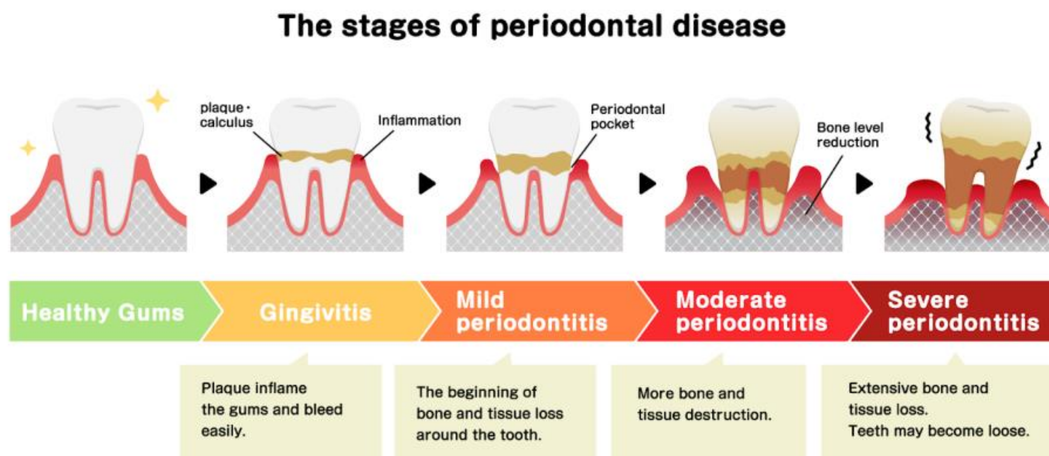


Fig. 2. The stages of periodontal disease. Source: https://www.divadentistry.com/periodontal-disease-stages/

1.3 Study of the oral microbiome with metagenomics and metatranscriptomics

Most microbial species in the human microbiome, including the oral microbiome, are non-culturable (Aas et al., 2005). The advent of massively parallel shotgun sequencing (high-throughput sequencing technologies) has substantially resolved the taxonomic composition of this microbial "dark matter" (Lloyd-Price et al., 2016). As a result, our understanding of composition of the human microbiome has undergone a profound transformation. Metagenomics and, more recently, metatranscriptomics studies have been employed to better understand periodontitis and unravel the molecular mechanisms involved in the disease (Huang et al., 2021). Metagenomics offers information on genes or microorganisms that are or have been present in the community. Metatranscriptomics could provide a broader perspective than metagenomics, as it can reveal details about transcriptionally active populations, functional characteristics of complex microbial communities in health or disease, and elucidate potential intervention targets (Huang et al., 2021; Ojala et al., 2023). Additionally, total RNA-seq has been shown to be more accurate in microbial identification accuracy than metagenomics, at equal sequencing depths and even at sequencing depths almost one order of magnitude lower than those of metagenomics (Hempel et al., 2022).

Most of the dental plaque materials originate from microbial cells and they are actively secreted into the surrounding environment. The major classes of these extracellular polymeric substances comprise carbohydrates, proteins, nucleic acids, and cell wall polymers, such as peptidoglycans and lipids (Flemming & Wingender, 2010). Within biofilms, approximately 10% of the dry mass consists of microorganisms, while the extracellular matrix may make up more than 90% (Flemming & Wingender, 2010). Furthermore, host cells that have undergone degradation, along with dietary components like fiber, polysaccharides, and proteins from the host's diet, may become embedded within dental plaque (Jakubovics et al., 2021). Lastly, microbial vesicles and

bacteriophages could also play important role in the ecology of biofilms (Jakubovics et al., 2021). In **Fig. 3**, a scanning electron micrograph of a subgingival dental plaque sample is shown.

During periods of gingival health, Gram-positive organisms such as *Actinomyces* and *Sterptococcus* are prevalent within the plaque biofilm. As plaque biofilm formation progresses, it undergoes a maturation process, marked by a transition towards Gram-negative anaerobes and motile organisms (Ezzo & Cutler, 2003).

Saliva and dental plaque (*i.e.* bacterial biofilm on the teeth) are readily accessible, making them ideal for investigating the onset and progression of periodontitis (Lasserre et al., 2018). Bacterial species that metagenomic and metatranscriptomics studies have associated with periodontal disease include *Porphyromonas gingivalis*, *Tannerella forsythia, Filifactor alocis*, *Prevotella intermedia*, *Treponema socranskii, Treponema denticola*, *Campylobacter rectus, Campylobacter showae, Fusobacterium nucleatum*, *Osenella uli*, *Parvimonas micra*, *Fretibacterium fastidiosum*, *Lactobacillus gasseri*, *Aggregatibacter actinomycetemcomitans*, and *Actinobacillus actinomycetemcomitans* (Huang et al., 2021)

Interestingly, studies have also shown significant overlap in the composition of microbial communities associated with healthy and periodontal disease (Huang et al., 2021; Jorth et al., 2014). This suggests that differences in periodontal status between sites can't be solely attributed to variations in the composition of the subgingival microbiome, and that microbial activity may play a critical role in driving disease progression (Duran-Pinedo et al., 2023; Huang et al., 2021).

Through metatranscriptomic methodology, increased levels of iron acquisition, lipopolysaccharide synthesis, amino acid catabolism, flagella synthesis, and TonB-dependent receptors, and reduced levels of carbohydrate metabolism, have been connected to periodontitis (Belstrøm et al., 2021; Jorth et al., 2014; Szafrański et al., 2015; Yost et al., 2015).

While metagenomics and metatranscriptomics have revolutionized our understanding of oral microbial communities and their functional potential, a number of important questions remain to be addressed. For instance, how do alterations in the oral microbial composition intricately contribute to the pathogenesis of periodontitis? Could the disease progression be attributed to the structural shifts of the entire microbial network, or to distinct components within the oral microbiota? Which are the most important microbial interactions during dental plaque development and disease progression?

1.4 Normalization of metagenomics data

Metagenomics data is compositional in nature as the identified experimental values are proportions of a whole. In the context of metagenomic data, each sample represents a microbial community, and the data typically consists of counts or relative abundances of different taxa (*e.g.* bacteria, archaea, viruses) or functional elements (*e.g.* genes) within that community. Therefore, the main aim of normalization is to adjust the observed data, ensuring that discrepancies in the absolute abundances between two

microbiomes are not influenced by differences in sampling fractions (Lin & Peddada, 2020).

The most common normalization technique, total-sum scaling (TSS), divides feature read counts by the total number of reads in each sample. TSS has been shown to bias differential abundance estimates in RNA-seq data (Bullard et al., 2010; Dillies et al., 2013) because certain measurements (*e.g.* species or genes) are disproportionately sampled as sequencing yield rises, leading them to exert an unwarranted influence on normalized counts. The false discovery rate (FDR) generated from TSS-based analyses has been reported to be unacceptably large (Lin & Peddada, 2020).

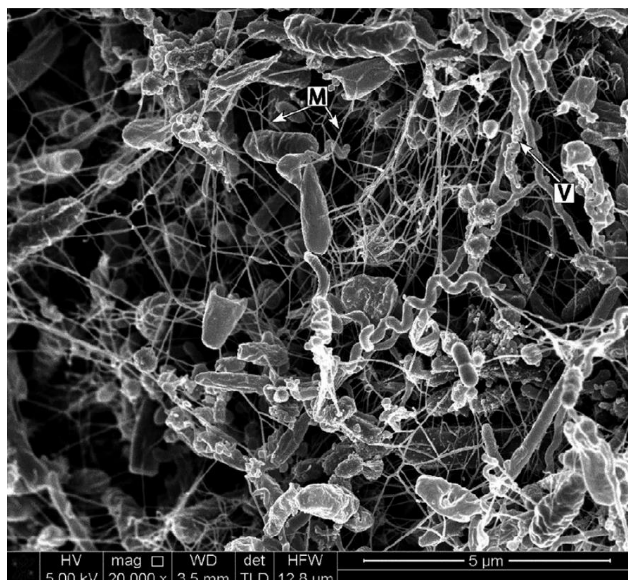


Fig. 3. Scanning electron micrograph of subgingival dental plaque on the surface of a tooth extracted due to periodontal disease. Microbial cells are connected by a meshwork of fibrous material (M), apparently the collapsed remains of a hydrated polymeric matrix. Some microbial cells are associated with small particles or vesicles (V) that may provide a source for matrix polymers. Adapted from (Jakubovics et al., 2021).

Other widely used normalization methods include the median normalization (MED) method used in DESeq2 (Love et al., 2014) and the Upper Quartile normalization (UQ) and the trimmed mean of M-values (TMM) used in edgeR (Robinson et al., 2009). In DESeq2, for each gene, the ratio of its count to the geometric mean of counts across all samples is calculated. Then, the median of these ratios is used as the normalization factor for each sample. MED assumes that the taxon of median absolute abundance is not differentially abundant, which may be a valid assumption in gene expression studies, but frequently it is not in microbiome studies (Lin & Peddada, 2020). EdgeR deals with outliers by not including (trimming) genes with extreme counts (both lower and higher) in the calculations of the scaling factor, and it is also based on the assumption that most genes are not differentially abundant. Overall, DESeq2 and edgeR have displayed satisfactory

behavior in the differential analysis of both real Illumina data and simulations (Dillies et al., 2013). In particular, in the presence of high-count genes, DESeq and TMM were able to maintain a relatively low false-positive rate without any loss of power.

An alternative normalization method is to scale across only the segment of the count distribution that is relatively invariant across samples (Dillies et al., 2013). In the cumulative-sum scaling (CSS) method employed by the metagenomeSeq R package, raw counts are divided by the cumulative sum of counts up to a specific quantile, under the assumption that sample abundances are independently and identically distributed up to this quantile (Dillies et al., 2013; Lin & Peddada, 2020). The quantiles are determined adaptively in a data-driven way, which relies on the change point of the distribution of cumulative sum switching from stability to instability (Lin & Peddada, 2020).

CSS has been reported to perform well compared to DESeq2 and edgeR in evaluations with simulated and real published microbiota data sets (Paulson et al., 2013) and it was the normalization method of choice in our meta-analysis.

1.5 Meta-analysis

The results of individual metagenomics and metatranscriptomics studies are often insufficient to provide confident answers, as they are not consistently reproducible (Armour et al., 2019; Huang et al., 2021). This situation is further impaired by the lack of standards in metagenomics and metatranscriptomics data generation and processing (Costea et al., 2017) and the relatively low number of samples used. These limitations of individual studies and the overwhelming increase in the amount of medical data generated, make it almost impossible to keep up to date with the scientific literature and to draw general conclusions (Walker et al., 2008).

Meta-analysis refers to the statistical technique of combining results from two or more studies to improve statistical power (Lee, 2018; Tufanaru et al., 2015). Furthermore, meta-analyses could investigate the source of variation and the different effects among subgroups (Lee, 2018). Overall, meta-analyses provide less biased estimates on clinical issues, and they are the best tools in evidence-based medicine (Lee, 2018).

According to (Walker et al., 2008) the critical factors that could affect the quality of a meta-analysis are the following:

1. Identification and selection of studies. The publication bias (*i.e.* the higher probability of reporting “positive” results) and the search bias (*i.e.* choosing which studies to be included) could highly influence the results.

2. Heterogeneity of results among individual studies. The degree of dissimilarity in the results of individual studies can either be traced back to inherent differences or might not be easily elucidated. As the level of heterogeneity increases, the justification for an integrated result becomes more difficult.

3. Availability of information. Most reports of individual studies include only summary results, such as means, standard deviations, proportions, odds ratios, and relative risks.

The lack of information can severely limit the type of analyses and conclusions that can be reached in a meta-analysis.

4. Analysis of the data. The statistical and bioinformatics tools that are used to analyze the data could affect the results. Importantly, the data from the individual studies can be analyzed using either a fixed-effect or a random-effect model. The fixed-effect model assumes that all studies included in the meta-analysis share a single true effect size, with any observed variation among studies attributed to sampling errors or chance (Lee, 2018). Conversely, the random-effects model assumes significant diversity among studies, acknowledging that the true effect size may vary from one study to another (Lee, 2018). Under the fixed-effects model, individual study results are combined using weights based on each study's sample size, whereas in the random-effects model, each study carries equal weight (Walker et al., 2008). Consequently, while the fixed-effect model evaluates only intra-study sampling errors (intra-study variation), the random-effects model accounts for both intra-study sampling errors and inter-study variance (Lee, 2018).

2 Materials and methods

2.1 Study inclusion and data acquisition

PubMed was used to retrieve studies containing oral shotgun metatranscriptome data from subgingival samples in both patients with periodontitis and healthy individuals. The search terms were “periodontitis" AND "metatranscriptome”, and results were collected up until June 1, 2023. In total, 48 articles were identified, 11 of which were review articles and were excluded from the analysis. An examination of the references cited in these review articles led to the addition of 1 more article (Duran-Pinedo et al., 2014) to our list. All search hits, except review articles, along with the additional study dataset, and the justification for exclusion or inclusion in our study are available in **Appendix 1**, **Supplementary Table S1.** Ultimately, the datasets from four articles were included in the meta-analysis and periodontitis samples from all studies had pocket depth (PD) ≥ 4.0 mm and clinical attachment loss (CAL) > 3 mm. The data were downloaded from either the National Library of Medicine (NCBI) Sequence Read Archive (SRA) or the Human Oral Microbiome Database (HOMD). The BioProject accession numbers were PRJNA678453 (Belstrøm et al., 2021) and PRJNA221620 (Jorth et al., 2014). The data from HOMD had the submission numbers 20141024 (Yost et al., 2015) and 20130522 (Duran-Pinedo et al., 2014). The Belstrøm et al., 2021 data was produced in an Illumina HiSeq2500 platform at a read length of 2 x 100 bp, the Yost et al., 2015 and Duran-Pinedo et al., 2014 data in an Illumina MiSeq v2 at read lengths of 2 x 75 and 2 x 150 bp, respectively, and the Jorth et al., 2014 data in an Illumina HiSeq2000 with 50-bp single-end reads length.

2.2 Data preprocessing

The Illumina 3′ adapters (5’AGATCGGAAGAGCACACGTCTGAACTCCAGTCA3’ from read 1, 5’AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT3’ from read 2 - if applicable -, and 5’CTGTCTCTTATACACATCT3’ from both read 1 and read 2 were removed from the sequencing reads by Atropos (Didion et al., 2017). A quality cutoff of 20 was used to trim low-quality 3' ends from reads before adapter removal. Then, rRNA reads were eliminated from the data by using sortmeRNA (Kopylova et al., 2012) version 2.1 and the default parameters. Human reads were removed from the data by aligning reads to the human genome GRCh38, NCBI RefSeq assembly GCF_000001405.40. For the alignment, HISAT2 (Kim et al., 2015) v.2.2.1 was employed and pairs that did not align concordantly were considered reads of microbial origin.

2.3 Taxonomic profiling, functional analysis and virulence factor genes detection

Taxonomic profiling of metatranscriptomes was performed with AGAMEMNON (Skoufos et al., 2022). For the creation of the database, all the reference and representative bacterial genomes (4264 in total), all complete archaeal (492) and viral (11284) genomes, available as of 31st of January 2023, and all available fungal transcriptomes (489), as of 24th of March 2023, were retrieved from the NCBI RefSeq database (O’Leary et al., 2016).

For the identification of bacterial virulence factors, the bacterial virulence factor full DNA dataset (VFDB) (Liu et al., 2022) was downloaded on the 13th of February, 2023. Then, preprocessed reads were mapped to the database with Bowtie 2 (Langmead & Salzberg, 2013). From the generated SAM files, samtools were used to extract the virulence factor reference names and the number of reads that were uniquely mapped to each of them.

HUMAnN 3 (Beghini et al., 2021) was used for functional analysis. Briefly, preprocessed reads were mapped using Bowtie2 to the full chocophlan pangenomes (version: v201901_v31). Unaligned reads were then blasted against UniRef90 (version: uniref90_201901b) using DIAMOND (Buchfink et al., 2015). Counts were assigned to gene families and were normalized for length (reads per kilobase).

2.4 Data handling

α-diversity and β-diversity were assessed using the Shannon diversity index and Bray-Curtis dissimilarity, respectively. These calculations were performed on data that had been normalized using the Cumulative Sum Score (CSS) method. The CSS normalization (Paulson et al., 2013) was executed in R, utilizing the “metagenomeSeq” package (version 1.40.0). To evaluate the significance of differences in Shannon diversities between the healthy group and the periodontitis group, a blocked Wilcoxon test was employed. This test was implemented using the R “coin” package (Hothorn et

al., 2006). Principal Coordinates Analysis (PCoA) was performed on the Bray-Curtis dissimilarity matrix to visualize and compare the microbial communities across samples and groups.

The core microbiota was identified based on two key criteria: species occurrence within the study group and in-sample relative abundance. Specifically, a species was included in the core microbiota if it met the following conditions: (a) the species had to be present in at least 80% of the samples, reflecting a high degree of occurrence within the study group, and (b) the species had to be within the top 25% of in-sample species relative abundances. These stringent criteria ensure that the core microbiota comprises species that are both prevalent across samples and abundant within individual samples.

For differential abundance analysis, MaAsLin2 (Mallick et al., 2021) was used. For microbial species, gene families and virulence factors, the study was set as a random effect and a negative binomial regression model was applied to the CSS normalized counts data. The prevalence threshold (min_prevalence parameter) for the microbial species was set at 0.5 (*i.e.*, 50% prevalence), while for gene families and the virulence factor genes at 0.25. The Benjamini-Hochberg (BH) method was used for multiple hypothesis correction and estimation of q-values (FDR values).

3 Results

3.1 Studies inclusion

In this meta-analysis, four studies containing human oral metatranscriptome data with healthy and periodontitis subgingival plaque samples, were used. All studies, as well as the type and the number of samples used, can be found in **Table 1**. In total, 54 subgingival plaque samples, 27 from healthy and 27 from periodontitis individuals, were used. The studies that were initially considered to be included in the analysis, along with the basis of exclusion of the excluded studies can be found in **Appendix 1**, **Supplementary Table S1.** The workflow of our meta-analysis is presented in **Fig. 4.**

3.2 Microbial diversity (alpha- and beta-diversity)

After preprocessing of the raw RNA-seq data, the abundance of microbial species was determined by AGAMEMNON by using an extensive database containing bacterial, archaeal, viral and fungal reference genomes. The microbial α-diversity, as measured by Shannon index, was not significantly different (p-value > 0.05) between the healthy and periodontitis study groups (**Fig. 5A**). This was also the case in each individual study of the meta-analysis, except for Yost et al. 2015, where the periodontitis group displayed greater α-diversity than the healthy group (**Fig. 5B**).

Table 1. Sampling details of each study included in the meta-analysis.

Study	Sample inclusion and definition of periodontitis	Number of samples	
		Healthy	Periodontitis
Duran-Pinedo et al., 2014	All healthy and periodontitis samples. Healthy sites had PD < 3 mm while periodontitis sites had PD > 5 mm and CAL > 3 mm.	6	7
Jorth et al., 2014	All healthy and periodontitis samples. Periodontitis sites had PD ≥ 5 mm and CAL ≥ 5 mm.	3	3
Yost et al., 2015	Samples from stable tooth sites from first visit (0 months) were defined as healthy and samples from progressive site from last visit (2 months) were defined as periodontitis (PD ≥ 4.0 mm and CAL ≥ 4 mm).	8	8
Belstrøm et al., 2021	All healthy and periodontitis samples. Periodontitis sites displayed PD ≥ 6 mm and CAL ≥4 mm.	10	9

PD=pocket depth, CAL=clinical attachment loss

The PCoA using Bray-Curtis dissimilarity matrix on CSS normalized species abundance data, can be seen in **Fig. 5C**. The samples on the PCoA tended to cluster together based on their study origin rather than their dental health status, suggesting a significant impact of the “study” factor on species composition. To further investigate the effect of the confounding factors, the proportion of the total variance that can be attributed to each confounding factor (study, normalized library size, smoking status) and the variance explained by dental health status was quantified (**Appendix 1**, **Supplementary Fig. S1)**. The analysis showed that the “study” factor had the strongest effect on the microbial species composition. This was expected, since the studies varied in both biological and technical aspects.

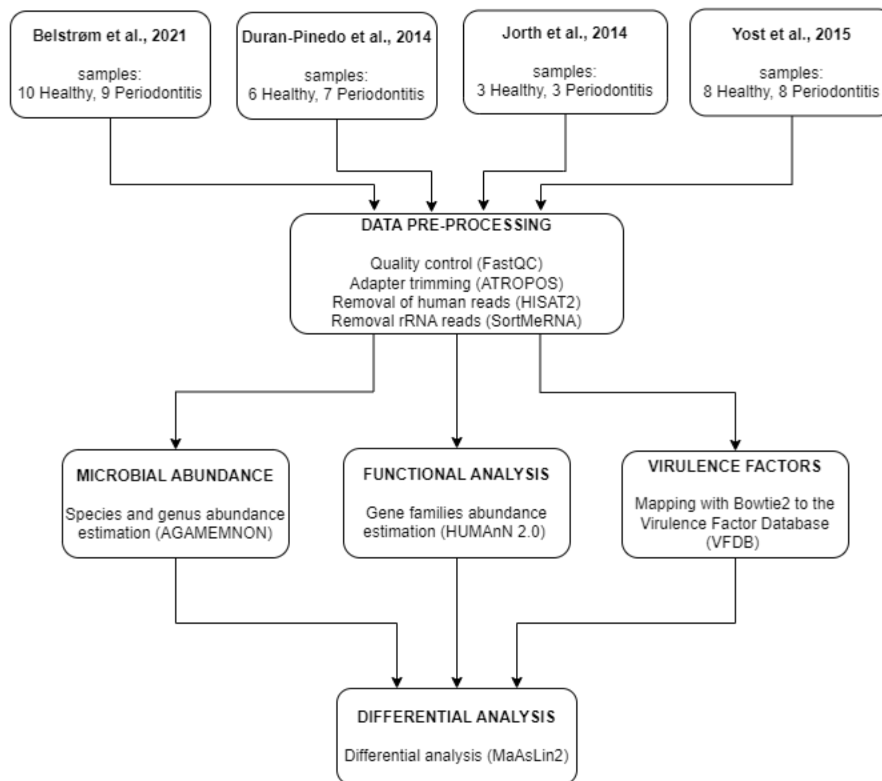


Fig. 4. Illustration of the meta-analysis workflow integrating metatranscriptomics data from four studies.

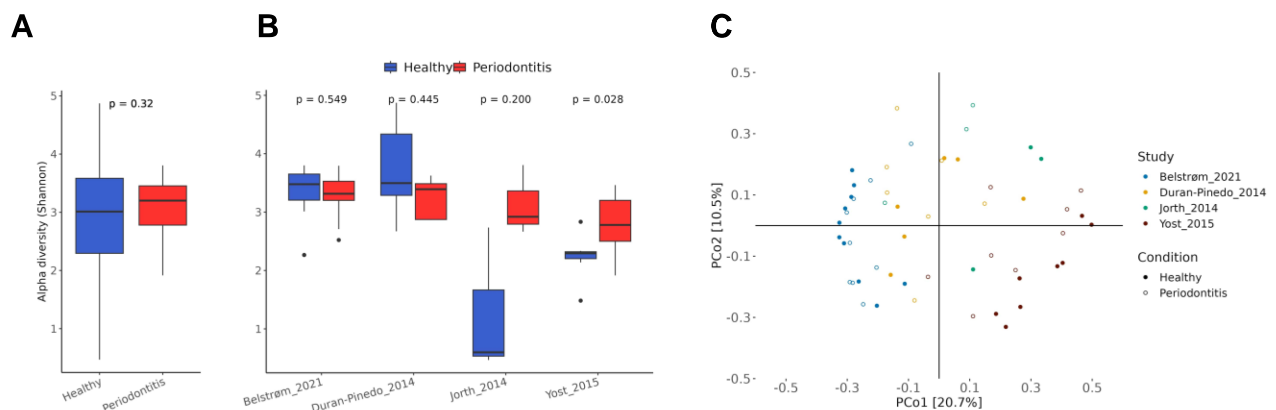


Fig. 5. Microbial species diversity. (**A**) Box plot of α-diversity (Shannon index) in healthy (blue) and periodontitis (red) groups of all studies combined. For statistical analysis, blocked Wilcoxon test was used to treat "study" as a blocking factor. The p-value is shown on the top of the figure. (**B**) Box plot of α-diversity in healthy (blue) and periodontitis (red) groups in each individual study. The p-values of the two-sided Wilcoxon tests are shown at the top of the figures. (**C**) Principal coordinate analysis (PCoA) with Bray-Curtis dissimilarity matrix of the CSS normalized species abundances.

3.3 Differentially abundant species and genera between healthy and periodontitis

The microbial genera that were differentially abundant (FDR ≤ 10^{-3}) between healthy and periodontitis were 25, all of which were bacterial (**Fig. 6A**). The only genera that were more abundant in the healthy group were *Delftia* and *Caulobacter,* while the remaining 23 were increased in periodontitis, including *Filifactor*, *Tannerella*, *Porphyromonas*, *Lactobacillus*, *Bacteroides*, *Treponema*, *Limosilactobacillus*, *Prevotella,* and *Hoylesella,* which exhibited the lowest calculated FDR values, in ascending order. To find out the number of individual studies in which each identified genus is differentially abundant, we performed a similar analysis for each study of the meta-analysis, separately. For the genera with the lowest FDR values, statistical significance was found in at least three of the studies (**Fig. 6A**).

On the species level, 26 species were found to be differentially abundant (FDR ≤ 10^{-3}), all of which were bacteria, and they were more abundant in the periodontitis group (**Fig. 6B**). A more extensive list of the differentially abundant microbial species with an FDR ≤ 0.05 can be found in **Appendix 1**, **Supplementary Table S2**. The top 10 species with the lowest FDR, in ascending order, were the following: *F. alocis, T. denticola, T. forsythia, P. intermedia, T. putidum, P. multiformis, P. gingivalis, P. oris, C. rectus, D. oralis.* All the identified bacterial species had very high prevalence among the samples (>50%) and have been previously associated with oral microbiome or periodontitis (Abusleme et al., 2013; Curtis et al., 2020; Huang et al., 2021; Lasserre et al., 2018; Pérez-Chaparro et al., 2014; Teles et al., 2013). The differentially abundant species were also found to be differentially abundant in most of the individual studies (**Fig. 6B**).

3.4 Core microbiota

The core microbiota for each dental health condition was also identified. This was defined with species that had at least an 80% occurrence within the dental health group and their in-sample relative abundance was among the top 25% of the samples. The number of identified microbial species in the core microbiota of healthy and periodontitis groups were 40 and 80, respectively (**Appendix 1**, **Supplementary Table S3**), meaning that the periodontitis group had a richer core microbiota. The number of shared species between the healthy and periodontitis core microbiota was 38, as only 2 bacterial species were exclusively found in the healthy core, namely, *L. mirabilis* and *N. sicca.* The number of species that were part of the core microbiota of either healthy and/or periodontitis and were found to be enriched in the periodontitis group were 18 (out of the 26 differentially abundant) (**Table 2**). Four species, *T. forsythia, P. gingivalis*, *S. periodonticum* and *P. micra,* were found in both core microbiota, while the remaining 14, only in the periodontitis core.

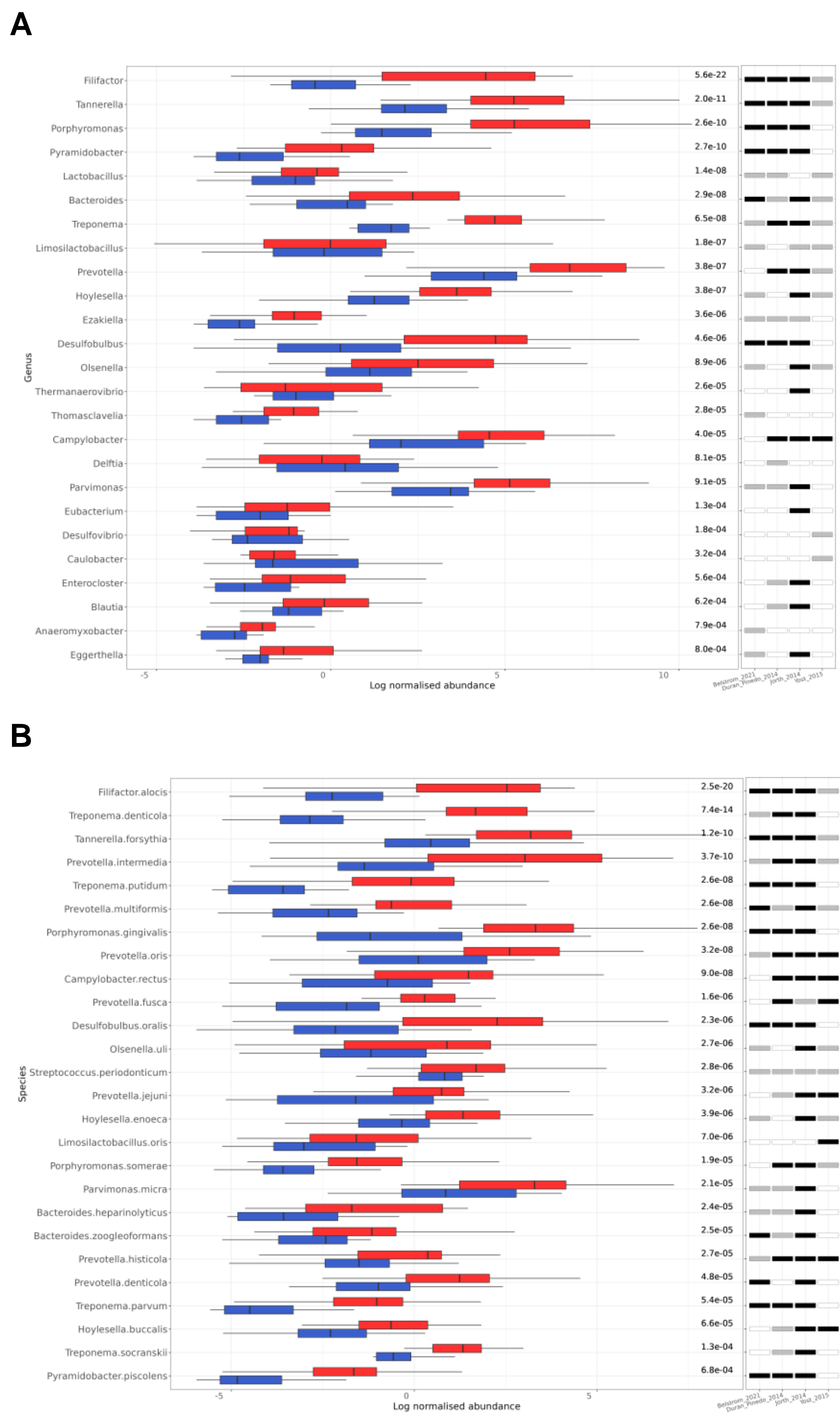


Fig. 6. Differentially abundant genera and species between healthy (blue) and periodontitis (red) groups identified by MaAsLin2. (**A**) The 25 differentially abundant genera (FDR ≤ 10^{-3}). (**B**) The 26 differentially abundant species (FDR ≤ 10^{-3}). The log of CSS normalized abundance is presented. The FDR values are shown on the right side of the bar plot. The statistical significance in each individual study included in the meta-analysis is also shown. A white fill indicates an FDR > 0.05, a grey fill indicates an FDR ≤ 0.05, and a black fill indicates FDR ≤ 10^{-3}.

Table 2. Bacterial species that are part of the core microbiota of healthy and/or periodontitis group, which have been found to be overrepresented in periodontitis. Species in bold were part of the core microbiota of both health status groups, while the rest of the species were only part of the periodontitis core microbiota. The mean relative abundance % of the species in each health group, and the FDR is also shown.

Species	Mean relative abundance %		FDR
	Healthy	Periodontitis	
Filifactor alocis	4.78×10^{-3}	6.93×10^{-1}	2.51×10^{-20}
Treponema denticola	9.28×10^{-3}	5.95×10^{-1}	7.40×10^{-14}
Tannerella forsythia	1.76×10^{-1}	3.90	1.23×10^{-10}
Prevotella intermedia	5.62×10^{-2}	2.22	3.72×10^{-10}
Treponema putidum	7.35×10^{-4}	1.14×10^{-1}	2.55×10^{-8}
Porphyromonas gingivalis	1.49×10^{-1}	3.30	2.62×10^{-8}
Prevotella oris	1.70×10^{-1}	1.02	3.25×10^{-8}
Campylobacter rectus	3.75×10^{-2}	3.27×10^{-1}	8.97×10^{-8}
Desulfobulbus oralis	1.54×10^{-1}	1.78	2.34×10^{-6}
Olsenella uli	3.28×10^{-2}	4.26×10^{-1}	2.71×10^{-6}
Streptococcus periodonticum	5.89×10^{-2}	7.37×10^{-1}	2.84×10^{-6}
Prevotella jejuni	3.96×10^{-2}	4.44×10^{-1}	3.19×10^{-6}
Hoylesella enoeca	4.63×10^{-2}	3.05×10^{-1}	3.89×10^{-6}
Parvimonas micra	2.70×10^{-1}	2.36	2.10×10^{-5}
Prevotella histicola	1.54×10^{-2}	5.05×10^{-2}	2.75×10^{-5}
Prevotella denticola	5.46×10^{-2}	1.23	4.77×10^{-5}
Hoylesella buccalis	7.51×10^{-3}	2.38×10^{-2}	6.56×10^{-5}
Treponema socranskii	5.66×10^{-2}	1.51×10^{-1}	1.30×10^{-4}

3.5 Differentially transcribed gene families (UniRef90) and virulence factors

Functional analysis showed that 52 gene families (UniRef90) were differentially transcribed between the healthy and periodontitis groups, with an FDR $\leq 10^{-2}$ and a total sample prevalence of at least 25 % (**Fig. 7**). The uncharacterized proteins that were found to be differentially transcribed are not presented, but they can be found in **Appendix 1**, **Supplementary Table S4**. Fifty of the differentially transcribed gene families were more abundant in periodontitis and they are involved in transmembrane transport and

secretion, amino acid metabolism, surface protein and flagella synthesis, energy metabolism, and DNA supercoiling. These genes are expressed by bacterial genera that are known to be involved in periodontitis such as *Capnocytophaga, Fusobacterium, Porphyromonas, Fretibacterium, Prevotella, Streptococcus, Synergistes, Tannerella, Parvimonas, Treponema and Desulfobulbus* (Y. Huang et al., 2021). Only two gene families, encoding the LPXTG-motif cell wall anchor domain protein (UniRef90_C0E5X3) from *Corynebacterium* and the PHB domain protein (UniRef90_K8MUM0) from *Streptococcus*, were found to be more abundant in healthy samples (**Fig. 8**). Interestingly, statistical significance (FDR ≤ 0.05) of the differential transcription of the identified gene families was not achieved in each study separately (except from elongation factor Tu – UniRef90_Q73PN3 and 34kDa membrane antigen – UniRef90_C8PTE1 in Jorth et al., 2014), further highlighting the importance of meta-analysis for the identification of gene families with an important role in periodontitis.

For the identification of bacterial virulence factor genes, the virulence factor genes database (VFDB) was used. Four virulence factor genes were found to be differentially transcribed (FDR ≤ 0.05) in periodontitis (**Fig. 9**), and they originate from known periodontal pathogens. In particular, these virulence factors were the TonB-dependent receptor from *P. gingivalis*, BspA from *T. forsythia*, and two from the *Streptococcus* genus: metal ABC transporter substrate-binding lipoprotein/adhesin PsaA, and the Type I glyceraldehyde-3-phosphate dehydrogenase (GAPDH).

4 Discussion

In our meta-analysis, we observed a higher number of microbial species in the core microbiota of the periodontitis group compared to the healthy group, with the former exhibiting twice the number of species (80 versus 40, respectively). Consistent with findings from Abusleme and co-authors, an increased number of core species in the periodontitis group can be attributed to the emergence of new species during the development of periodontitis, without the replacement of primary health-associated species (Abusleme et al., 2013). In the healthy state, the Gram-positive bacteria *Actinomyces*, *Streptococcus*, *Rothia*, and *Gemella,* and to a lesser extent, the Gram-negative bacteria *Veillonella, Neisseria, Leptotrichia*, *Fusobacterium,* and *Lautropia* dominated the dental microbiota. However, under disease conditions, there was a notable shift towards Gram-negative bacteria with a pathogenic role such as, *Tannerella*, *Prevotella*, *Porphyromonas*, *Treponema*, *Campylobacter*, *Pyramidobacter, Bacteroides* and *Desulfobulbus*, which aligns closely with findings from (Ezzo & Cutler, 2003). Other bacterial genera with eminent presence in periodontal disease included *Capnocytophaga*, *Fusobacterium*, *Filifactor*, *Leptotrichia* and *Hoylesella.*

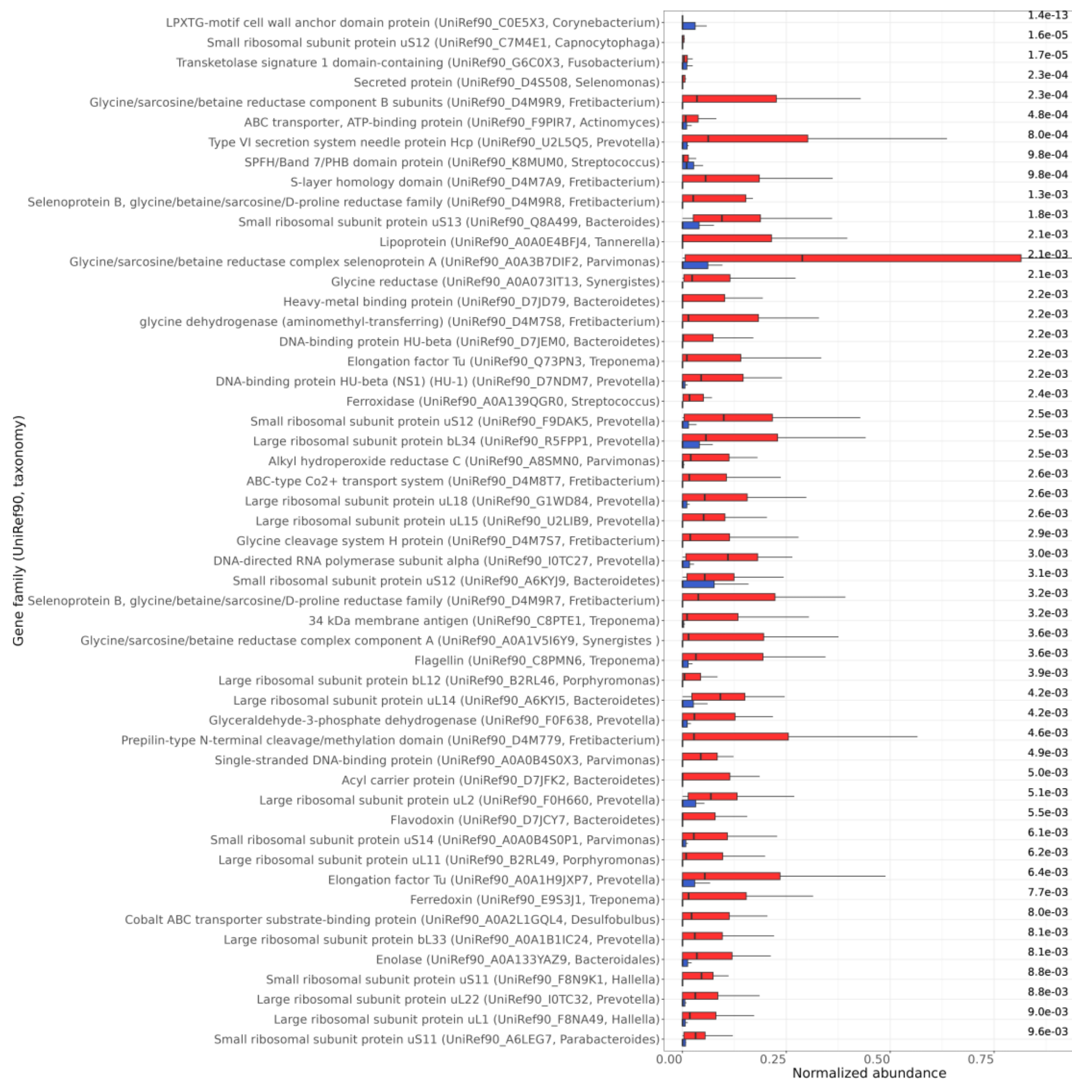


Fig. 7. Differentially transcribed gene families (UniRef90, taxonomy) between healthy (blue) and periodontitis (red) health groups identified by MaAsLin2. The normalized abundance of the 52 gene families, with FDR ≤ 10^{-2}, are presented. The FDR are shown on the right side of the bar plot.

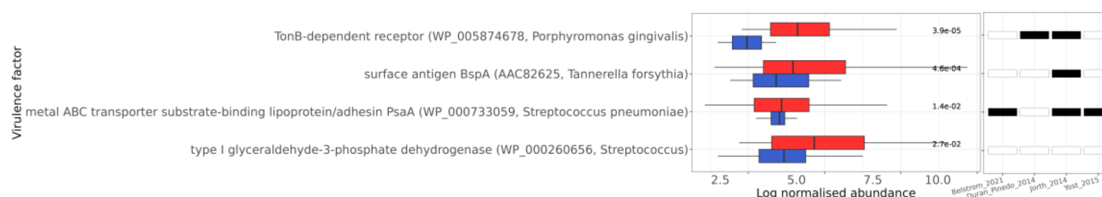


Fig. 8. Differentially transcribed virulence factors between healthy (blue) and periodontitis (red) health groups. The log of normalized abundance of the 4 virulence factor genes with FDR ≤ 0.05 are presented. The name of the genes and the NCBI accession (in parenthesis) are shown. The FDR are shown on the right side of the bar plot. The statistical significance in each individual study included in the meta-analysis is also shown. White fill indicates FDR > 0.05, while black fill indicates FDR ≤ 0.05.

As expected, prominent periodontal pathogens, including *F. alocis*, *T. forsythia*, *P. gingivalis*, *T. denticola*, *P. intermedia, C. rectus, D. oralis and S. periodonticum* were significantly more abundant in periodontitis. However, none of these species was differentially abundant in all individual studies with an FDR threshold of 10^{-3}, while statistical significance (FDR ≤ 0.05) was achieved only for seven of these species across all studies. This highlights the importance of meta-analysis for the identification of microbial species with important role in periodontitis.

The functional analysis identified 50 gene families (UniRef90) with significantly increased transcription levels in the periodontitis group. Most of these gene families originate from bacterial genera with known periodontal pathogen members, and they are involved in various processes of bacterial metabolism. In particular, ABC transporters from *Actinomyces* and *Fretibacterium*, and metal ABC transporters from *Desulfobulbus* (cobalt) and *Streptococcus* (manganese) displayed higher transcription levels in periodontitis. Increased transcription levels of ABC transporters in periodontitis have been previously reported (Duran-Pinedo et al., 2014; Szafrański et al., 2015; Yost et al., 2015). Besides their role as transporters, these proteins may also function as adhesins, for the attachment to the host cells, as shown for the virulence factor PsaA, which is part of the ABC-type Mn^{+2} transport protein complex of *S. pneumoniae* (Rajam et al., 2008). A number of surface proteins, with known virulence functions, were also found to be enriched in periodontitis. Increased levels of TonB-dependent receptors, which are involved in hemin and hemoglobin utilization (Dashper et al., 2000; Simpson et al., 2000), were identified for *P. gingivalis*, congruent with Yost et al., 2015 and Szafrański et al., 2015. Higher transcription of prepilin from *Fretibacterium* was detected in periodontitis. Pili (or fimbriae) are filamentous structures on the bacterial cell surface, with crucial role in virulence and host cell attachment (Ellison et al., 2022; Yoshimura et al., 2009). The pathogenic role of type IV pilins has been reported for the periodontitis associated bacteria *P. gingivalis* (Choi et al., 2012) and *Eikenella corrodens* (Hood & Hirschberg, 1995), and their enhanced expression under periodontal disease conditions has been previously reported (Szafrański et al., 2015). Transcription of flagellin from *Treponema* was also enhanced in periodontal disease. The flagella confer the characteristic spiral morphology of spirochetes and play an important role in their motility and host tissue infection (Kurniyati et al., 2023), and their up-regulation in disease has been reported (Duran-Pinedo et al., 2014; Szafrański et al., 2015; Yost et al., 2015). Other up-regulated surface proteins with a potential pathogenic role, identified in the periodontitis state, were the BspA of *T. forsynthia* (Mahalakshmi et al., 2018; Sharma et al., 1998), the s-layer homology protein of *Fretibacterium*, and a lipoprotein cluster of *Tannerella. Prevotella* is known to secrete proteins in both planktonic and biofilm lifestyles (Karched et al., 2022) and the increased transcription of the type VI secretion system protein Hcp suggests an important role of the *Prevotella* secretome in periodontitis. Elongation factor thermal unstable (EF-Tu) is one of the most abundant proteins in bacteria, with an important role in protein synthesis. However, these proteins could also display ‘moonlighting’ functions on the surface of the bacterial cells, where they can operate as adhesins and stimulate

immune response (Harvey et al., 2019). EF-Tu proteins from *Treponema* and *Prevotella* were up-regulated, indicating that these proteins may contribute to the pathogenicity of *Treponema* and *Prevotella* in periodontitis.

Periodontal pockets are known to contain high levels of peptides and amino acids which emanate from the degradation of host serum or damaged tissue, by the proteolytic activity of the microbial communities and the immune response (Teles et al., 2013). Several differentially transcribed genes under the periodontal disease state are involved in glycine metabolism. More specifically, glycine reductases from bacterial genera *Fretibacterium*, *Synergistes*, and *Parvimonas,* and glycine cleavage system H- and T-proteins from the *Fretibacterium* genus*,* exhibited increased transcription. Glycine reductases are involved in metabolic pathways that lead to energy production from glycine catabolism. Glycine reductases consist of 3 protein fractions, A and B selenoproteins and fraction C, and they partake in ATP production through the conversion of glycine to acetyl phosphate by using inorganic phosphate and the reducing potential of thioredoxin (Andreesen, 2004). In line with our results, the growth of the periodontal pathogen *T. denticola* has been reported to be impaired by the inhibition of selenium metabolism (Jackson-Rosario & Self, 2009). The glycine cleavage system comprises 4 proteins called T-, P-, L- and H-protein, and catalyzes the formation of serine from glycine (Szafrański et al., 2015). Serine is converted to pyruvate by the serine dehydratase, which can then be further metabolized by the ferredoxin oxidoreductase, or by the flavodoxin oxidoreductase under iron-limiting conditions (Buckel & Thauer, 2018). Ferredoxin from *Treponema* and flavodoxin from *Bacteroidetes,* were also enriched in periodontitis, in agreement with Szafrański et al., 2015. Interestingly, glycine has been shown to have an anti-inflammatory effect in gingival epithelial cells (Schaumann et al., 2013). Overall, the results indicate the important role of glycine catabolism and energy metabolism rewiring, for subgingival survival and for periodontal disease development.

Enhanced transcription of the DNA-binding protein HU from *Prevotella* and *Bacteroidetes* was identified under the periodontal disease state. HU has been reported to affect the transcription of surface polysaccharides in *P. gingivalis* (Alberti-Segui et al., 2010), and therefore, our results connote that this DNA-binding protein might function as a global regulator of genes involved in periodontitis.

Metatranscriptomics excludes the genomic material of dead and inactive cells and could generate more relevant information on the role of microbes in human health (Hempel et al., 2022; Ojala et al., 2023). By utilizing four distinct metatranscriptome datasets, our meta-analysis has produced robust and reliable results, pinpointing signature microbial species, gene families, virulence factor genes, and microbial relationships crucial in periodontal disease. The bacterial species and gene families identified as pivotal in the context of periodontitis, may serve as focal points for future strategies aimed at disease mitigation and drug development. To further advance our understanding of periodontal disease and its progression, longitudinal studies in conjunction with multiomics studies combining transcriptomics, proteomics, and metabolomics, could be used to associate genes, functional proteins and metabolites involved in periodontitis (Huang et al., 2021). Additionally, host-pathogen interaction

studies could elucidate the interplay between the microbial community and the host and its immune response (Cugini et al., 2021).

5 Code availability

Bioinformatic open-source tools and parameters used in this present study are defined in the “Methods” section and scripts can be found in the following Github repository: https://github.com/dianalabgr/Oral_Metatranscriptome_MetaAnalysis.

Appendix 1

Supplementary material

Table S1. Studies that were considered for our meta-analysis. The study details, the decision on inclusion outcome (YES/NO) and the justification of exclusion are shown.

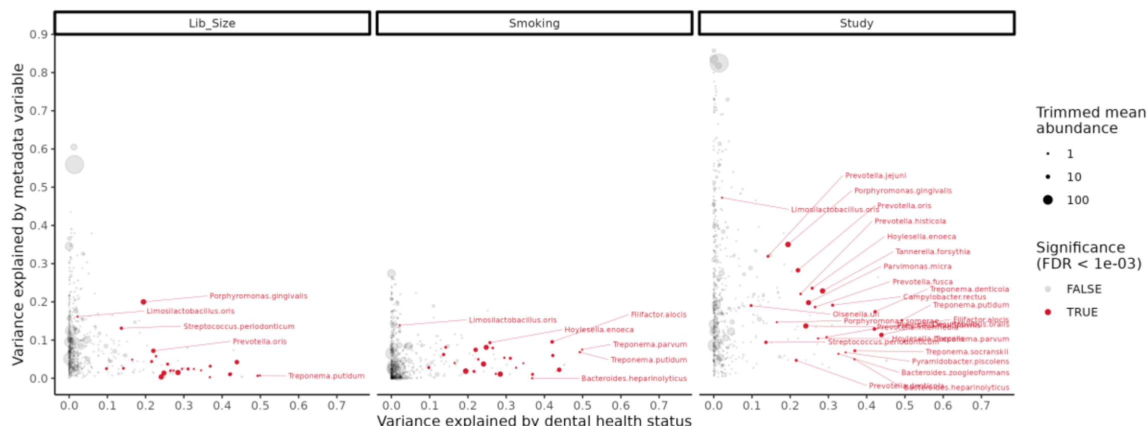
A/A	Authors	Title	Year	Journal	DOI	Inclusion (YES/NO)	Reason of Exclusion
1	Zhong et al.	Metagenome and Metatranscriptome Analyses Using Protein Family Profiles.	2016	PLoS Comput. Biol.	10.1371/journal.pcbi.1004991	NO	A homology detection algorithm is presented. New sequencing data was not generated.
2	Duran-Pinedo et al.	Subgingival host-microbiome metatranscriptomic changes following scaling and root planing in grade II/III periodontitis.	2023	J. Clin. Periodontol.	10.1111/jcpe.13737	NO	Stable and progressive dental sites are studied. No clearly healthy and periodontitis samples were produced.
3	Yost et al.	Potassium is a key signal in host-microbiome dysbiosis in periodontitis.	2017	PLoS Pathog.	10.1371/journal.ppat.1006457	NO	Same data used as in Yost et al., 2015.
4	Frias-Lopez & Duran-Pinedo	Effect of periodontal pathogens on the metatranscriptome of a healthy multispecies biofilm model.	2012	J. Bacteriol.	10.1128/JB.06328-11	NO	*In vitro* experimental data.
5	May et al.	metaModules identifies key functional subnetworks in microbiome-related disease.	2016	Bioinformatics	10.1093/bioinformatics/btv526	NO	MetaModules tool was introduced. No metatranscriptome data produced.
6	Belstrøm et al.	Transcriptional Activity of Predominant *Streptococcus* Species at Multiple Oral Sites Associate With Periodontal Status.	2021	Front. Cell Infect. Microbiol.	10.3389/fcimb.2021.752664	NO	No original metatranscriptome data produced.
7	Lewin et al.	A quantitative framework reveals traditional laboratory growth is a highly accurate model of human oral infection.	2022	Proc Natl Acad Sci U S A.	10.1073/pnas.2116637119	NO	No original metatranscriptome data produced.
8	Grischke et al.	Removable denture is a risk indicator for peri-implantitis and facilitates expansion of specific periodontopathogens: a cross-sectional study.	2021	BMC Oral Health	10.1186/s12903-021-01529-9	NO	Peri-implantitis was studied.

9	Aleti et al.	Identification of the Bacterial Biosynthetic Gene Clusters of the Oral Microbiome Illuminates the Unexplored Social Language of Bacteria during Health and Disease.	2019	mBio	10.1128/mBio.00321-19	NO	No metatranscriptome data produced.
10	Yost & Duran-Pinedo	The contribution of Tannerella forsythia dipeptidyl aminopeptidase IV in the breakdown of collagen.	2018	Mol. Oral. Microbiol.	10.1111/omi.12244	NO	No metatranscriptome data was produced.
11	Jorth et al.	Metatranscriptomics of the Human Oral Microbiome during Health and Disease.	2014	mBio	10.1128/mBio.01012-14	YES	
12	Belstrøm et al.	Periodontitis associates with species-specific gene expression of the oral microbiota.	2021	NPJ Biofilms and Microbiomes	10.1038/s41522-021-00247-y	YES	
13	Nowicki et al.	Microbiota and Metatranscriptome Changes Accompanying the Onset of Gingivitis.	2018	mBio	10.1128/mBio.00575-18	NO	The samples are from dental sites with gingivitis and not periodontitis.
14	Edlund et al.	Uncovering complex microbiome activities via metatranscriptomics during 24 hours of oral biofilm assembly and maturation.	2018	Microbiome	10.1186/s40168-018-0591-4	NO	*In vitro* oral biofilm used. Only first 24 h of biofilm development was studied.
15	Komatsu et al.	Discriminating Microbial Community Structure Between Peri-Implantitis and Periodontitis With Integrated Metagenomic, Metatranscriptomic, and Network Analysis.	2020	Frontiers in Cellular and Infection Microbiology	10.3389/fcimb.2020.596490	NO	Lack of healthy samples. Only periodontitis and peri-implantitis samples.
16	Duran-Pinedo et al.	Small RNA Transcriptome of the Oral Microbiome during Periodontitis Progression.	2015	AEM journal	10.1128/AEM.01782-15	NO	Same samples as in Duran-Pinedo et al., 2014
17	Shiba et al.	Distinct interacting core taxa in co-occurrence networks enable discrimination of polymicrobial oral diseases with similar symptoms.	2016	Scientific Reports	10.1038/srep30997	NO	Lack of healthy samples. Only periodontitis and peri-implantitis samples.
18	Carda-Diéguez et al.	Functional changes in the oral microbiome after use of fluoride and arginine containing dentifrices: a metagenomic and metatranscriptomic study.	2022	Microbiome	10.1186/s40168-022-01338-4	NO	Study of fluoride-containing dentifrice on dental caries. Supragingival dental plaque samples
19	Duran-Pinedo et al.	Community-wide transcriptome of the oral microbiome in subjects with and without periodontitis.	2014	The ISME Journal	10.1038/ismej.2014.23	YES	
20	Zhang et al.	Metatranscriptomic analysis of an in vitro biofilm model reveals strain-specific	2019	J. Oral Microbiol.	10.1080/20002297.2019.1599670	No	An *in vitro* polymicrobial subgingival plaque model was used.

		interactions among multiple bacterial species					
21	Yost et al.	Functional signatures of oral dysbiosis during periodontitis progression revealed by microbial metatranscriptome analysis.	2015	Genome Medicine	10.1186/s13073-015-0153-3	YES	
22	Nemoto et al.	Discrimination of Bacterial Community Structures amongHealthy, Gingivitis, and Periodontitis Statuses through Integrated Metatranscriptomic and Network Analyses.	2021	mSystems	10.1128/msystems.00886-21	NO	Raw data not available.
23	Szafrański et al.	Functional biomarkers for chronic periodontitis and insights into the roles of Prevotella nigrescens and Fusobacterium nucleatum; a metatranscriptome analysis.	2015	NPJ Biofilms and Microbiomes	10.1038/npjbiofilms.2015.17	NO	Raw data not available.
24	Ram-Mohan & Meyer	Comparative Metatranscriptomics of Periodontitis Supports a Common Polymicrobial Shift in Metabolic Function and Identifies Novel Putative Disease-Associated ncRNAs	2020	Front. Microbiol.	10.3389/fmicb.2020.00482	No	No new sequencing data was produced.
25	Belstrøm et al.	Metagenomic and metatranscriptomic analysis of saliva reveals disease-associated microbiota in patients with periodontitis and dental caries	2017	NPJ Biofilms and Microbiomes	10.1038/s41522-017-0031-4	No	Saliva samples were used.
26	Beall et al.	Genomics of the Uncultivated, Periodontitis-Associated Bacterium *Tannerella* sp. BU045 (Oral Taxon 808)	2018	mSystems	10.1128/mSystems.00018-18	No	No metatranscriptome data was produced.
27	Joseph et al.	The Murine Oral Metatranscriptome Reveals Microbial and Host Signatures of Periodontal Disease	2023	J. Dent. Res.	10.1177/00220345221149675	No	No human samples.
28	Lewin et al.	A quantitative framework reveals traditional laboratory growth is a highly accurate model of human oral infection	2022	Proc. Natl Acad. Sci. USA		No	No new sequencing data was produced.
29	Shah et al.	The making of a miscreant: tobacco smoke and the creation of pathogen-rich biofilms	2017	NPJ Biofilms and Microbiomes	10.1038/s41522-017-0033-2	No	The article does not study periodontitis, but the effect of smoking

30	Edlund et al.	High-Quality Draft Genome Sequence of Low-pH-Active Veillonella parvula Strain SHI-1, Isolated from Human Saliva within an In Vitro Oral Biofilm Model	2016	Genome Announc.	10.1128/genomeA.01684-15	No	No metatranscriptome and subgingival samples.
31	Deng et al.	Worlds Apart – Transcriptome Profiles of Key Oral Microbes in the Periodontal Pocket Compared to Single Laboratory Culture Reflect Synergistic Interactions	2018	Front. Microbiol.	10.3389/fmicb.2018.00124	No	No metatranscriptome data was produced
32	Zou et al.	TrkA serves as a virulence modulator in *Porphyromonas gingivalis* by maintaining heme acquisition and pathogenesis	2022	Front. Cell. Infect. Microbiol.	10.3389/fcimb.2022.1012316	No	No metatranscriptome data was produced.
33	Fried et al.	Use of unbiased metagenomic and transcriptomic analyses to investigate the association between feline calicivirus and feline chronic gingivostomatitis in domestic cats	2021	Am. J. Vet. Res.	10.2460/ajvr.82.5.381	No	Human periodontitis was not studied.
34	Wang G.P.	Defining functional signatures of dysbiosis in periodontitis progression	2015	Genome Med.	10.1186/s13073-015-0165-z	No	No sequencing data produced
35	Duran-Pinedo et al.	The effect of the stress hormone cortisol on the metatranscriptome of the oral microbiome	2018	NPJ Biofilms and Microbiomes	10.1038/s41522-018-0068-z	No	Periodontitis was not studied.
36	Stashenko et al.	The Oral Mouse Microbiome Promotes Tumorigenesis in Oral Squamous Cell Carcinoma	2019	mSystems	10.1128/mSystems.00323-19	No	No human metatranscriptome data was produced.
37	Szafrański et al.	Quorum sensing of Streptococcus mutans is activated by Aggregatibacter actinomycetemcomitans and by the periodontal microbiome	2017	BMC Genomics	10.1186/s12864-017-3618-5	No	No periodontal metatranscriptome data was produced. Biofilms of of *S. mutans*, *A. actinomycetemcomitans were studied*
38	Wattimena et al.	An in vitro model for studies of attenuation of antibiotic-inhibited growth of *Aggregatibacter actinomycetemcomitans* Y4 by polyamines	2021	Mol. Oral Microbiol.	10.1111/omi.12353	No	Periodontitis was not studied and subgingival metatranscriptome data was not produced

A



B

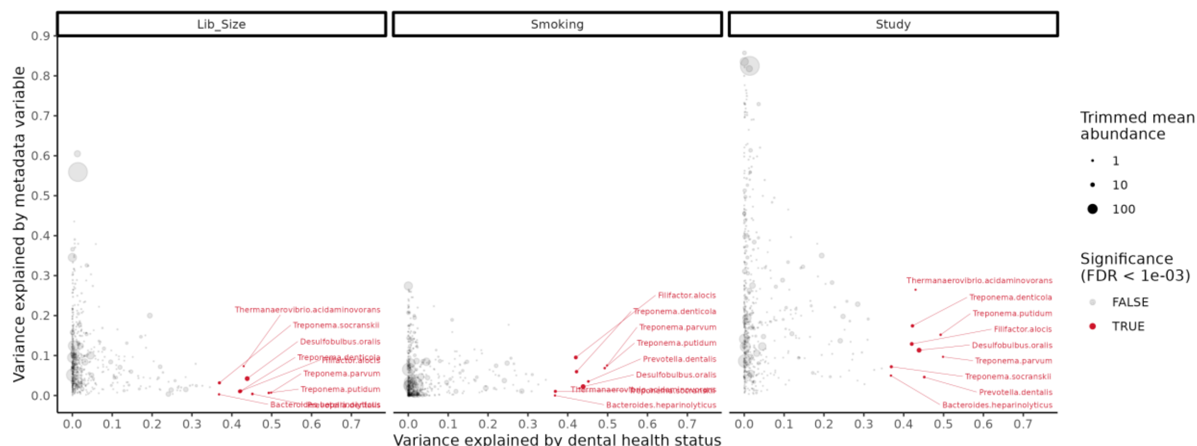


Fig. 9. The variance explained by the confounding factors (normalized library sizes, smoking status and study origin) against the variance explained by the oral health status, with (**A**) and without (**B**) setting the study factor as a random effect. The analysis of the confounding factors for species abundance was conducted following the methodology described by (Wirbel et al., 2019). Briefly, the total variance within the abundance of a given microbial species was compared to the variance explained by disease status and the variance explained by the confounding factor, with a linear model assumption. Variance calculations were performed with a non-parametric approach to account for non-Gaussian distribution of microbiome abundance data. The library size potential confounder, with continuous values, was transformed into categorical data as quartiles. The size of the spots represents the trimmed mean abundance. The species that have been found to have statistically significant abundance (qval ≤ 10^{-3}), after differential analysis with MaAsLin2, are shown in red.

Table S2. The 61 differentially abundant species between healthy and periodontitis groups identified by MaAsLin2 (FDR > 0.05).

Species	coef	stderr	N	N.not.0	pval	qval
Filifactor.alocis	4.347	0.439	54	47	4.60E-23	2.51E-20
Treponema.denticola	4.130	0.505	54	44	2.72E-16	7.40E-14
Tannerella.forsythia	2.826	0.393	54	54	6.80E-13	1.23E-10
Prevotella.intermedia	3.344	0.478	54	51	2.73E-12	3.72E-10
Prevotella.multiformis	2.878	0.456	54	48	2.72E-10	2.55E-08
Treponema.putidum	4.127	0.654	54	38	2.81E-10	2.55E-08
Porphyromonas.gingivalis	3.593	0.572	54	53	3.37E-10	2.62E-08
Prevotella.oris	2.225	0.357	54	54	4.77E-10	3.25E-08
Campylobacter.rectus	2.826	0.467	54	48	1.48E-09	8.97E-08
Prevotella.fusca	2.833	0.510	54	49	2.85E-08	1.56E-06
Desulfobulbus.oralis	3.593	0.658	54	48	4.73E-08	2.34E-06
Olsenella.uli	2.772	0.511	54	49	5.98E-08	2.71E-06
Streptococcus.periodonticum	2.074	0.384	54	51	6.76E-08	2.84E-06
Prevotella.jejuni	2.182	0.407	54	52	8.19E-08	3.19E-06
Hoylesella.enoeca	2.096	0.394	54	50	1.07E-07	3.89E-06
Limosilactobacillus.oris	4.939	0.951	54	28	2.06E-07	7.01E-06
Porphyromonas.somerae	3.027	0.606	54	42	5.79E-07	1.85E-05
Parvimonas.micra	2.023	0.408	54	52	6.94E-07	2.10E-05
Bacteroides.heparinolyticus	2.620	0.532	54	41	8.47E-07	2.43E-05
Bacteroides.zoogleoformans	3.245	0.661	54	42	9.30E-07	2.53E-05
Prevotella.histicola	1.842	0.377	54	49	1.06E-06	2.75E-05
Prevotella.denticola	2.452	0.515	54	53	1.93E-06	4.77E-05
Treponema.parvum	3.819	0.808	54	37	2.26E-06	5.36E-05
Hoylesella.buccalis	1.752	0.374	54	46	2.89E-06	6.56E-05
Treponema.socranskii	1.571	0.347	54	49	5.95E-06	1.30E-04
Pyramidobacter.piscolens	2.962	0.713	54	39	3.25E-05	6.81E-04
Fusobacterium.mortiferum	1.892	0.481	54	36	8.44E-05	1.70E-03
Porphyromonas.crevioricanis	2.701	0.696	54	33	1.03E-04	2.01E-03
Cutibacterium.granulosum	-1.764	0.457	54	52	1.11E-04	2.10E-03
Streptococcus.cristatus	-1.274	0.332	54	54	1.25E-04	2.28E-03
Streptococcus.anginosus	1.422	0.372	54	54	1.32E-04	2.33E-03
Ralstonia.insidiosa	-2.404	0.631	54	39	1.40E-04	2.39E-03
Delftia.lacustris	-2.163	0.580	54	37	1.93E-04	3.18E-03
Eikenella.exigua	1.662	0.448	54	46	2.08E-04	3.33E-03

Fusobacterium.gastrosuis	1.723	0.478	54	42	3.12E-04	4.83E-03
Parascardovia.denticolens	2.962	0.823	54	36	3.19E-04	4.83E-03
Ralstonia.wenshanensis	-2.599	0.724	54	28	3.31E-04	4.88E-03
Fusobacterium.necrophorum	2.397	0.672	54	41	3.59E-04	5.15E-03
Prevotella.dentalis	1.515	0.441	54	47	6.00E-04	8.38E-03
Brevundimonas.albigilva	-1.693	0.497	54	34	6.61E-04	9.01E-03
Aminipila.luticellarii	1.982	0.591	54	36	7.97E-04	1.06E-02
Pseudoprevotella.muciniphila	3.984	1.202	54	30	9.20E-04	1.19E-02
Campylobacter.gracilis	1.248	0.379	54	53	9.80E-04	1.24E-02
Streptococcus.gordonii	-1.153	0.358	54	54	1.29E-03	1.60E-02
Ralstonia.mannitolilytica	-2.239	0.698	54	28	1.34E-03	1.62E-02
Thermanaerovibrio.acidaminovorans	2.099	0.677	54	30	1.93E-03	2.29E-02
Prevotella.veroralis	2.123	0.692	54	52	2.17E-03	2.52E-02
Scardovia.inopinata	2.819	0.924	54	36	2.28E-03	2.59E-02
Fusobacterium.canifelinum	1.111	0.366	54	52	2.43E-03	2.59E-02
Roseburia.hominis	2.938	0.966	54	32	2.34E-03	2.59E-02
Treponema.phagedenis	2.602	0.858	54	35	2.42E-03	2.59E-02
Porphyromonas.asaccharolytica	1.980	0.655	54	39	2.49E-03	2.61E-02
Comamonas.thiooxydans	-1.971	0.657	54	37	2.71E-03	2.78E-02
Dialister.massiliensis	1.985	0.665	54	32	2.85E-03	2.87E-02
Schlegelella.aquatica	-1.900	0.638	54	33	2.90E-03	2.87E-02
Brevundimonas.pondensis	-2.739	0.936	54	32	3.44E-03	3.34E-02
Streptococcus.sobrinus	-1.801	0.620	54	47	3.67E-03	3.51E-02
X.Ruminococcus.lactaris	2.882	1.002	54	30	4.02E-03	3.78E-02
Corynebacterium.kroppenstedtii	1.626	0.575	54	40	4.70E-03	4.27E-02
Leptotrichia.trevisanii	1.385	0.489	54	51	4.63E-03	4.27E-02
Actinomyces.naeslundii	-1.021	0.362	54	54	4.78E-03	4.27E-02

Table S3. The 82 bacterial species that are part of the core microbiota of healthy and/or periodontitis group.

	Mean relative abundance	
Species	**Healthy**	**Periodontitis**
Actinomyces.faecalis	1.19E-02	1.23E-02
Actinomyces.gaoshouyii	6.42E-04	1.24E-03
Actinomyces.howellii	8.43E-03	6.13E-03
Actinomyces.marmotae	1.30E-03	1.20E-03
Actinomyces.naeslundii	3.90E-02	1.37E-02
Actinomyces.oris	6.13E-02	8.30E-02
Actinomyces.pacaensis	1.48E-04	4.22E-04
Actinomyces.procaprae	6.27E-04	4.27E-04
Actinomyces.qiguomingii	4.13E-04	8.11E-04
Actinomyces.radicidentis	4.31E-03	4.29E-03
Actinomyces.respiraculi	1.45E-03	1.64E-03
Actinomyces.slackii	7.48E-04	1.50E-03
Actinomyces.trachealis	7.26E-04	6.33E-04
Actinomyces.viscosus	5.31E-03	3.64E-03
Actinomyces.weissii	2.03E-03	2.36E-03
Arachnia.propionica	5.62E-03	3.72E-03
Campylobacter.gracilis	1.50E-03	2.97E-03
Campylobacter.rectus	3.75E-04	3.27E-03
Campylobacter.showae	1.34E-03	2.41E-03
Capnocytophaga.gingivalis	8.90E-03	1.74E-02
Capnocytophaga.sputigena	3.00E-03	2.96E-03
Desulfobulbus.oralis	1.54E-03	1.78E-02
Eikenella.corrodens	1.13E-03	1.13E-03
Filifactor.alocis	4.78E-05	6.93E-03
Fusobacterium.canifelinum	2.70E-03	7.94E-03
Fusobacterium.hwasookii	3.06E-03	4.55E-03
Fusobacterium.nucleatum	1.40E-02	1.53E-02
Fusobacterium.pseudoperiodonticum	9.11E-04	1.83E-03
Gemella.haemolysans	2.75E-03	4.20E-03
Gemella.morbillorum	7.55E-03	6.19E-03
Gemella.sanguinis	1.28E-03	2.07E-03
Granulicatella.adiacens	8.21E-03	4.60E-03
Hoylesella.buccalis	7.51E-05	2.38E-04
Hoylesella.enoeca	4.63E-04	3.05E-03
Kingella.oralis	4.09E-03	3.97E-03
Lachnoanaerobaculum.umeaense	5.03E-03	4.88E-03
Lancefieldella.parvula	1.11E-03	3.44E-03

Lautropia.mirabilis	1.46E-02	2.82E-03
Leptotrichia.trevisanii	1.63E-03	3.51E-03
Leptotrichia.wadei	8.58E-03	1.09E-02
Mogibacterium.diversum	8.30E-04	2.26E-03
Neisseria.sicca	2.58E-03	1.28E-03
Olsenella.uli	3.28E-04	4.26E-03
Parvimonas.micra	2.70E-03	2.36E-02
Porphyromonas.gingivalis	1.49E-03	3.30E-02
Prevotella.dentalis	1.86E-04	8.96E-04
Prevotella.denticola	5.46E-04	1.23E-02
Prevotella.histicola	1.54E-04	5.05E-04
Prevotella.intermedia	5.62E-04	2.22E-02
Prevotella.jejuni	3.96E-04	4.44E-03
Prevotella.melaninogenica	2.18E-03	3.68E-03
Prevotella.nigrescens	1.01E-02	1.05E-02
Prevotella.oris	1.70E-03	1.02E-02
Prevotella.veroralis	2.46E-03	4.53E-03
Rothia.aeria	6.16E-03	7.23E-03
Rothia.dentocariosa	5.85E-02	5.44E-02
Schaalia.cardiffensis	1.83E-03	4.34E-03
Schaalia.meyeri	3.56E-03	3.08E-03
Selenomonas.sputigena	8.28E-03	1.01E-02
Streptococcus.anginosus	9.63E-04	6.75E-03
Streptococcus.australis	2.21E-03	4.97E-03
Streptococcus.cristatus	2.87E-02	1.01E-02
Streptococcus.gordonii	3.38E-02	1.94E-02
Streptococcus.gwangjuense	9.32E-03	1.17E-02
Streptococcus.himalayensis	3.51E-04	2.56E-04
Streptococcus.intermedius	4.33E-03	1.15E-02
Streptococcus.koreensis	2.69E-03	3.75E-03
Streptococcus.lactarius	3.11E-03	8.98E-03
Streptococcus.mutans	1.34E-02	1.68E-02
Streptococcus.oralis	5.69E-02	4.37E-02
Streptococcus.periodonticum	5.89E-04	7.37E-03
Streptococcus.pseudopneumoniae	5.37E-03	5.88E-03
Streptococcus.thermophilus	2.31E-03	2.89E-03
Streptococcus.toyakuensis	9.78E-03	1.40E-02
Tannerella.forsythia	1.76E-03	3.90E-02
Treponema.denticola	9.28E-05	5.95E-03
Treponema.putidum	7.35E-06	1.14E-03
Treponema.socranskii	5.66E-04	1.51E-03

Veillonella.dispar	1.89E-03	2.46E-03
Veillonella.nakazawae	2.29E-03	2.65E-03
Veillonella.parvula	3.20E-02	3.91E-02
Veillonella.rodentium	6.60E-04	6.06E-04

Table S4. The 81 differentially transcribed gene families (UniRef90, taxonomy) between healthy and periodontitis health groups identified by MaAsLin2 (FDR ≤ 10^{-2}).

Gene_family	coef	stderr	N	N.not.0	pval	qval
UniRef90_C0E5X3	-0.308	0.036	54	16	9.41E-18	1.41E-13
UniRef90_C7M4E1	1.028	0.172	54	15	2.10E-09	1.58E-05
UniRef90_G6C0X3	1.283	0.217	54	22	3.34E-09	1.67E-05
UniRef90_UPI0009428330	4.840	0.843	54	20	9.44E-09	3.54E-05
UniRef90_T2NCD3	4.722	0.864	54	23	4.61E-08	1.39E-04
UniRef90_D4M9R9	3.943	0.742	54	21	1.09E-07	2.33E-04
UniRef90_D4S508	1.287	0.242	54	14	1.02E-07	2.33E-04
UniRef90_F9PIR7	1.046	0.203	54	25	2.55E-07	4.79E-04
UniRef90_U2L5Q5	3.097	0.615	54	28	4.78E-07	7.98E-04
UniRef90_D4M7A9	3.420	0.691	54	20	7.47E-07	9.81E-04
UniRef90_E8JUY6	-0.529	0.107	54	29	7.46E-07	9.81E-04
UniRef90_M2BDX0	3.701	0.749	54	22	7.84E-07	9.81E-04
UniRef90_D4M9R8	3.758	0.775	54	19	1.25E-06	1.34E-03
UniRef90_R5M7M5	2.970	0.612	54	20	1.20E-06	1.34E-03
UniRef90_Q8A499	2.119	0.443	54	33	1.78E-06	1.78E-03
UniRef90_A0A073IT13	3.602	0.767	54	24	2.68E-06	2.12E-03
UniRef90_A0A0E4BFJ4	3.656	0.778	54	19	2.64E-06	2.12E-03
UniRef90_A0A3B7DIF2	2.340	0.498	54	34	2.66E-06	2.12E-03
UniRef90_Q73NG0	3.844	0.815	54	17	2.38E-06	2.12E-03
UniRef90_A0A0E2IUZ9	3.242	0.696	54	40	3.16E-06	2.18E-03
UniRef90_A0A3A0QCD4	3.020	0.650	54	23	3.42E-06	2.18E-03
UniRef90_D4M7S8	3.472	0.748	54	19	3.48E-06	2.18E-03
UniRef90_D7JD79	4.892	1.051	54	16	3.22E-06	2.18E-03
UniRef90_D7JEM0	4.771	1.023	54	17	3.08E-06	2.18E-03
UniRef90_D7NDM7	3.186	0.689	54	23	3.82E-06	2.21E-03
UniRef90_Q73PN3	3.978	0.860	54	20	3.74E-06	2.21E-03
UniRef90_A0A139QGR0	3.421	0.745	54	25	4.32E-06	2.40E-03
UniRef90_A8SMN0	3.423	0.750	54	25	5.08E-06	2.48E-03
UniRef90_B2RH89	4.789	1.052	54	15	5.29E-06	2.48E-03

UniRef90_F9DAK5	2.513	0.550	54	30	4.84E-06	2.48E-03
UniRef90_R5FPP1	2.243	0.492	54	28	5.14E-06	2.48E-03
UniRef90_S3JQ41	3.340	0.733	54	21	5.24E-06	2.48E-03
UniRef90_D4M8T7	3.231	0.712	54	21	5.74E-06	2.61E-03
UniRef90_G1WD84	2.469	0.546	54	27	6.05E-06	2.63E-03
UniRef90_G8UHU2	3.862	0.855	54	15	6.23E-06	2.63E-03
UniRef90_U2LIB9	2.540	0.563	54	24	6.47E-06	2.63E-03
UniRef90_UPI000C1C7680	3.753	0.831	54	16	6.30E-06	2.63E-03
UniRef90_D4M7S7	3.993	0.892	54	18	7.51E-06	2.89E-03
UniRef90_S5ZR48	3.595	0.802	54	19	7.41E-06	2.89E-03
UniRef90_I0TC27	2.192	0.491	54	30	7.89E-06	2.96E-03
UniRef90_A0A2A6E736	4.348	0.976	54	16	8.38E-06	3.00E-03
UniRef90_C3J8F0	4.433	0.994	54	19	8.23E-06	3.00E-03
UniRef90_A6KYJ9	2.077	0.468	54	32	8.88E-06	3.10E-03
UniRef90_C8PTE1	3.378	0.763	54	21	9.52E-06	3.18E-03
UniRef90_D4M9R7	3.480	0.786	54	22	9.46E-06	3.18E-03
UniRef90_A0A2A6EBH5	3.717	0.843	54	17	1.04E-05	3.38E-03
UniRef90_A0A1V5I6Y9	4.058	0.926	54	23	1.16E-05	3.59E-03
UniRef90_A7G9R9	2.927	0.667	54	20	1.13E-05	3.59E-03
UniRef90_C8PMN6	2.920	0.666	54	23	1.17E-05	3.59E-03
UniRef90_G5SRK6	3.442	0.787	54	25	1.24E-05	3.71E-03
UniRef90_B2RL46	3.616	0.831	54	20	1.34E-05	3.94E-03
UniRef90_A6KYI5	1.875	0.433	54	33	1.46E-05	4.22E-03
UniRef90_F0F638	2.443	0.565	54	26	1.52E-05	4.24E-03
UniRef90_Q7M7G8	3.752	0.867	54	16	1.51E-05	4.24E-03
UniRef90_D4M779	3.376	0.784	54	21	1.67E-05	4.56E-03
UniRef90_A0A0B4S0X3	3.085	0.720	54	24	1.84E-05	4.93E-03
UniRef90_D7JFK2	5.707	1.334	54	15	1.90E-05	5.00E-03
UniRef90_A0A0A0X218	3.227	0.755	54	17	1.94E-05	5.03E-03
UniRef90_D9RTJ2	2.983	0.700	54	23	2.01E-05	5.11E-03
UniRef90_F0H660	1.911	0.449	54	32	2.09E-05	5.15E-03
UniRef90_Q7M7E2	6.850	1.610	54	14	2.09E-05	5.15E-03
UniRef90_D7JCY7	3.955	0.934	54	17	2.28E-05	5.52E-03
UniRef90_A0A0B4S0P1	2.474	0.588	54	26	2.56E-05	6.10E-03
UniRef90_B2RL49	3.970	0.945	54	16	2.66E-05	6.25E-03
UniRef90_A0A1H9JXP7	2.099	0.501	54	28	2.76E-05	6.39E-03
UniRef90_A0A0Y2EE40	1.249	0.299	54	16	3.03E-05	6.77E-03
UniRef90_D6LIF4	3.622	0.869	54	20	3.07E-05	6.77E-03
UniRef90_U2J131	1.865	0.447	54	28	3.03E-05	6.77E-03
UniRef90_A0A0E9FDV8	2.805	0.674	54	20	3.18E-05	6.92E-03
UniRef90_F5TD07	2.818	0.679	54	23	3.35E-05	7.20E-03

UniRef90_E9S3J1	4.348	1.053	54	16	3.66E-05	7.74E-03
UniRef90_F9PPY8	2.847	0.691	54	35	3.80E-05	7.92E-03
UniRef90_A0A174XW16	3.018	0.733	54	21	3.87E-05	7.96E-03
UniRef90_A0A2L1GQL4	3.182	0.774	54	18	3.93E-05	7.98E-03
UniRef90_A0A133YAZ9	2.207	0.538	54	27	4.10E-05	8.07E-03
UniRef90_A0A1B1IC24	2.517	0.614	54	23	4.14E-05	8.07E-03
UniRef90_Q2FW88	2.707	0.660	54	15	4.11E-05	8.07E-03
UniRef90_F8N9K1	3.889	0.955	54	22	4.64E-05	8.82E-03
UniRef90_I0TC32	2.313	0.568	54	27	4.61E-05	8.82E-03
UniRef90_F8NA49	2.489	0.612	54	25	4.77E-05	8.96E-03
UniRef90_A6LEG7	2.640	0.652	54	28	5.19E-05	9.63E-03

References

Aas, J. A., Paster, B. J., Stokes, L. N., Olsen, I., & Dewhirst, F. E. (2005). Defining the normal bacterial flora of the oral cavity. *Journal of Clinical Microbiology*, *43*(11), 5721–5732. https://doi.org/10.1128/JCM.43.11.5721-5732.2005

Abusleme, L., Dupuy, A. K., Dutzan, N., Silva, N., Burleson, J. A., Strausbaugh, L. D., Gamonal, J., & Diaz, P. I. (2013). The subgingival microbiome in health and periodontitis and its relationship with community biomass and inflammation. *ISME Journal*, *7*(5), 1016–1025. https://doi.org/10.1038/ismej.2012.174

Alberti-Segui, C., Arndt, A., Cugini, C., Priyadarshini, R., & Davey, M. E. (2010). HU protein affects transcription of surface polysaccharide synthesis genes in Porphyromonas gingivalis. *Journal of Bacteriology*, *192*(23), 6217–6229. https://doi.org/10.1128/JB.00106-10

Andreesen, J. R. (2004). Glycine reductase mechanism. *Current Opinion in Chemical Biology*, *8*(5), 454–461. https://doi.org/10.1016/j.cbpa.2004.08.002

Armour, C. R., Nayfach, S., Pollard, K. S., & Sharpton, T. J. (2019). A Metagenomic Meta-analysis Reveals Functional Signatures of Health and Disease in the Human Gut Microbiome. *MSystems*, *4*(4), 1–15. https://doi.org/https://doi.org/10.1128/msystems.00332-18

Barone, M., D’Amico, F., Brigidi, P., & Turroni, S. (2022). Gut microbiome–micronutrient interaction: The key to controlling the bioavailability of minerals and vitamins? In *BioFactors* (Vol. 48, Issue 2, pp. 307–314). John Wiley and Sons Inc. https://doi.org/10.1002/biof.1835

Beghini, F., McIver, L. J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., Mailyan, A., Manghi, P., Scholz, M., Thomas, A. M., Valles-Colomer, M., Weingart, G., Zhang, Y., Zolfo, M., Huttenhower, C., Franzosa, E. A., & Segata, N. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with biobakery 3. *ELife*, *10*. https://doi.org/10.7554/eLife.65088

Belstrøm, D., Constancias, F., Drautz-Moses, D. I., Schuster, S. C., Veleba, M., Mahé, F., & Givskov, M. (2021). Periodontitis associates with species-specific gene expression of the oral microbiota. *Npj Biofilms and Microbiomes*, *7*(1), 76. https://doi.org/10.1038/s41522-021-00247-y

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60. https://doi.org/10.1038/nmeth.3176

Buckel, W., & Thauer, R. K. (2018). Flavin-based electron bifurcation, ferredoxin, flavodoxin, and anaerobic respiration with protons (Ech) or NAD+ (Rnf) as electron acceptors: A historical review. *Frontiers in Microbiology*, *9*(MAR), 1–24. https://doi.org/10.3389/fmicb.2018.00401

Bullard, J. H., Purdom, E., Hansen, K. D., & Dudoit, S. (2010). *Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments*. http://www.biomedcentral.com/1471-2105/11/94

Choi, Y. S., Moon, J. H., Park, J. H., & Lee, J. Y. (2012). Characterization of FimA in Porphyromonas gingivalis genotype IV. *FEMS Immunology and Medical Microbiology*, *65*(3), 497–500. https://doi.org/10.1111/j.1574-695X.2012.00955.x

Costea, P. I., Zeller, G., Sunagawa, S., Pelletier, E., Alberti, A., Levenez, F., Tramontano, M., Driessen, M., Hercog, R., Jung, F. E., Kultima, J. R., Hayward, M. R., Coelho, L. P., Allen-Vercoe, E., Bertrand, L., Blaut, M., Brown, J. R. M., Carton, T., Cools-Portier, S., … Bork, P. (2017). Towards standards for human fecal sample processing in metagenomic studies. *Nature Biotechnology*, *35*(11), 1069–1076. https://doi.org/10.1038/nbt.3960

Cugini, C., Ramasubbu, N., Tsiagbe, V. K., & Fine, D. H. (2021). Dysbiosis From a Microbial and Host Perspective Relative to Oral Health and Disease. *Frontiers in Microbiology*, *12*, 1–23. https://doi.org/10.3389/fmicb.2021.617485

Curtis, M. A., Diaz, P. I., & Van Dyke, T. E. (2020). The role of the microbiota in periodontal disease. *Periodontology 2000*, *83*(1), 14–25. https://doi.org/10.1111/prd.12296

Dashper, S. G., Hendtlass, A., Slakeski, N., Jackson, C., Cross, K. J., Brownfield, L., Hamilton, R., Barr, I., & Reynolds, E. C. (2000). Characterization of a Novel Outer Membrane Hemin-Binding Protein of Porphyromonas gingivalis. *JOURNAL OF BACTERIOLOGY*, *182*(22), 6456–6462. http://www.ncbi.nlm.nih.gov.

Datar, U., Angadi, P. V., Hallikerimath, S., & Kale, A. D. (2013). Cytological assessment of barr bodies using aceto-orcein and papanicolaou stains in buccal mucosal smears and their sex estimation efficacy in an Indian sample. *Acta Cytologica*, *57*(5), 516–521. https://doi.org/10.1159/000353216

Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C. R., Yu, W. H., Lakshmanan, A., & Wade, W. G. (2010). The human oral microbiome. *Journal of Bacteriology*, *192*(19), 5002–5017. https://doi.org/10.1128/JB.00542-10

Diaz, P. I., & Valm, A. M. (2020). Microbial Interactions in Oral Communities Mediate Emergent Biofilm Properties. *Journal of Dental Research*, *99*(1), 18–25. https://doi.org/10.1177/0022034519880157

Didion, J. P., Martin, M., & Collins, F. S. (2017). Atropos: Specific, sensitive, and speedy trimming of sequencing reads. *PeerJ*, *5*, 1–19. https://doi.org/10.7717/peerj.3720

Dillies, M. A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, N. S., Castel, D., Estelle, J., Guernec, G., Jagla, B., Jouneau, L., Laloë, D., Le Gall, C., Schaëffer, B., Le Crom, S., Guedj, M., & Jaffrézic, F. (2013). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in Bioinformatics*, *14*(6), 671–683. https://doi.org/10.1093/bib/bbs046

Dominy, S. S., Lynch, C., Ermini, F., Benedyk, M., Marczyk, A., Konradi, A., Nguyen, M., Haditsch, U., Raha, D., Griffin, C., Holsinger, L. J., Arastu-Kapur, S., Kaba, S., Lee, A., Ryder, M. I., Potempa, B., Mydel, P., Hellvard, A., Adamowicz, K., … Potempa, J. (2019). Porphyromonas gingivalis in

Alzheimer's disease brains: Evidence for disease causation and treatment with small-molecule inhibitors. *Sci Adv .*, *5*(1), 1–21.

Duran-Pinedo, A. E., Chen, T., Teles, R., Starr, J. R., Wang, X., Krishnan, K., & Frias-Lopez, J. (2014). Community-wide transcriptome of the oral microbiome in subjects with and without periodontitis. *ISME Journal*, *8*(8), 1659–1672. https://doi.org/10.1038/ismej.2014.23

Duran-Pinedo, A. E., Solbiati, J., Teles, F., & Frias-Lopez, J. (2023). Subgingival host–microbiome metatranscriptomic changes following scaling and root planing in grade II/III periodontitis. *Journal of Clinical Periodontology*, *50*(3), 316–330. https://doi.org/10.1111/jcpe.13737

Ellison, C. K., Whitfield, G. B., & Brun, Y. V. (2022). Type IV Pili: Dynamic bacterial nanomachines. *FEMS Microbiology Reviews*, *46*(2), 1–14. https://doi.org/10.1093/femsre/fuab053

Ezzo, P. J., & Cutler, C. W. (2003). Microorganisms as risk indicators for periodontal disease. *Periodontology 2000*, *32*, 24–35.

Faust, K., Sathirapongsasuti, J. F., Izard, J., Segata, N., Gevers, D., Raes, J., & Huttenhower, C. (2012). Microbial co-occurrence relationships in the Human Microbiome. *PLoS Computational Biology*, *8*(7), 1–17. https://doi.org/10.1371/journal.pcbi.1002606

Flemming, H. C., & Wingender, J. (2010). The biofilm matrix. In *Nature Reviews Microbiology* (Vol. 8, Issue 9, pp. 623–633). https://doi.org/10.1038/nrmicro2415

Frank, D. N., St Amand, A. L., Feldman, R. A., Boedeker, E. C., Harpaz, N., & Pace, N. R. (2007). *Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases*. www.pnas.org/cgi/content/full/

Frencken, J. E., Sharma, P., Stenhouse, L., Green, D., Laverty, D., & Dietrich, T. (2017). Global epidemiology of dental caries and severe periodontitis – a comprehensive review. *Journal of Clinical Periodontology*, *44*, S94–S105. https://doi.org/10.1111/jcpe.12677

Genco, R. J., Graziani, F., & Hasturk, H. (2020). Effects of periodontal disease on glycemic control, complications, and incidence of diabetes mellitus. In *Periodontology 2000* (Vol. 83, Issue 1, pp. 59–65). Blackwell Munksgaard. https://doi.org/10.1111/prd.12271

Graham, D. B., & Xavier, R. J. (2023). Conditioning of the immune system by the microbiome. In *Trends in Immunology* (Vol. 44, Issue 7, pp. 499–511). Elsevier Ltd. https://doi.org/10.1016/j.it.2023.05.002

Harvey, K. L., Jarocki, V. M., Charles, I. G., & Djordjevic, S. P. (2019). The diverse functional roles of elongation factor tu (Ef-tu) in microbial pathogenesis. *Frontiers in Microbiology*, *10*, 1–19. https://doi.org/10.3389/fmicb.2019.02351

Hempel, C. A., Wright, N., Harvie, J., Hleap, J. S., Adamowicz, S. J., & Steinke, D. (2022). Metagenomics versus total RNA sequencing: most accurate data-processing tools, microbial identification accuracy and perspectives for ecological assessments. *Nucleic Acids Research*, *50*(16), 9279–9293. https://doi.org/10.1093/nar/gkac689

Hood, B. L., & Hirschberg, R. (1995). Purification and Characterization of Eikenella corrodens Type IV Pilin. *INFECTION AND IMMUNITY*, *63*(9), 3693–3696.

Hothorn, T., Hornik, K., Van De Wiel, M. A., & Zeileis, A. (2006). A lego system for conditional inference. *American Statistician*, *60*(3), 257–263. https://doi.org/10.1198/000313006X118430

Huang, R., Li, M., & Gregory, R. L. (2011). Bacterial interactions in dental biofilm. *Virulence*, *2*(5), 435–444. https://doi.org/10.4161/viru.2.5.16140

Huang, Y., Zhao, X., Cui, L., & Huang, S. (2021). Metagenomic and Metatranscriptomic Insight Into Oral Biofilms in Periodontitis and Related Systemic Diseases. *Frontiers in Microbiology*, *12*, 1–12. https://doi.org/10.3389/fmicb.2021.728585

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., Creasy, H. H., Earl, A. M., Fitzgerald, M. G., Fulton, R. S., Giglio, M. G., Hallsworth-Pepin, K., Lobos, E. A., Madupu, R., Magrini, V., Martin, J. C., Mitreva, M., Muzny, D. M., Sodergren, E. J., … White, O. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, *486*(7402), 207–214. https://doi.org/10.1038/nature11234

Iacob, S., Iacob, D. G., & Luminos, L. M. (2019). Intestinal microbiota as a host defense mechanism to infectious threats. In *Frontiers in Microbiology* (Vol. 10, Issue JAN). Frontiers Media S.A. https://doi.org/10.3389/fmicb.2018.03328

Jackson-Rosario, S., & Self, W. T. (2009). Inhibition of selenium metabolism in the oral pathogen Treponema denticola. *Journal of Bacteriology*, *191*(12), 4035–4040. https://doi.org/10.1128/JB.00164-09

Jakubovics, N. S., Goodman, S. D., Mashburn-Warren, L., Stafford, G. P., & Cieplik, F. (2021). The dental plaque biofilm matrix. *Periodontology 2000*, *86*(1), 32–56. https://doi.org/10.1111/prd.12361

Jorth, P., Turner, K. H., Gumus, P., Nizam, N., Buduneli, N., & Whiteley, M. (2014). Metatranscriptomics of the Human Oral Microbiome during Health and Disease. *MBio*, *5*(2). https://doi.org/10.1128/mBio.01012-14

Karched, M., Bhardwaj, R. G., Qudeimat, M., Al-Khabbaz, A., & Ellepola, A. (2022). Proteomic analysis of the periodontal pathogen Prevotella intermedia secretomes in biofilm and planktonic lifestyles. *Scientific Reports*, *12*(5636). https://doi.org/10.1038/s41598-022-09085-0

Karlsson, F. H., Tremaroli, V., Nookaew, I., Bergström, G., Behre, C. J., Fagerberg, B., Nielsen, J., & Bäckhed, F. (2013). Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature*, *498*(7452), 99–103. https://doi.org/10.1038/nature12198

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360. https://doi.org/10.1038/nmeth.3317

Kopylova, E., Noé, L., & Touzet, H. (2012). SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, *28*(24), 3211–3217. https://doi.org/10.1093/bioinformatics/bts611

Kullaa, A. M., Asikainen, P., Herrala, M., Ukkonen, H., & Mikkonen, J. J. W. (2014). Microstructure of oral epithelial cells as an underlying basis for salivary mucosal pellicle. *Ultrastructural Pathology*, *38*(6), 382–386. https://doi.org/10.3109/01913123.2014.944732

Kullaa-Mikkonen, A., & Sorvari, T. E. (1985). A scanning electron microscopy study of the dorsal human tongue. *Acta Anat.*, *123*, 114–120.

Kurniyati, K., Chang, Y., Guo, W., Liu, J., Malkowski, M. G., & Li, C. (2023). Anti-s28 Factor FlgM Regulates Flagellin Gene Expression and Flagellar Polarity of Treponema denticola. *Journal of Bacteriology*, *205*(2), 1–16. https://doi.org/10.1128/jb.00463-22

Kurtz, Z. D., Müller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J., & Bonneau, R. A. (2015). Sparse and Compositionally Robust Inference of Microbial Ecological Networks. *PLoS Computational Biology*, *11*(5), 1–25. https://doi.org/10.1371/journal.pcbi.1004226

Lambring, C. B., Siraj, S., Patel, K., Sankpal, U. T., Mathew, S., & Basha, R. (2019). Impact of the microbiome on the immune system. *Critical Reviews in Immunology*, *39*(5), 313–328. https://doi.org/10.1615/CritRevImmunol.2019033233

Langmead, B., & Salzberg, S. L. (2013). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lasserre, J. F., Brecx, M. C., & Toma, S. (2018). Oral microbes, biofilms and their role in periodontal and peri-implant diseases. *Materials*, *11*(10), 1–17. https://doi.org/10.3390/ma11101802

Lee, Y. H. (2018). An overview of meta-analysis for clinicians. *Korean Journal of Internal Medicine*, *33*(2), 277–283. https://doi.org/10.3904/kjim.2016.195

Lin, H., & Peddada, S. Das. (2020). Analysis of microbial compositions: a review of normalization and differential abundance analysis. In *npj Biofilms and Microbiomes* (Vol. 6, Issue 1). Nature Research. https://doi.org/10.1038/s41522-020-00160-w

Liu, B., Zheng, D., Zhou, S., Chen, L., & Yang, J. (2022). VFDB 2022: A general classification scheme for bacterial virulence factors. *Nucleic Acids Research*, *50*(Database issue), D912–D917. https://doi.org/10.1093/nar/gkab1107

Lloyd-Price, J., Abu-Ali, G., & Huttenhower, C. (2016). The healthy human microbiome. *Genome Medicine*, *8*(1), 1–11. https://doi.org/10.1186/s13073-016-0307-y

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12). https://doi.org/10.1186/s13059-014-0550-8

Mager, D. L., Ximenez-Fyvie, L. A., Haffajee, A. D., & Socransky, S. S. (2003). Distribution of selected bacterial species on intraoral surfaces. *Journal of Clinical Periodontology*, *30*(7), 644–654. https://doi.org/10.1034/j.1600-051X.2003.00376.x

Mahalakshmi, K., Krishnan, P., & Chandrasekaran, S. C. (2018). Detection of Tannerella forsythia bspA and prtH genotypes among periodontitis patients and healthy subjects—A case—Control study. *Archives of Oral Biology*, *96*, 178–181. https://doi.org/10.1016/j.archoralbio.2018.09.012

Mallick, H., Rahnavard, A., McIver, L. J., Ma, S., Zhang, Y., Nguyen, L. H., Tickle, T. L., Weingart, G., Ren, B., Schwager, E. H., Chatterjee, S., Thompson, K. N., Wilkinson, J. E., Subramanian, A., Lu, Y., Waldron, L., Paulson, J. N., Franzosa, E. A., Bravo, H. C., & Huttenhower, C. (2021). Multivariable association discovery in population-scale meta-omics studies. *PLoS Computational Biology*, *17*(11), 1–27. https://doi.org/10.1371/journal.pcbi.1009442

Mark Welch, J. L., Ramírez-Puebla, S. T., & Borisy, G. G. (2020). Oral Microbiome Geography: Micron-Scale Habitat and Niche. In *Cell Host and Microbe* (Vol. 28, Issue 2, pp. 160–168). Cell Press. https://doi.org/10.1016/j.chom.2020.07.009

Molina, A., Huck, O., Herrera, D., & Montero, E. (2023). The association between respiratory diseases and periodontitis: A systematic review and meta-analysis. In *Journal of Clinical Periodontology* (Vol. 50, Issue 6, pp. 842–887). John Wiley and Sons Inc. https://doi.org/10.1111/jcpe.13767

Morillo-Lopez, V., Sjaarda, A., Islam, I., Borisy, G. G., & Mark Welch, J. L. (2022). Corncob structures in dental plaque reveal microhabitat taxon specificity. *Microbiome*, *10*(1). https://doi.org/10.1186/s40168-022-01323-x

Nwizu, N., Wactawski-Wende, J., & Genco, R. J. (2020). Periodontal disease and cancer: Epidemiologic studies and possible mechanisms. In *Periodontology 2000* (Vol. 83, Issue 1, pp. 213–233). Blackwell Munksgaard. https://doi.org/10.1111/prd.12329

Ojala, T., Kankuri, E., & Kankainen, M. (2023). Understanding human health through metatranscriptomics. *Trends in Molecular Medicine*, *29*(5), 376–389. https://doi.org/10.1016/j.molmed.2023.02.002

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., … Pruitt, K. D. (2016). Reference sequence

(RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, *44*(Database issue), D733–D745. https://doi.org/10.1093/nar/gkv1189

Paulson, J. N., Colin Stine, O., Bravo, H. C., & Pop, M. (2013). Differential abundance analysis for microbial marker-gene surveys. *Nature Methods*, *10*(12), 1200–1202. https://doi.org/10.1038/nmeth.2658

Pérez-Chaparro, P. J., Gonçalves, C., Figueiredo, L. C., Faveri, M., Lobão, E., Tamashiro, N., Duarte, P., & Feres, M. (2014). Newly identified pathogens associated with periodontitis: A systematic review. *Journal of Dental Research*, *93*(9), 846–858. https://doi.org/10.1177/0022034514542468

Peschel, S., Müller, C. L., Von Mutius, E., Boulesteix, A. L., & Depner, M. (2021). NetCoMi: Network construction and comparison for microbiome data in R. In *Briefings in Bioinformatics* (Vol. 22, Issue 4, pp. 1–18). Oxford University Press. https://doi.org/10.1093/bib/bbaa290

Petersen, C., & Round, J. L. (2014). Defining dysbiosis and its influence on host immunity and disease. *Cellular Microbiology*, *16*(7), 1024–1033. https://doi.org/10.1111/cmi.12308

Radaic, A., & Kapila, Y. L. (2021). The oralome and its dysbiosis: New insights into oral microbiome-host interactions. *Computational and Structural Biotechnology Journal*, *19*, 1335–1360. https://doi.org/10.1016/j.csbj.2021.02.010

Rajam, G., Anderton, J. M., Carlone, G. M., Sampson, J. S., & Ades, E. W. (2008). Pneumococcal surface adhesin A (PsaA): A review. In *Critical Reviews in Microbiology* (Vol. 34, Issues 3–4, pp. 131–142). https://doi.org/10.1080/10408410802275352

Relvas, M., Regueira-Iglesias, A., Balsa-Castro, C., Salazar, F., Pacheco, J. J., Cabral, C., Henriques, C., & Tomás, I. (2021). Relationship between dental and periodontal health status and the salivary microbiome: bacterial diversity, co-occurrence networks and predictive models. *Scientific Reports*, *11*(929). https://doi.org/10.1038/s41598-020-79875-x

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Rowland, I., Gibson, G., Heinken, A., Scott, K., Swann, J., Thiele, I., & Tuohy, K. (2018). Gut microbiota functions: metabolism of nutrients and other food components. In *European Journal of Nutrition* (Vol. 57, Issue 1). Dr. Dietrich Steinkopff Verlag GmbH and Co. KG. https://doi.org/10.1007/s00394-017-1445-8

Schaumann, T., Kraus, D., Winter, J., Wolf, M., Deschner, J., & Jäger, A. (2013). Potential immune modularly role of glycine in oral gingival inflammation. *Clinical and Developmental Immunology*, *2013*, 1–9. https://doi.org/10.1155/2013/808367

Sender, R., Fuchs, S., & Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biology*, *14*(8), 1–14. https://doi.org/10.1371/journal.pbio.1002533

Sharma, A., Sojar, H. T., Glurich, I., Honma, K., Kuramitsu, H. K., & Genco, R. J. (1998). Cloning, Expression, and Sequencing of a Cell Surface Antigen Containing a Leucine-Rich Repeat Motif from Bacteroides forsythus ATCC 43037. *INFECTION AND IMMUNITY*, *66*(12), 5703–5710.

Simpson, W., Olczak, T., & Genco, C. A. (2000). Characterization and Expression of HmuR, a TonB-Dependent Hemoglobin Receptor of Porphyromonas gingivalis. *JOURNAL OF BACTERIOLOGY*, *182*(20), 5737–5748.

Skoufos, G., Almodaresi, F., Zakeri, M., Paulson, J. N., Patro, R., Hatzigeorgiou, A. G., & Vlachos, I. S. (2022). AGAMEMNON: an Accurate metaGenomics And MEtatranscriptoMics quaNtificatiON analysis suite. *Genome Biology*, *23*(39), 1–27. https://doi.org/10.1186/s13059-022-02610-4

Socransky, Haffaiee A. D., Cugini M. A., Smith C., & Kent R. L. Jr. (1998). Microbial complexes in subgingival plaque. *J Clin Periodontol*, *25*, 134–144.

Szafrański, S. P., Deng, Z. L., Tomasch, J., Jarek, M., Bhuju, S., Meisinger, C., Kühnisch, J., Sztajer, H., & Wagner-Döbler, I. (2015). Functional biomarkers for chronic periodontitis and insights into the roles of Prevotella nigrescens and Fusobacterium nucleatum; a metatranscriptome analysis. *Npj Biofilms and Microbiomes*, *1*(15017), 1–13. https://doi.org/10.1038/npjbiofilms.2015.17

Teles, R., Teles, F., Frias-Lopez, J., Paster, B., & Haffajee, A. (2013). Lessons learned and unlearned in periodontal microbiology. *Periodontology 2000*, *62*, 95–162.

Tufanaru, C., Munn, Z., Stephenson, M., & Aromataris, E. (2015). Fixed or random effects meta-analysis? Common methodological issues in systematic reviews of effectiveness. *International Journal of Evidence-Based Healthcare*, *13*(3), 196–207. https://doi.org/10.1097/XEB.0000000000000065

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The Human Microbiome Project. *Nature*, *449*(7164), 804–810. https://doi.org/10.1038/nature06244

Valdes, A. M., Walter, J., Segal, E., & Spector, T. D. (2018). Role of the gut microbiota in nutrition and health. *BMJ (Online)*, *361*, 36–44. https://doi.org/10.1136/bmj.k2179

Van Dyke, T. E., Kholy, K. El, Ishai, A., Takx, R. A. P., Mezue, K., Abohashem, S. M., Ali, A., Yuan, N., Hsue, P., Osborne, M. T., & Tawakol, A. (2021). Inflammation of the periodontium associates with risk of future cardiovascular events. *Journal of Periodontology*, *92*(3), 348–358. https://doi.org/10.1002/JPER.19-0441

Walker, E., Herbandez, A. V., & Kattan, M. W. (2008). Meta-analysis: Its strengths and limitations. *Cleveland Clinic Journal of Medicine*, *75*(6), 431–439.

Wirbel, J., Pyl, P. T., Kartal, E., Zych, K., Kashani, A., Milanese, A., Fleck, J. S., Voigt, A. Y., Palleja, A., Ponnudurai, R., Sunagawa, S., Coelho, L. P., Schrotz-King, P., Vogtmann, E., Habermann, N., Niméus, E., Thomas, A. M., Manghi, P., Gandini, S., … Zeller, G. (2019). Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nature Medicine*, *25*(4), 679–689. https://doi.org/10.1038/s41591-019-0406-6

Xu, X., He, J., Xue, J., Wang, Y., Li, K., Zhang, K., Guo, Q., Liu, X., Zhou, Y., Cheng, L., Li, M., Li, Y., Li, Y., Shi, W., & Zhou, X. (2015). Oral cavity contains distinct niches with dynamic microbial communities. *Environmental Microbiology*, *17*(3), 699–710. https://doi.org/10.1111/1462-2920.12502

Yoshimura, F., Murakami, Y., Nishikawa, K., Hasegawa, Y., & Kawaminami, S. (2009). Surface components of Porphyromonas gingivalis. *Journal of Periodontal Research*, *44*(1), 1–12. https://doi.org/10.1111/j.1600-0765.2008.01135.x

Yost, S., Duran-Pinedo, A. E., Teles, R., Krishnan, K., & Frias-Lopez, J. (2015). Functional signatures of oral dysbiosis during periodontitis progression revealed by microbial metatranscriptome analysis. *Genome Medicine*, *7*(27), 1–19. https://doi.org/10.1186/s13073-015-0153-3

Zhang, Y., Li, Y., Yang, Y., Wang, Y., Cao, X., Jin, Y., Xu, Y., Li, S. C., & Zhou, Q. (2022). Periodontal and Peri-Implant Microbiome Dysbiosis Is Associated With Alterations in the Microbial Community Structure and Local Stability. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/fmicb.2021.785191

Zheng, D., Liwinski, T., & Elinav, E. (2020). Interaction between microbiota and immunity in health and disease. In *Cell Research* (Vol. 30, Issue 6, pp. 492–506). Springer Nature. https://doi.org/10.1038/s41422-020-0332-7