



Quantitative Insights into Microbial Ecology software allows the dissection of oral microbial diversity

Nik Norliza Nik Hassan^{1*}, Rozita Hassan², Zuliani Mahmood², NurWaliyuddin Hanis Zainal Abidin¹, Kaamini Ravy¹

¹ School of Health Sciences, Health Campus, Universiti Sains Malaysia, 16150 Kelantan

² School of Dentistry, Health Campus, Universiti Sains Malaysia, 16150 Kelantan

Abstract

Objective: Recent approach based on the 16S rRNA gene sequence, which bypasses the need for culturing was being used to assess the diversity and relative abundance of bacteria in the carious patients. Here, we report the use of 16S rRNA V3-V4 marker gene approach to dissect the oral microbial diversity in patients' caries. **Methods:** A total of 8 subjects with caries aged between 20-30 years old from Dental Clinic, Hospital Universiti Sains Malaysia, Kelantan were recruited. The DNA from collected plaque samples were examined by high-throughput sequencing of the V3-V4 hypervariable region of the 16S rRNA gene using MiSeq sequencing platform. For 16S rRNA amplicon analysis, the Quantitative Insights into Microbial Ecology (QIIME) programme was utilised and the Greengenes database was used for taxonomy classification. **Results:** By using this platform, we identified six top communities of microbial in the caries samples. Protobacteria OTUs were most dominant in all samples with abundance value of (47.4%), followed by Firmicutes (21%), Bacteroidetes (15.6%), Actinobacteria (7.5%) and Fusobacteria (5.2%) and TM7 (3.2%). **Conclusion:** The QIIME is a robust platform that is able to support a wide range of microbial community analyses and to classify the oral microbial taxa in the taxonomic hierarchy i.e. from kingdom, phylum, class, order, family, genus to species level.

Keywords: Metagenomic analysis, open-reference Operational taxonomic units (OTU), Quantitative insights into microbial ecology (QIIME), taxonomy

Introduction

Caries is among the major dental diseases caused by oral bacteria. It is a progressive disease that expands from the initial focus of enamel degradation and the subsequent exposure of the underlying dentine to a point where the microorganism gains access to the tubular network of dentine that extent to the dental pulp (Love et al., 2000). Extension of the lesion can lead to necrosis of the pulp tissue by toxic action and if left untreated, it can involve the supporting bone where cellulitis is the potential outcome (Love et al., 2002). Even though it is not exactly infectious in the classical sense because it is not contagious, many studies have reported the oral bacteria as an aetiological agent for other oral conditions such as alveolar osteitis, tonsillitis and linked to certain

systemic diseases (aspiration pneumonia and cardiovascular) (Duran-Pinedo & Frias-Lopez, 2015). Nonetheless, many of these oral bacteria remained uncharacterized (Martin et al., 2002).

Characterisation of microbial genetic profiles using culture-independent 16S ribosomal RNA (16S rRNA) gene sequencing offers a room to understand the microbial community structure. The small subunit 16S rRNA gene is targeted because of its universal presence in cellular organisms. The 16S rRNA is uniquely found in prokaryotes and has become the important biomarker to identify specific microbes using the nine variable regions, which approximately 1,500 bp long (Ahn et al., 2011a; Chaudhary et al., 2015). These polymorphic regions are frequently used in phylogenetic classification. The 16S rRNA sequence analysis will provide an insight into the diversity and community structure of the human microbiome in relation to health and disease.

Received: 17 January 2022; **Accepted revised manuscript:**

16 February 2022 **Published online:** 30 March 2022

***Corresponding author:** Nik Norliza Nik Hassan,
BioMedicine Programme, School of Health Sciences, Universiti
Sains Malaysia, 16150 Kota Bharu, Kelantan, Malaysia

Tel: +6097677624 **Email:** nnorliza@usm.my

The prokaryotic 16S rRNA gene analysis is commonly performed using metagenomic approach. Metagenomic approach has the advantage of providing much richer data especially on the functional potential present in microbial communities. In this study, the diversity of caries microbiota was estimated by sequencing the V3 and V4 regions of 16S rRNA that allows the population analysis without cultivation of individual microbes.

We also discussed on the bioinformatic software used to analyse the sequencing data that emanate from both target amplicon and metagenomic sequencing analysis. To the best of our knowledge, this is the first study conducted among Kelantan population employing 16S rRNA metagenomics platform that targets V3-V4 variable region for dissection of oral microbial diversity. To identify the microbes, 16S rRNA amplicon sequencing was used to sequence the V3-V4 hypervariable region, and all the effective tags of all samples were clustered, with those sequences with over 97% similarity were considered as one Operational Taxonomic Unit (OTU) for further analysis using Quantitative Insights into Microbial Ecology (QIIME). The primary reason for using the V3 region is because it contains the most nucleotide heterogeneity and has the greatest discriminatory power (Chakravorty et al., 2007). According to (Graspeuntner et al., 2018), the V3-V4 area detects more taxa and has more variety than the V1-V2 region. In addition, the V3-V4 area identifies community state types with distinct species that are lacking in the V1-V2 region.

QIIME is an open-source software pipeline built using the PyCogent toolkit to address the problem of taking sequencing data from raw sequences for interpretation and database deposition. The software support wide range of microbial community analyses and visualisation that have been central to several recent high-profile studies including network analysis, histograms of within- or between samples diversity. QIIME aids this study by providing graphical displays that allows us to interact with the data. This modulatory allows alternative component for functionalities such as choosing OTUs, sequence alignment, inferring phylogenetic trees and phylogenetic and taxon-based analysis of diversity within (alpha diversity) and between samples (beta diversity).

Materials and methods

Ethics approval

All related research protocol and procedures were approved by the Human Research Ethics Committee of Universiti Sains Malaysia (USM/JEPeM/18120787). All subjects were provided with written informed consent.

Subjects and sampling

A total of eight subjects with caries were selected from patients attending the Dental Clinic, Hospital Universiti Sains Malaysia for dental treatment after considering the inclusion and exclusion criteria. The inclusion criteria include individuals between the age of 20-30 years old and possess carious tooth score of five and six according to the International Caries Detection and Assessment System (ICDAS). Those subjects who had systemic diseases, craniofacial syndrome or wearing oral appliances were excluded. Plaque samples were removed by dental practitioner and stored in deionized water and kept at -20°C until further analysis.

DNA extraction and quantification

DNA was extracted from plaque samples using MasterPure™ Complete DNA and RNA Purification Kit (Lucigen, USA) as described in the manufacturer's protocol. The extracted DNA samples were viewed using gel electrophoresis on a 1% agarose gel along with a λ *Hind*III DNA ladder at 85 V for 35 minutes. The concentration and purity of genomic DNA samples were evaluated spectrophotometrically using NanoDrop 2000™ (Thermo Fisher Scientific, USA) based on the ratio of A260/A280.

Creation of the 16S rRNA V3-V4 amplicon library

First stage PCR

An amplicon library from individual samples was created by PCR amplification with unique barcoded primers of the 16S rRNA V3-V4 gene region. The primer sequences of 16S amplicon were shown in Table 1.

Table 1. Primer sequences for 16S rRNA of V3 -V4 gene regions.

16S amplicon primer	PCR Primer Sequence (5'-3')
Forward primer	TCGTCGGCAGCGTC AGATGTGTATAAGA GACAGCCTACGGGN GGCWGCAG
Reverse primer	GTCTCGTGGGCTCG GAGATGTGTATAAG AGACAGGACTACHV GGGTATCTAATCC

PCR was carried out using 2x KAPA HiFi HotStart kit with the thermal profile of 1 cycle at 95°C for 3 minutes followed by 25 cycles at 95°C for 30 seconds, 55°C for 30 seconds and extension at 72°C for another 30 seconds. Another additional of extension step were applied at 72°C for another 5 minutes before the reaction was kept on hold at 4°C.

First PCR clean-up, Index PCR and second PCR clean-up

The 16S rRNA V3-V4 amplicon were subjected to clean-up phase using AMPure XP beads (Beckman Coulter, USA). The index PCR refers to the attachment of dual indices and Illumina sequencing adapters using the Nextera XT Index kit. The second PCR was carried out following the thermal profile of 1 cycle at 95°C for 3 minutes and another 8 cycles at 95°C for 30 seconds, primer annealing at 55°C for 30 seconds and extension at 72°C for 30 seconds. An additional extension phase was applied for 5 minutes at 72°C and followed by the second PCR clean-up using AMPure XP beads.

Library quantification, normalization, and pooling

Fluorometric quantification method was performed to quantify the DNA libraries using dsDNA binding dyes. DNA concentration (nM) was calculated based on the size of DNA amplicons as determined by an Agilent Technologies 2100 Bioanalyzer trace:

$$\frac{\text{Concentration (ng/ul)}}{(660\text{g/mol} \times \text{average library size})} \times 10^6 = \text{concentration (nM)}$$

Library denaturing and MiSeq Sample pooling

Sequencing of the amplified V3-V4 gene region was performed using MiSeq sequencing platform. A minimum of 5% PhiX was served as an internal

control for the low diversity libraries. The complete protocol of the sequencing is as mentioned in the user manual of Illumina MiSeq System (cat number: 15044223). In brief, sequencing on MiSeq using 300 bp paired reads and MiSeq Reagents V3 kit. The ends of reads are overlapped to generate high quality and full length reads of V3 and V4 regions. The single sequencing run was completed in 65 hours.

MiSeq Reporter Metagenomics Workflow

The bioinformatics data analysis of 16S metagenomics involved qualitative and quantitative methods to assess diversity through 16S rRNA. The pre-processing steps included were the quality filtering, adapter removal, error-correction and chimera removal (Fig. 1). The raw sequences data underwent strict quality and size filtering, removing beads shorter than 150bp as well as those mismatches and poor-quality scores using FASTX Toolkit. Sequences were error-corrected by using de novo chimera detection with USEARCH against RDP Gold v9 database. Chimeras were filtered with 8.7% accuracy. The verified sequences were characterised for diversity and taxonomic composition using QIIME suite version 1.9. Alpha and beta diversity measurements and significant tests were also performed within the samples.

The secondary data analysis was executed by using MiSeq Reporter software (MSR) following the 16S metagenomic protocols. The workflow classified organisms from the V3-V4 amplicons based on the Greengenes database (<http://greengenes.lbl.gov/>). Sequences were first clustered into OTUs using phyton (.py) scripts. The OTUs picking were based on the sequence similarities and assembled the OTU abundances in each sample with taxonomic identifiers for each OTU. The output of this step was formatted in Biological Observation Matrix (BIOM) file. The BIOM file is the classification of reads at several taxonomic levels such as kingdom, phylum, class, order, family, genus, and species.

Statistical analysis

QIIME was used to create the heatmaps and estimate the following alpha-diversity metrics: raw number of OTU's per samples, Chao 1 estimator, Shannon entropy, non-metric dimensional scaling; and Bray-Curtis distance metric. The Chao1 index method was used to estimate the richness based on the singletons (OTUs with single appearance) and doubletons (OTUs that appeared twice). It was used to estimate the missing species since missing species information is mostly focused on the low frequency counts.

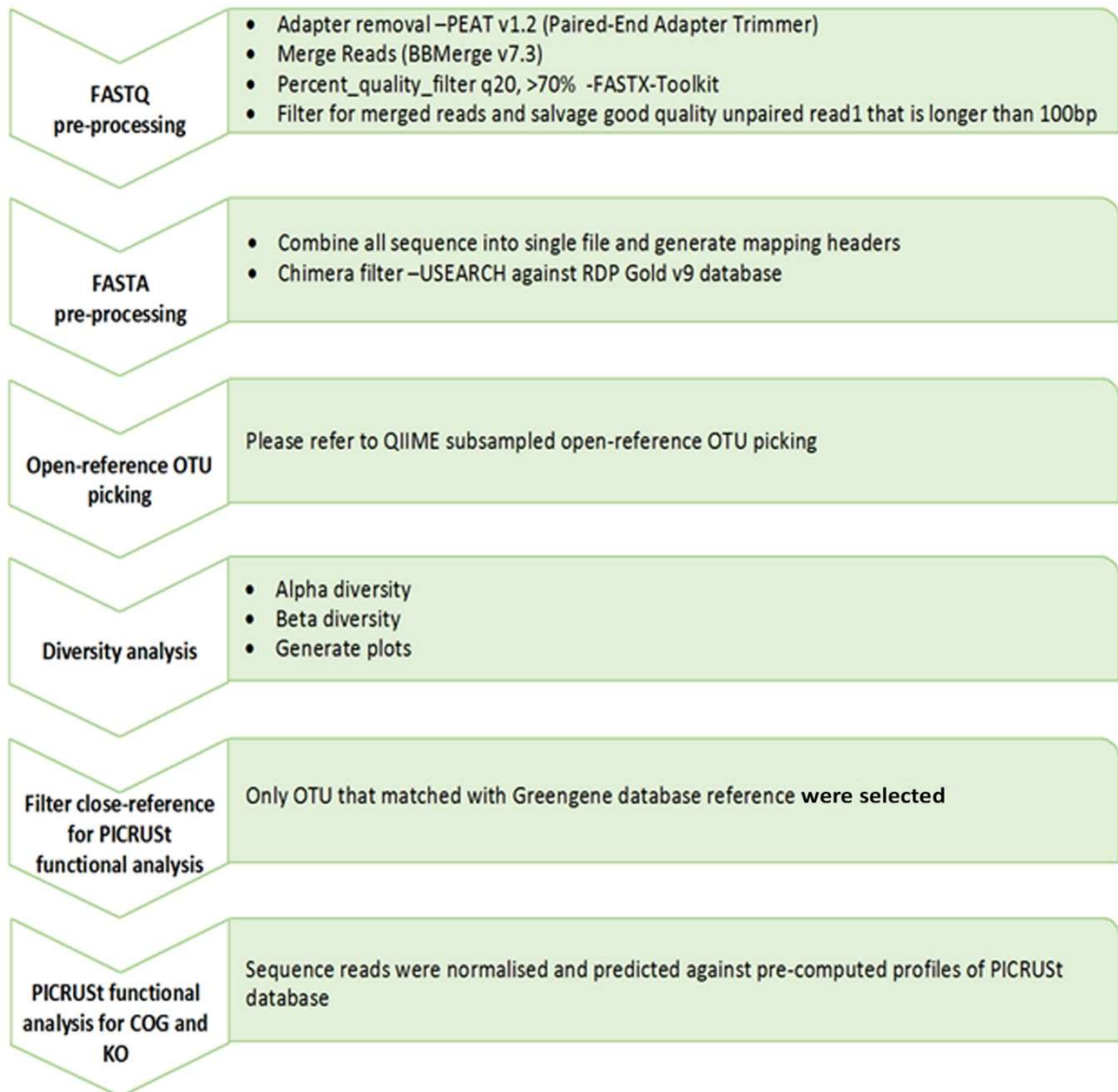


Figure 1. Steps involved in the 16S rRNA data processing. A typical analysis path for metagenomic analysis when raw sequences data passed through quality filtering known as chimera to minimize the effects of sequencing artifacts. The resulting filtered sequence reads are clustered into operational taxonomic units (OTUs), representing similar organisms, phylogenetic and taxonomic identity. The abundance of the various OTUs were then subjected to a variety of multivariate analysis and visualization procedures to elucidate the structure and patterns of the microbial communities.

Results and Discussion

Here, we present a report on the identification of predominant bacteria flora of the carious oral cavity by using 16S rRNA MiSeq sequencing of the V3-V4 hypervariable region. We acquired a set of 7,556,908 raw reads from eight samples used in the study (Table 2). The number of sequences per samples ranged from to 516,300 to 1,294,105, with average of 944,613 read sequences. Table 2 shows

the total number of OTUs for each sample tested before and after the quality control steps.

Table 2. Total number of sequences and Operational Taxonomic Units (OTU) of each sample.

Sample ID	Total number of sequences	Total number of OTUs (before QC)	Total number of OTUs (after QC)
2	1,003,136	1,316,136	1,134,742
5	1,114,608	1,566,526	1,293,702
6	516,300	1,469,857	1,217,619
A	1,294,105	1,762,178	1,472,292
B	1,233,761	1,661,815	1,416,017
C	727,086	1,389,354	1,195,449
D	724,729	1,540,312	1,306,731
WAA	943,183	1,674,861	1,441,425

The relative abundance values (%) and taxonomic profiles of microbes at the phyla level for each sample tested are depicted in Table 3 and Figure 2, respectively. Overall, Proteobacteria OTUs were the most dominant in all samples with total average abundance value of 47.5%. This was followed by Firmicutes (21%), Bacteroidetes (15.6%), Actinobacteria (7.5%), Fusobacteria (5.2%) and TM7 (3.2%). The abundant bacterial groups found in our study are coherent to those found in most other studies (Aas et al., 2005; Ahn

et al., 2011b; Dewhirst et al., 2010; Peterson et al., 2013). Similarity in research papers in terms of bacterial community indicates relatively steady bacterial community structure, suggesting that disease status may not markedly influence bacterial composition.

Individually, Protobacteria was the most dominant in majority of the samples tested with the highest abundance reported in samples WAA and B (91.7 and 73.9%, respectively). Firmicutes and Bacteroidetes were the most abundance identified in sample 6 (68.8%) and 5 (37.2%), respectively. There were three predominant microbes reported in sample A, namely Fusobacteria, Firmicutes, and Protobacteria.

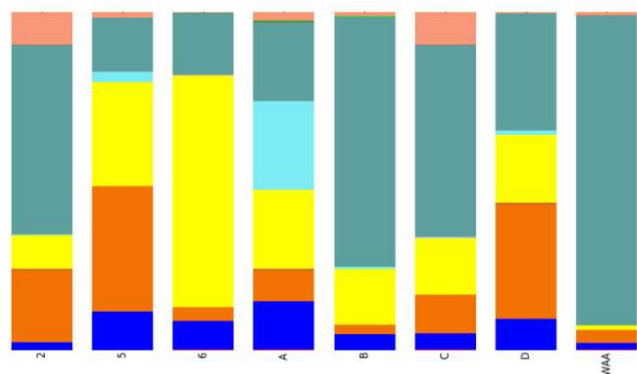


Figure 2. Taxonomic profiles at the phyla levels of eight samples.

Table 3. The percentage of microbial present in each sample.

Dominant phyla	Percentage of microbe in each sample, %								
	2	5	6	A	B	C	D	WAA	Total Average
k_Bacteria;p_Proteobacteria	56.0	16.0	18.6	23.5	73.9	56.8	34.8	91.7	47.5
k_Bacteria;p_Firmicutes	10.2	30.8	68.8	23.6	16.5	16.7	20.3	1.2	21.0
k_Bacteria;p_Bacteroidetes	21.5	37.2	4.1	9.4	2.9	11.5	34.1	3.9	15.6
k_Bacteria;p_Actinobacteria	2.4	11.4	8.7	14.5	4.7	4.9	9.3	2.2	7.5
k_Bacteria;p_usobacteria	0.1	3.2	0	26.3	0.7	0.3	1.2	0.0	5.2
k_Bacteria;p_TM7	9.6	1.6	0	2.5	1.2	9.6	0.2	0.8	3.2

Abbreviation: TM7 is *Candidatus saccharibacteria*.

Proteobacteria is a major phylum of Gram-negative bacteria which includes wide variety of pathogenic genera such as *Escherichia coli*, *Salmonella*, *Vibrio*, *Helicobacter*, *Yersinia*, and *Legionella*.

The Firmicutes most of which possess Gram-positive cell wall structure. Firmicutes also called the low-G+C group, having round cells (singular cocci or sometimes rod-like forms known as bacillus). Many Firmicutes produce endospores,

which are resistant to desiccation and able to survive extreme conditions. They are found in various environments, and the group includes some notable pathogens. Firmicutes play an important role in beer, wine, and cider spoilage (Haakensen et al., 2008).

The phylum Bacteroidetes is composed of three large classes of Gram-negative, non-spore forming, anaerobic or aerobic, and rod-shaped bacteria that are widely distributed in the environment, including in soil, sediments, and sea

water, as well as in the guts and on the skin of animals. Although some *Bacteroides* spp. can be opportunistic pathogens, many of them are symbiotic species highly adjusted to the gastrointestinal tract. They are highly abundant in intestines, reaching up to 10^{11} cells g^{-1} of intestinal material. *Bacteroidetes* colonization also found in the infants gastrointestinal (Rajilić-Stojanović & de Vos, 2014).

The Actinobacteria are a phylum of Gram-positive bacteria with high G+C in their DNA. It is one of the dominant bacterial phyla and contains one of the largest of bacterial genera, *Streptomyces*. *Streptomyces* and other actinobacteria are major contributors to biological buffering of soils. They are also the source of many antibiotics. Most Actinobacteria of medical or economic significance are in subclass Actinobacteriacea and belong to the order Actinomycetales. While many of these cause disease in human, *Streptomyces* is notable as a source of antibiotics, especially *Streptomyces* spp., are recognized as the producers of many bioactive metabolites especially in medicine (anti-bacterials, anti-fungals, anti-virals) (Ventura et al., 2007).

Fusobacterium is a genus of anaerobic, Gram-negative, non-spore forming bacteria, similar to *Bacteroides*. Individual cells are slender rod-shaped bacilli with pointed ends. Strains of *Fusobacterium* cause several human diseases, including periodontal diseases, Lemierre's syndrome, and topical skin ulcers. Although older sources state that *Fusobacterium* is part of the normal flora of the human oropharynx, the current consensus is that *Fusobacterium* should always be treated as a pathogen (Bennett & Eley, 1993).

Phyla TM7 also known as *Nanosynbacter lyticus* type strain TM7x HMT 952 is the only member of the candidate phylum that has been cultivated successfully from the human oral cavity. The cultivated oral taxon is designated as Saccharibacteria oral taxon TM7x. TM7x has a unique lifestyle in comparison to other bacteria that are associated with human. It is an obligate epibiont parasite, or an "epiparasite", growing on the surface of its host bacterial species *Actinomycesodontolyticus* subspecies actinosynbacter strain XH001, which is referred to as the "basibiont". *Actinomyces* species are one of the early microbial colonisers in the oral cavity. Together with *Actinomyces*, they exhibit parasitic epibiont symbiosis (He et al., 2015; Bor et al., 2017).

It is well documented that the oral bacteria is able to colonise different sites of the oral cavity due to the specific attachment of the bacterial surface to complementary specific receptors (Aas et al, 2005).

For instance, Mager et al. (2003) reported that 40 different profiles of culturable bacterial species had been detected on different oral soft tissue surfaces, saliva, supra-gingival and sub-gingival plaques of healthy individual. The crucial reason researchers are interested in oral microbes is in their potential to serve as biomarkers for systemic diseases.

The microbial diversity score obtained indicated that the oral microbiome in samples 5 and A were more diverse in terms of species richness compared to other subjects. The presence of *Candidatus saccharibacteria* (TM7) in the subjects; was less abundant but relatively higher in two samples with ID of 2 and C. The exclusive phyla found in all the subjects indicated that the individual variability of the oral microbiome was due to the persistence of subject-specific taxa whose frequency fluctuated between the time points, as shown in previous studies. The composition of the microbial communities on and within the human body varies between individuals. Inter-individual variation had been shown in various studies for healthy intestinal tract (Bik et al., 2010).

Although it is universally acknowledged that bacteria play a key role in dental caries, they are not the only factor. Microbiomes of other types can also influence disease progression. Fungal-bacterial ecological interactions have attracted researchers' attention in recent years. For example, *Candida albicans* was reported to have a strong effect on early children's caries and root caries (Jiang et al., 2019). Further studies are needed to investigate other microbiomes and their relationships in the caries process. Furthermore, environmental conditions such as temperature, pH conditions, salinity, redox potential; and access to oxygen or nutrients, all have some impact on the microbial composition of biofilms. Temporary changes in the oral microbiome may result from diet, salivary flow; and long-term use of antibiotics (Struzycka, 2014). However, the number of subjects in this study was relatively small, thus inter-individual differences associated with gender, age or ethnicity might become apparent if larger number of subjects were studied.

Conclusion

This study had demonstrated the importance of most bacterial taxa that had never been detected in our environment using traditional culture methods; however, by employing a metagenomics approach such as QIMME, we can now gain insight into a wide range of microbial community analyses and be able to classify oral microbial taxa in the taxonomic hierarchy, i.e.,

from kingdom, phylum, class, order, family, genus, and species level. Characterising the taxonomic and functional characteristics of the human microbiome, despite considerable variation in methodology, is providing a better picture of our 'normal' microbial symbionts. As a result, analysing the oral microbiome at an early stage of chronic oral disorders like dental caries could aid in earlier detection and therapy. As there is an association between the oral microbiome and the occurrence of other diseases such as heart and liver problems, identifying and characterisation of the causal organisms could be useful in maintaining overall health.

In conclusion, scientific advancements have revealed new information regarding the nature of the oral microbiome in health and, to a lesser degree, illness. More study is needed to better understand the functional consequences of the oral microbiome in terms of illness diagnosis and risk assessment, as well as therapeutic techniques to restore the oral ecosystem's health.

Acknowledgements

We would like to take the opportunity to thank those involved in this project: the participants, dental officers and nurses from School of Dental Science, Universiti Sains Malaysia, Health Campus during the sample collection. Special thanks also to ScienceVision, Malaysia and Institute for Genome Malaysia (MGI) for their assistance in data analysis. We also acknowledge the financial support by Universiti Sains Malaysia (Bridging Grant: 304/PPSK/6316245).

Conflicts of interest

The authors have no conflicts of interest with the subject matter or materials discussed in this manuscript. The funders have no role in study design, data collections and analysis, decision to publish or preparation of the manuscripts.

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), pp. 5721–5732. doi: 10.1128/JCM.43.11.5721-5732.2005.
- Ahn, J., Yang, L., Paster, B. J., Ganly, I., Morris, L., Pei, Z., & Hayes, R. B. (2011) 'Oral microbiome profiles: 16s rRNA pyrosequencing and microarray assay comparison', *PLoS ONE*, 6(7), pp. 1–6. doi: 10.1371/journal.pone.0022788.
- Bennett, K. W. and Eley, A. (1993) 'Fusobacteria: New taxonomy and related diseases', *Journal of*

Medical Microbiology, 39(4), pp. 246–254. doi: 10.1099/00222615-39-4-246.

Bik, E. M., Long, C. D., Armitage, G. C., Loomer, P., Emerson, J., Mongodin, E. F., Nelson, K. E., Gill, S. R., Fraser-Liggett, C. M., & Relman, D. A. (2010) 'Bacterial diversity in the oral cavity of 10 healthy individuals', *ISME Journal*, 4(8), pp. 962–974. doi: 10.1038/ismej.2010.30.

Chakravorty, S., Helb, D., Burday, M., & Connell, N. (2007) 'A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria', *J Microbiol Methods*, 69(2), pp. 330–339. doi: 10.1016/j.mimet.2007.02.005.A.

Chaudhary, N., Sharma, A. K., Agarwal, P., Gupta, A., & Sharma, V. K. (2015) '16S classifier: A tool for fast and accurate taxonomic classification of 16S rRNA hypervariable regions in metagenomic datasets', *PLoS ONE*, 10(2), pp. 1–13. doi: 10.1371/journal.pone.0116106.

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), pp. 5002–5017. doi: 10.1128/JB.00542-10.

Graspeuntner, S., Loeper, N., Künzel, S., Baines, J. F., & Rupp, J. (2018) 'Selection of validated hypervariable regions is crucial in 16S-based microbiota studies of the female genital tract', *Scientific Reports*, (June), pp. 4–10. doi: 10.1038/s41598-018-27757-8.

Jiang, Q., Liu, J., Chen, L., Gan, N., & Yang, D. (2019) 'The oral microbiome in the elderly with dental caries and health', *Frontiers in Cellular and Infection Microbiology*, 9(JAN). doi: 10.3389/fcimb.2018.00442.

Peterson, S. N., Snesrud, E., Liu, J., Ong, A. C., Kilian, M., Schork, N. J., & Bretz, W. (2013) 'The Dental Plaque Microbiome in Health and Disease', *PLoS ONE*, 8(3). doi: 10.1371/journal.pone.0058487.

Rajilić-Stojanović, M. and de Vos, W. M. (2014) 'The first 1000 cultured species of the human gastrointestinal microbiota', *FEMS Microbiology Reviews*, 38(5), pp. 996–1047. doi: 10.1111/1574-6976.12075.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G. F., Chater, K. F., & van Sinderen, D. (2007) 'Genomics of Actinobacteria: Tracing the Evolutionary History of an Ancient Phylum', *Microbiology and Molecular Biology Reviews*, 71(3), pp. 495–548. doi: 10.1128/mmbr.00005-07.