

Influence of DNA extraction methods and 16S rRNA targeted hypervariable regions on microbial diversity profiling of dental plaque and saliva

Sree Sankar Darveekaran Nair^{a,b}, Pengfei Zhu^a, Yogendra Bhaskar^{a,b}, Fei Teng^{a,c,*}

^a Single-Cell Center, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, Shandong 266101 China

^b University of Chinese Academy of Science, Beijing 100049 China

^c Department of Stomatology, Qingdao Municipal Hospital, Qingdao, Shandong 266101 China

*Corresponding author, e-mail: tengfei@qibebt.ac.cn

Received 7 Jun 2020

Accepted 3 May 2021

ABSTRACT: The application of high-throughput sequencing has revolutionized our ability to investigate the microbiota composition of complex oral ecosystems. One such approach is through 16S rRNA sequencing. Deciphering this diverse microbial community using such approach is more accurate than traditional culture-based methods. Numerous factors such as primer selection, DNA extraction methods, and sequencing platforms may compromise the results. Therefore, these factors need to be considered carefully for studies such as microbial diversity analysis and microbial abundance determination. Here, we evaluated the effect of two DNA extraction methods (soil and swab DNA extraction kits) and targeted 16S rRNA hypervariable regions (V1–V3 and V3–V4) on plaque and saliva sample types to investigate the diverse microbes in oral samples of five Chinese citizens. Both of the extraction methods were successful in determining the relative abundance of microbes; however, relative abundance percentage of the species reported by the soil extraction method with the hypervariable region V1–V3 is higher in both sample types. The relative taxonomical abundance of microbes at OTU level of species represented by the V3–V4 region is lower; nonetheless, the number of unshared species identified is higher using this region. Both of the hypervariable regions identified unshared species in plaque and saliva sample types. Thus, choice of DNA extraction methods and hypervariable regions greatly influence the representation of oral microbial diversity in plaque and saliva. This study provides a guide to future studies to portray and quantify the human oral microbiota.

KEYWORDS: plaque, saliva, 16S regions, DNA extraction methods, microbiome, next generation sequencing

INTRODUCTION

The human body is enclosed within a plethora of self-originating microbes in multiple sites [1–4]. Among the vast diversity expressed in human body, modern scientists showed a keen interest in oral cavity, as it harbours unknown and unculturable microbiome with exceptional physiological and ecological characters [5, 6], plaque [7] and saliva [8] are the two most explored areas in oral cavity, as they are repositories of microbes including viruses, fungi, protozoa, bacteria, and archaea [9]. Oral diseases, such as periodontal diseases and dental caries, are mainly associated with oral microbial dysbiosis [10, 11].

Distinct niches, such as plaque and saliva, occur within the oral cavity, which harbours disease specific microbes [12, 13]. One of the successful and

frequent approaches to study the diversity observed in these niches is extraction of DNA through high throughput sequencing of 16S rRNA from microbial samples obtained from humans [14, 15] and environment [16, 17]. This technique is helpful in microbial diversity analysis; however, experimental biases occur in the characterization of microbiome during the sampling steps [18].

The choice of 16S regions and DNA extraction techniques is crucial during the analysis of microbiome samples [19–21]. Higher contamination in human metagenomic samples affect DNA extraction [22]. Moreover, microbial composition present in the samples are affected by the selection of sample sites from the oral cavity [23]. Since DNA extraction methods influence the identification of microbial profiles, a careful evaluation is needed to detect the biases in experimental procedures [24].

16S rRNA gene sequencing is used extensively in taxonomic studies [25]. The accurate assessment of a microbiota completely depends on the selected 16S region and primer sequences; however, errors can occur with the amplification of non-representative genomic region during taxonomic based studies [26]. Significant correlations between oral bacterial composition of the V1–V3 and V3–V4 regions, represented by β -diversity metrics, have been reported by past studies [27]. However, bacterial profiling remains a challenging factor for its inconsistent quantification and reproducible identification, leading to explore different 16S regions [28].

In this study, two DNA extraction methods (soil extraction and swab extraction) and two hypervariable regions (V1–V3 and V3–V4) were evaluated by exploring the oral microbial diversity present in plaque and saliva samples obtained from five Chinese citizens. The soil extraction method has been used to evaluate the environmental samples and also to study anaerobic microbes [29, 30]. Moreover, the swab extraction method was used on studies associated with gene polymorphisms in chronic periodontitis [31]. Here, we systematically evaluated two extraction methods and two hypervariable regions to probe the oral microbial diversity that could retrieve better taxonomic relative abundance of the microbes present in plaque and saliva samples.

MATERIALS AND METHODS

Sample collection

Three children of age 4–5 and two adults of age 29 and 28 from Qingdao, China were selected. A total of 3 saliva and 2 plaque samples were obtained individually from 3 children and 2 adults, respectively; and the sampling was repeated four times to get 4 replicates each of 5 samples. For 12 h before sampling, both groups of subjects were prohibited from any oral hygiene including tooth brushing, flossing, and mouth rinsing. Unstimulated saliva samples produced during 5 min were collected in sterile tubes from the children. The plaque samples from adults were collected using swabs. The swab heads with cohering plaque were removed using sterile scissors and washed several times in a 50 ml tube with 10 ml sterile distilled water. All samples were stored under the temperature of -20°C until DNA extraction experiment.

DNA extraction

Two DNA extraction methods using TIANamp soil DNA extraction kit and TIANamp swab DNA extraction kit (Tiangen biotech, Beijing) were chosen (Table 1). Final DNA concentrations were estimated by Qubit dsDNA HS assay kit (ThermoFisher, Carlsbad, USA). Detailed steps of the two DNA extraction methods followed manufacturer's procedures: soil extraction kit (<http://www.tiangen.com/asset/imsupload/up0718754001433125246.pdf>) and swab extraction kit (<http://www.tiangen.com/asset/imsupload/up0968755001348195137.pdf>).

TIANamp soil DNA extraction kit

Saliva samples (2 ml) or plaque samples (removed the swab heads) were put to each tube. SA buffer (750 μl) and 0.25 g of glass beads was added to two microcentrifuge tubes. SC buffer (60 μl) was added to the samples and mixed. After centrifuging for 1 min, the two supernatants were transferred to a new tube. Buffer HA (250 μl) was added and mixed, then incubated at 4°C for 5 min. After 1 min centrifugation, the supernatants from the two tubes were transferred to a new tube. Then incubated at 4°C for 5 min after adding 200 μl of Buffer HB, centrifuged again for 1 min and buffer GF of 1200 μl was added to the supernatants and mixed, then 700 μl solutions were transferred to a spin column and centrifuged at 12 000 rpm for 30 s and the flow-through was discarded. Buffer PWS (500 μl) was added and centrifuged at 12 000 rpm and the flow through was discarded. Then 500 μl of 70% ethanol was added to the spin column and centrifuged at 12 000 rpm for 30 s. The flow through was discarded and centrifuged at 12 000 rpm for 2 min. To completely dry the residual washing buffer in the columns, the spin columns were incubated at room temperature (15 – 25°C) for several minutes. After placing the spin column to a new microcentrifuge tube, 50–100 μl TE buffer was pipetted directly to the centre of the membrane and incubated at room temperature (15 – 25°C) for 2–5 min. Then centrifuged at 12 000 rpm for 2 min. and the flow-through was collected in a microcentrifuge tube.

TIANamp swab DNA extraction kit

Saliva samples of 2 ml or plaque samples (removed the swab heads) were put to each tube. GA buffer (400 μl) was dispensed into two microcentrifuge tubes. After that 20 μl Proteinase K was added and mixed, then incubated at 56°C for 60 min. For every

Table 1 Extraction methods, primers, sample subjects, and sample type used in this study.

Extraction method	16S regions	Sample subject	Sample type	Forward sequence	Reverse sequence
Swab-DNA	V1–V3	Adult Child	Plaque Saliva	TGGAGAGTTTGATCCTGGCTCAG	TACCGCGGCTGCTGGCAC
Swab-DNA	V3–V4	Adult Child	Plaque Saliva	GTACTCCTACGGGAGGAGCA	GTGGACTACHVGGGTWTCTAAT
Soil-DNA	V1–V3	Adult Child	Plaque Saliva	GGAGAGTTTGATCCTGGCTCAG	TACCGCGGCTGCTGGCAC
Soil-DNA	V3–V4	Adult Child	Plaque Saliva	CCTACGGRRBGCASCAGKVRVGAAT	GGACTACNVGGGTWTCTAATCC

15 min, the tubes were vortexed. GB Buffer (400 μ l) was added and mixed for 15 s and incubated at 70 °C for 60 min. To remove the drops inside the lid, the tubes were centrifuged. To extract the lysate, squeezing was done and the entire lysates were transferred to a clean tube. After the centrifugation, 200 μ l of 100% ethanol was added and mixed. The entire lysates from the two tubes were transferred to two spin column and centrifuged at 12 000 rpm for 30 s. GD buffer (500 μ l) was added after discarding the filtrates, centrifuged briefly at 12 000 rpm and discarded the filtrate. PW buffer (600 μ l) was added and the filtrates were discarded. Then repeated this step and centrifuged at 12 000 rpm for 2 min. After discarding the filtrate, the spin columns were incubated at room temperature (15–25 °C) for several minutes to dry the membrane. TB buffer (50 μ l) was pipetted to the membrane centre after placing the spin column in a microcentrifuge tube. Then, kept the tube at room temperature for 2–5 min and centrifuged at 12 000 rpm for 2 min.

Targeted 16S rRNA amplification

Amplicon libraries of two chosen hypervariable regions including V1–V3 and V3–V4 in 16S rRNA genes were prepared from individual extracted DNA samples, and barcode sequenced using the 2 \times 300 PE Illumina MiSeq platform [32] (Table 1).

Bioinformatics analysis

Qiime2 version 2020.6.0 was used for quality assessment of raw reads [33]. The minimum and the maximum screen lengths were identified using a random sampling of 10 000 out of 488 346 sequences without replacement, and after sub-sampling we found that the minimum base pair length is 240, and the maximum is 600. MAFFT aligner was used to perform multiple sequence alignment [34]. Taxa were

assigned for each sequence through oral “CORE” reference database [35] with a confidence threshold of 80%. Microbial community diversities, including α and β diversities, were analyzed at the OTU level. Shannon, Simpson, Chao1, and Pielou’s diversity indices, that measure richness and evenness, were employed to estimate diversity within each of the experimental protocols above. For β diversity, Unifrac unweighted measures were applied to assess the dissimilarity in microbial diversity found in sample subjects and sample types.

Statistical analysis

Wilcoxon test and False discovery rate (FDR) were employed for pairwise comparison and p -value adjustment, respectively. To visualize the differences between the extraction methods and 16S regions and whether they show an effect in differentiating between the sample types, boxplots were created based on unweighted unifrac along with PERMANOVA analysis and Wilcoxon test. To compare the quantitative data in α and β diversity analyses, the Kruskal-Wallis rank sum test was used with FDR adjustment for pairwise comparisons.

RESULTS

In this study, plaque and saliva samples were obtained from adults and children, respectively. We compared the 16S rRNA sequencing results from two DNA extraction procedures and two pairs of 16S rRNA hypervariable regions (Table 1; Fig. 1a). A total of 20 replicates, 3 saliva samples (12 replicates) and 2 plaque samples (8 replicates) were extracted with respect to the 16S regions and the extraction methods, which yielded inclusively 2 119 678 bacterial 16S rRNA gene sequences. After passing the stringent quality control, 33 167 non-chimeric reads per sample remained.

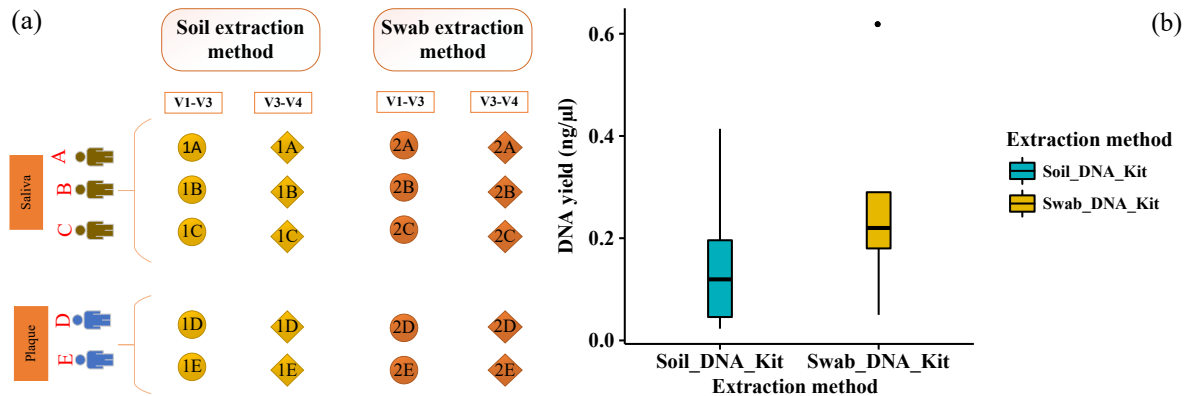


Fig. 1 Workflow of the study design and DNA yield processed by the two DNA extraction methods. (a) Representation of 3 saliva samples (12 replicates) and 2 plaque samples (8 replicates), repeated 4 times to get a total of 20 replicates that were sequenced by two DNA extraction methods and followed by amplification of two 16S rRNA gene target regions. 1A, 1B, 1C, 1D, and 1E represent the replicates used by soil extraction method; and 2A, 2B, 2C, 2D, and 2E represent the replicates used by swab extraction method. (b) Both DNA extraction methods produced high-quality metagenomics DNA from the oral samples. Swab DNA extraction kit yielded the higher amount of metagenomic DNA ($p < 0.05$, Wilcoxon signed rank test).

DNA concentrations significantly differ between two extraction methods

The presence of metagenomic DNA was confirmed by the application of gel electrophoresis. The two DNA extraction methods produced high-quality metagenomic DNA from both sample types. Pairwise comparison using Wilcoxon test showed significant difference between the two DNA extraction methods ($p < 0.05$, Fig. 1b). However, there was no significant difference in terms of DNA concentration from amplification of the two variable regions ($p > 0.05$).

Soil extraction method and ‘V1–V3’ 16S hypervariable region significantly represent the microbial diversity

To identify the microbiota features that are associated with the extraction method and hypervariable regions, PERMANOVA analysis was conducted in unweighted unifrac distance matrix (Fig. 2a). The extraction methods and 16S regions exhibited strong effects in oral microbiota structures (PERMANOVA $p < 0.05$, $F = 2.49$, $F = 1.86$) (Fig. 2a, Table S2), but no significant effect was seen in sample subjects and sample types. PCoA plot was created through Bray-Curtis distance to show the variation among DNA extraction methods and 16S regions; and the results revealed that the two extraction methods and the 16S regions had significant variation among each other ($p < 0.01$, Wilcoxon signed rank test) (Fig. S1), whereas both

extraction methods and 16S regions showed no variation to distinguish the microbial structure in different sample types ($p = 0.8$, Wilcoxon signed rank test) (Fig. S1). Moreover, pairwise comparison between the two extraction methods and the 16S regions showed significant variance (Fig. 2b, Table S2). Lower β -diversities were found in soil extraction method and V1–V3 regions, suggesting higher reproducibility of oral microbiota structure, which confirms that the soil extraction method along with V1–V3 hypervariable region is more suitable for determining the best oral microbiome structure.

The effect of both extraction methods and 16S regions based on α -diversity was assessed by Shannon, Simpson, Chao1, and Pielou’s evenness indices. Significant variance was observed between the two extraction methods ($p < 0.05$, Kruskal-Wallis test; Fig. 3), whereas 16S regions revealed no significant difference in Simpson ($p = 0.7$, Kruskal-Wallis test; Fig. 3), Chao1 ($p = 0.4$, Kruskal-Wallis test; Fig. 3), Shannon ($p = 0.4$, Kruskal-Wallis test; Fig. 3), and Pielou’s evenness indices ($p = 0.5$, Kruskal-Wallis test). This confirms that the two extraction methods have significant effect in determining species evenness, whereas the 16S regions have none.

β -diversity and α -diversity analyses confirmed significant representation of microbial structure and evenness in plaque and saliva by soil extraction method with V1–V3 region. Permanova confirmed the effect sizes of both extraction methods and

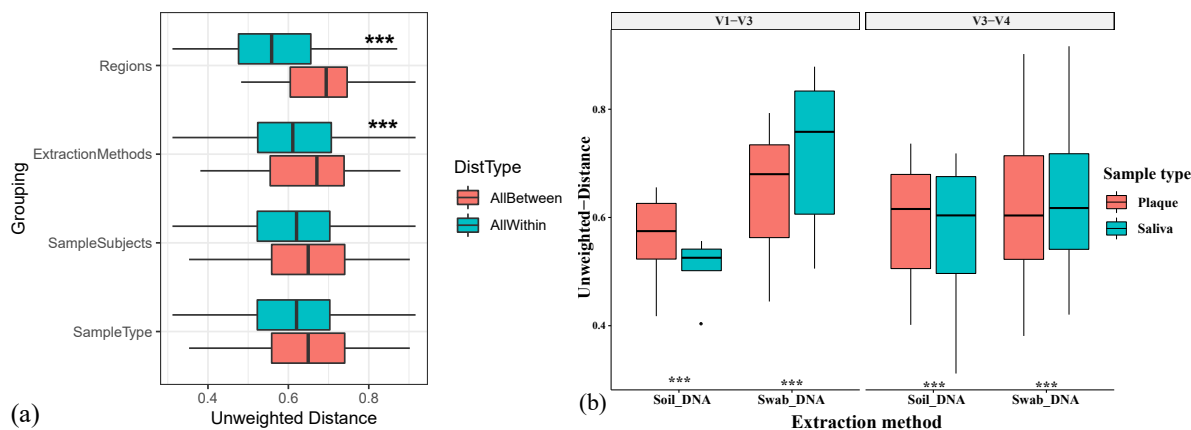


Fig. 2 The major factors that affect the structure of oral bacteria. (a) The oral microbiome structure within and between DNA extraction methods, hypervariable regions, sample types, and sample subjects. The extraction methods ($p < 0.001$, $F = 16.45$, PERMANOVA) and 16S regions ($F = 7.32$, PERMANOVA) exhibited strong effects on the oral microbial structures ($p < 0.01$, PERMANOVA), whereas sample types ($p = 0.29$, $F = 1.05$, PERMANOVA) and sample subjects ($p = 0.30$, $F = 1.05$, PERMANOVA) showed no significant effects. (b) DNA extraction methods and regions showed significant variance ($p < 0.01$, Wilcoxon signed rank test). Soil DNA extraction method and V1-V3 region produced lower β -diversity, indicating higher reproducibility of oral microbiota structure (all $p < 0.01$, Wilcoxon signed rank test).

16S regions. Lower β -diversity indicated the reproducibility of the experimental procedures, and species evenness is confirmed by Shannon, Simpson, and Chao1 indices.

Soil extraction method significantly identifies abundant microbial profiles

Taxonomic compositions in plaque and saliva samples were analyzed to determine the efficiency of the experimental procedures for the identification of different species at OTU level. Microbes with relative abundance less than 0.0001 were discarded as a stringent threshold to screen the less abundant taxonomic profiles. Total number of reported phyla (Fig. 4a), genera (Fig. 4b), and species (Fig. 5a) were 10, 65, and 118, respectively. Pairwise comparisons between the two extraction methods and the 16S hypervariable regions, illustrated by a heat map, were to explore the variation in taxonomical relative abundances of species. Both, extraction methods and hypervariable regions were able to identify similar species in plaque and saliva; however, relative taxonomical abundance reported by the extraction methods and 16S hypervariable regions showed significant difference between each other (Kruskal-Wallis, $p < 0.001$; Fig. 5a).

The species identified by soil DNA extraction method with the hypervariable region V1-V3 were *L. mirabilis* (17.44%) and *bacterium_HMT_348*

(17.47%) (Fig. 5a), and with V3-V4 were *L. mirabilis* (13.3%) and *L. mirabilis* (4.22%) (Fig. 5a). The highly abundant species reported by swab DNA extraction method with V1-V3 hypervariable region in plaque and saliva were *A. HMT_473* (21%) and *S. bacterium_HMT_348* (10.67%) (Fig. 5a), and with V3-V4 were *A. HMT_473* (7.70%) and *P. pasteri* (9.41%) (Fig. 5a).

The overall species abundance represented by soil DNA extraction method with the hypervariable region V1-V3 was higher, and the reported taxonomical profiles in plaque and saliva samples were consistent and abundant. Therefore, soil DNA extraction method together with the hypervariable region V1-V3 is capable of identifying high abundant as well as low abundant taxa from plaque and saliva.

16S hypervariable region 'V1-V3' is superior to 'V3-V4' in probing the microbial diversity

Soil DNA extraction method produced the maximum taxonomical relative abundance in plaque and saliva sample types. To detect the unique (unshared) and common (shared) species in both sample types, hypervariable regions V1-V3 and V3-V4 were further investigated. We categorized the unique (unshared) and common (shared) species in such a way that, out of the 20 samples if a

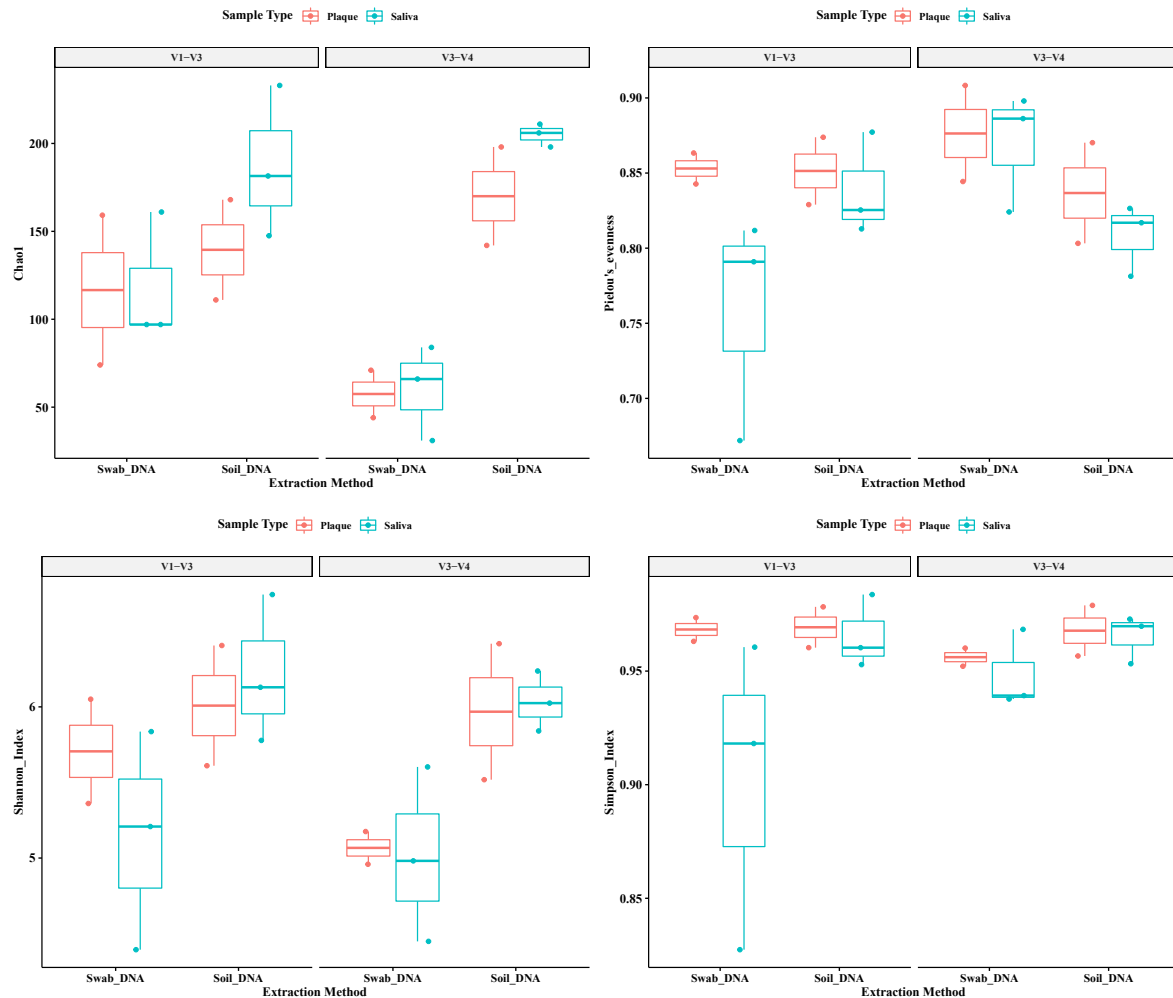


Fig. 3 The effect of species richness and evenness observed between extraction methods and 16S regions. Box plots represent α -diversity based on variance in species evenness. Significant variance was observed between the two extraction methods ($p < 0.01$, Kruskal-Wallis test), whereas no significant difference between the two hypervariable regions was found in all the four indices of Simpson ($p = 0.7$, Kruskal-Wallis test), Shannon ($p = 0.4$, Kruskal-Wallis test), Chao1 ($p = 0.4$, Kruskal-Wallis test), and Pielou's ($p = 0.5$, Kruskal-Wallis test).

species is present only in one sample type, it is considered as a unique (unshared) species; and if a species is present in both sample types, then it is considered as a common (shared) species. Pair-wise comparison between the hypervariable regions was conducted to assess the variance seen in the abundance showed by the two regions. Both hypervariable regions showed similar bacterial profiles; however, there was a significant difference in the relative abundance ($p < 0.05$, Wilcoxon test). Both regions identified a total of 115 species, in which 99 common (shared) species were found in plaque and saliva. V1-V3 region detected 28 unique (unshared) species in plaque, 8 unique (unshared) species in

saliva, and 45 common (shared) species (Fig. S2). Similarly, V3-V4 region detected 30 unique (unshared) species from saliva, 13 unique (unshared) species from plaque, and 54 common (shared) species (Fig. S2). The identified species using V1-V3 region were more taxonomically abundant, and the V3-V4 region reported higher number of unique species. For instance, the highest abundant unique species represented by both regions in plaque sample was *L. mirabilis*. But the relative abundance of this species varied in V1-V3 and V3-V4 regions were 17% and 13%, respectively (Fig. 5b). The highest abundant unique species identified by the V1-V3 and V3-V4 regions in plaque were *O. HMT_894*

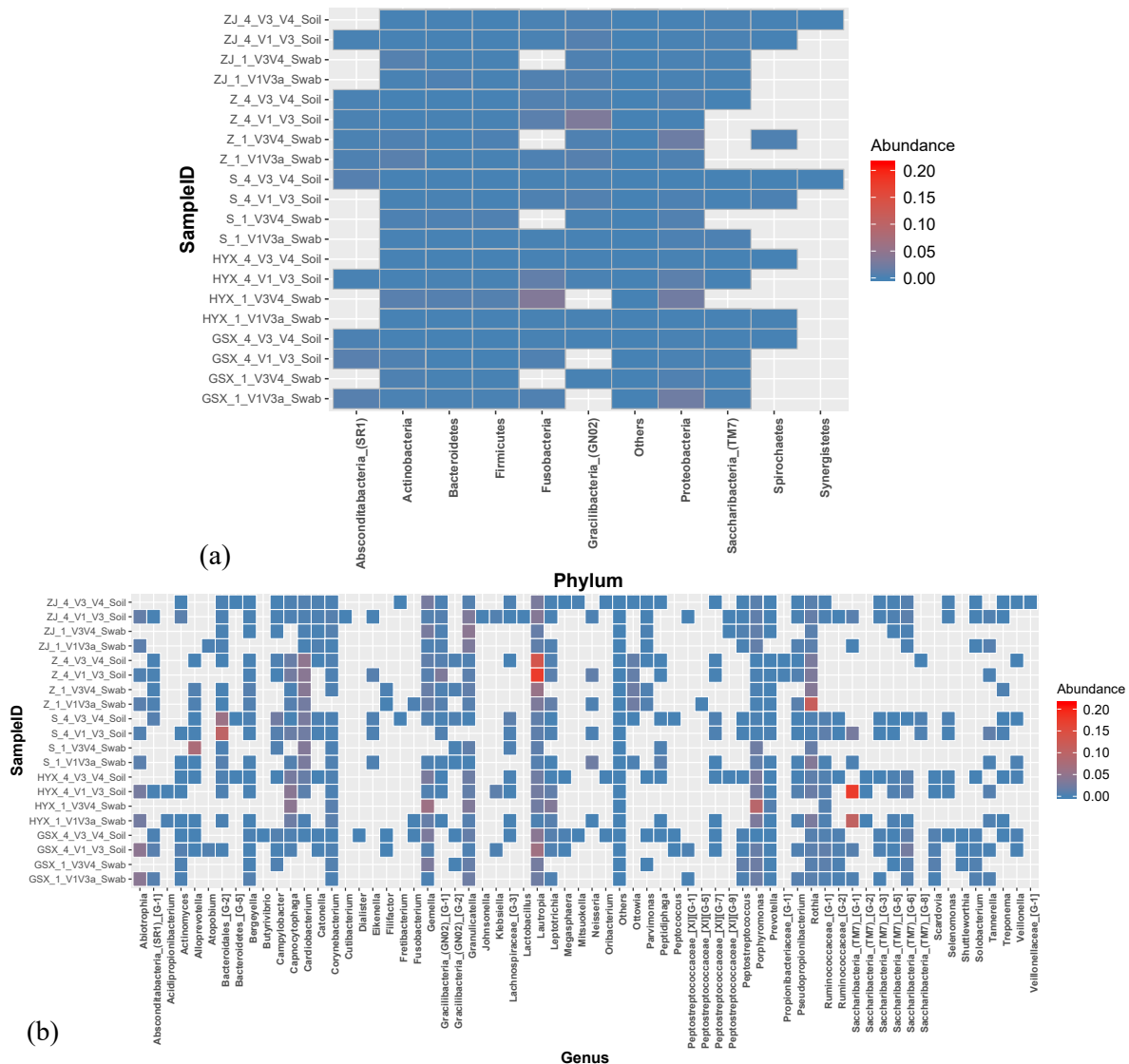


Fig. 4 Heat map showing relative abundance of 10 phyla and 65 genera. Extraction methods and hypervariable regions showed significant difference representing the microbial compositions at phylum and genus level. Soil extraction method with V1–V3 region produced the highest relative abundance represented by all taxa ($p < 0.05$, Wilcoxon signed rank test): (a) phyla detected from plaque and saliva samples; (b) genera detected from plaque and saliva samples.

(1.43%) and *P. nigrescens* (1.11%), respectively; and in saliva were *G. elegans* (1.4%) and *L. HMT_221* (0.69%), respectively (Fig. 5b). This depicts that species abundance identified by the hypervariable region V1–V3 is superior to the V3–V4 region. Therefore, hypervariable region V1–V3 is considered a suitable choice for identifying abundant oral microbial profile in plaque and saliva.

DISCUSSION

An accurate description of oral microbial community is a necessity to get a detailed idea about the

influence of oral microbiome in health and disease. Many characteristics of microbes, such as the structure of cell wall and physiological state of cells, can affect the experimental procedures conducted to extract DNA and cause biased results, specifically in phylogenetic studies of microbes [36]. To the best of our knowledge, these DNA extraction methods have never been used for oral microbial diversity studies. Here, in order to characterize the possible bias-generating steps in the experimental procedures and to disparagingly assess data quality, we examined potential sources of variation from DNA extraction

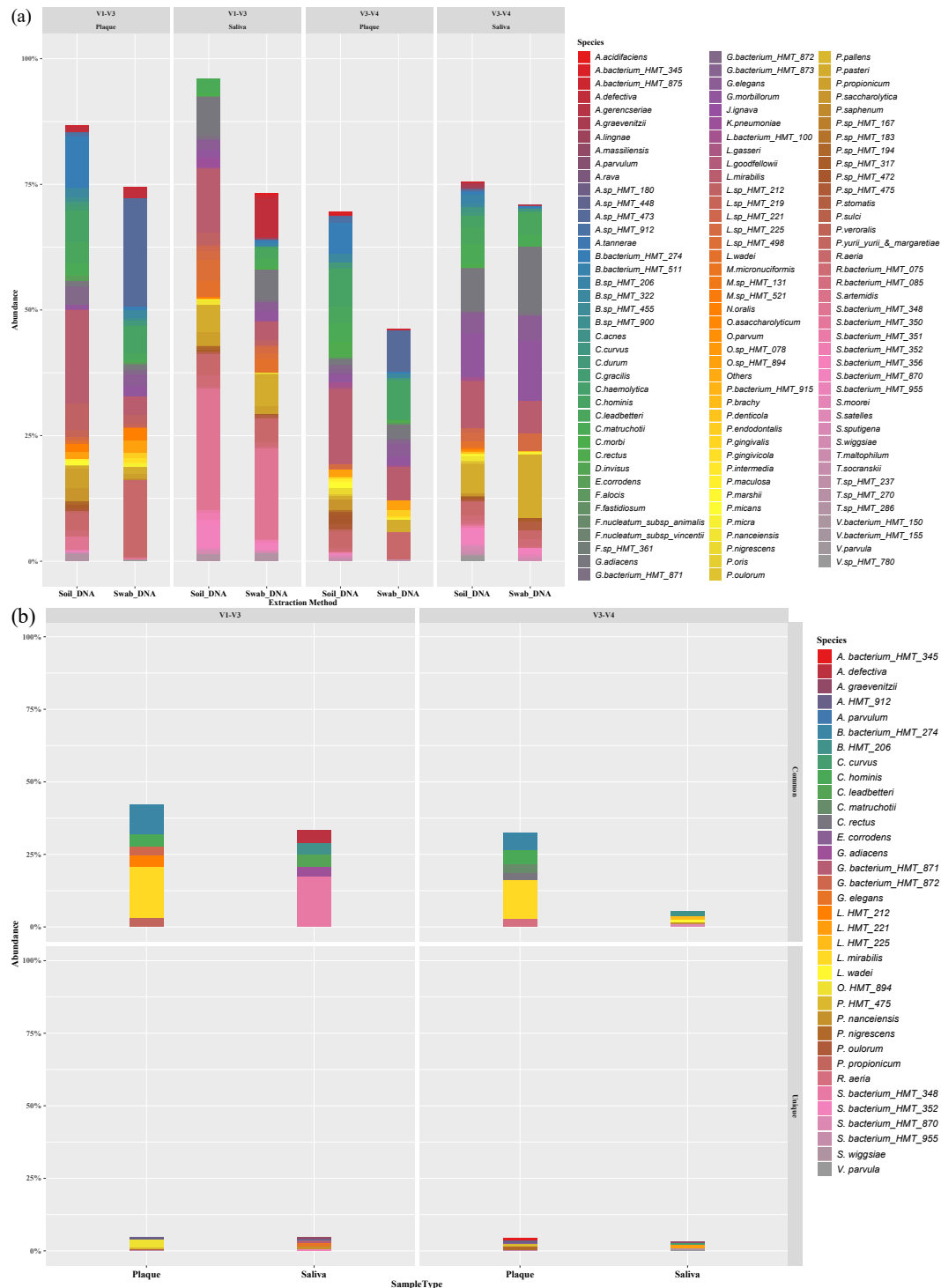


Fig. 5 Microbial relative abundance represented at species level by extraction methods. (a) Extraction methods and hypervariable regions showed significant difference representing the microbial diversities at species level. Each colour key represents a taxon obtained from the metagenomic samples. Soil extraction method with V1–V3 region produced the highest relative abundance represented by all taxa ($p < 0.05$, Kruskal-Wallis test). (b) Soil extraction method was tested separately with the two hypervariable regions, and top 35 bacterial profiles were generated. Both V1–V3 and V3–V4 hypervariable regions detected unique taxa from both sample types. However, V1–V3 region with soil extraction method reported the highest relative abundances of common bacterial species, and V3–V4 reported more unique species than V1–V3 (Wilcoxon test, $p < 0.05$).

methods and targeted 16S rRNA hypervariable regions to determine the nature and magnitude of their impacts on oral microbiome diversity profiling in plaque and saliva.

A primary concern with recent experimental procedures in microbiota profiling was how to quantitatively assess the sources of variation. Our results implicated that experimental procedures contributed up to 98% variation of the microbiota structures. DNA extraction method was the main source of variation ($F = 16.45$), which was contradicting by the hypervariable region ($F = 7.32$) results. This result was in agreement with previously reported studies, i.e., the final DNA yield and sequencing results were greatly influenced by the DNA extraction procedures [37]. In our study, we found that both extraction methods offered acceptable, yet varied results, in the extraction of metagenomic DNA from plaque and saliva samples. Notably, our results were in agreement with previous study that detected phyla, genera and species from both plaque and saliva [38]. This implicates the crucial role of DNA extraction methods in defining the variation of microbiota structures in different niches.

In previous studies, various targeted 16S hypervariable regions exhibited different degrees of variability and attributed that no individual hypervariable region by itself can differentiate all known microbial progenies [19]. Therefore, the choice of hypervariable regions affects the estimation of microbial diversity during microbiome sequencing. Moreover, considerable variations among targeted 16S rRNA gene regions have been documented by previous studies [20], yet no agreement with the 16S hypervariable regions was established. As previously reported [39] and confirmed in this study, selection of hypervariable region had a significant impact on the microbial structure, while its impact was relatively minor as compared with the DNA extraction methods. Although our experimental results exhibited ambiguity in the choice of hypervariable regions, we found that V1–V3 region seemed to produce more reproducible results than V3–V4, which is consistent with studies on taxonomical profiling of bacteria [40]. However, speculation of these conclusions to different oral bacteria studies will require further assessment in experimental design and sample collection techniques, as while 16S regions exhibited different degrees of variability among species [39]. Moreover, selection of hypervariable regions produced no obvious variations on the species evenness, as represented by α -diversity index value, though showed varying

effect in β -diversity. The selection of a hypervariable region, which is suitable and optimal, is of great importance, specifically when the aim is to track differences across sampling sites, time scales, or to compare results obtained from different laboratories.

On a similar context, Lerma et al [41] identified differences in bacterial community structure in soil and saliva samples that focused only in different 16S-rRNA regions, such as V1–V3, V3–V4, V4–V5, and V6–V8. Variations in microbial structure produced by different 16S regions were different in saliva and soil samples. In soil samples, V1–V3, V3–V4, and V6–V8 regions similarly detected the most abundant phyla, but V4–V5 differed. In addition, V4–V5 domain was biased towards the detection of various abundant phyla. In the current study, we mainly focused on the identification of different bacterial community in plaque and saliva samples using both 16S-rRNA regions (V1–V3 and V3–V4) and DNA extraction methods (soil and swab extraction kits). We found through PERMANOVA analysis that DNA extraction methods showed higher correlation factor and effect size in representing microbial structure, whereas 16S-rRNA regions had a minimal effect in the representation of microbial structure in both saliva and plaque samples. Our results are in agreement with the study conducted by Lerma et al where the results showed that 16S regions had no effect in saliva samples. In our study, we found that V1–V3 region produced more reproducible results than V3–V4, which was again in accordance with the study by Lerma et al showing that V1–V3 region detected the highest number of genera in soil and saliva samples, and it was the only one that detected the template DNA from all members in mock community.

As our sample selection was only limited to Chinese citizens, assessment of our study in determining the optimal choice of extraction method and the hypervariable region is confined within the evaluated samples. Necessary precautions should be maintained if these methods are to be introduced into other oral microbiome-based studies.

Appendix A. Supplementary data

Supplementary data associated with this article can be found at <http://dx.doi.org/10.2306/scienceasia1513-1874.2021.073>.

Acknowledgements: We thank all the authors for their contributions in this study. This work is funded by grant 81430011, 81670979 and 31600099 from National Natu-

ral Science Foundation of China, grant 16-5-1-67-jch and 16-5-1-64-jch from Qingdao Natural Science Foundation.

REFERENCES

- Aas JA, Paster BJ, Stokes LN, Olsen I, Dewhirst FE (2005) Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol* **43**, 5721–5732.
- Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, et al (2005) Diversity of the human intestinal microbial flora. *Science* **308**, 1635–1638.
- Gao Z, Tseng C-h, Pei Z, Blaser MJ (2007) Molecular analysis of human forearm superficial skin bacterial biota. *Proc Natl Acad Sci USA* **104**, 2927–2932.
- Bik EM, Eckburg PB, Gill SR, Nelson KE, Purdom EA, Francois F, Perez-Perez G, Blaser MJ, et al (2006) Molecular analysis of the bacterial microbiota in the human stomach. *Proc Natl Acad Sci USA* **103**, 732–737.
- Adler CJ, Browne GV, Sukumar S, Hughes T (2017) Evolution of the oral microbiome and dental caries. *Curr Oral Health Rep* **4**, 264–269.
- Jia G, Zhi A, Lai P, Wang G, Xia Y, Xiong Z, Zhang H, Che N, et al (2018) The oral microbiota: A mechanistic role for systemic diseases. *Br Dent J* **124**, 447–455.
- Coit P, Mumcu G, Ture-Ozdemir F, Unal AU, Alpar U, Bostanci N, Ergun T, Direskeneli H, et al (2016) Sequencing of 16S rRNA reveals a distinct salivary microbiome signature in Behcet's disease. *Clin Immunol* **169**, 28–35.
- Yang F, Zeng X, Ning K, Liu K-L, Lo C-C, Wang W, Chen J, Wang D, et al (2012) Saliva microbiomes distinguish caries-active from healthy human populations. *ISME J* **6**, 1–10.
- Jiang W, Zhang J, Chen H (2013) Pyrosequencing analysis of oral microbiota in children with severe early childhood dental caries. *Curr Microbiol* **67**, 537–542.
- Huang S, Li R, Zeng X, He T, Zhao H, Chang A, Bo C, Chen J, et al (2014) Predictive modeling of gingivitis severity and susceptibility via oral microbiota. *ISME J* **8**, 1768–1780.
- Teng F, Yang F, Huang S, Bo C, Xu ZZ, Amir A, Knight R, Ling J, et al (2015) Prediction of early childhood caries via spatial-temporal variations of oral microbiota. *Cell Host Microbe* **18**, 296–306.
- Xu X, He J, Xue J, Wang Y, Li K, Zhang K, Guo Q, Liu X, et al (2015) Oral cavity contains distinct niches with dynamic microbial communities. *Environ Microbiol* **17**, 699–710.
- Peterson SN, Snesrud E, Liu J, Ong AC, Kilian M, Schork NJ, Bretz W (2013) The dental plaque microbiome in health and disease. *PLoS One* **8**, e58487.
- Frémont M, Coomans D, Massart S, De Meirleir K (2013) High-throughput 16S rRNA gene sequencing reveals alterations of intestinal microbiota in myalgic encephalomyelitis/chronic fatigue syndrome patients. *Anaerobe* **22**, 50–56.
- Chen H, Jiang W (2014) Application of high-throughput sequencing in understanding human oral microbiome related with health and disease. *Front Microbiol* **5**, ID 508.
- Solomon S, Kachiprath B, Jayanath G, Sajeevan T, Singh IB, Philip R (2016) High-quality metagenomic DNA from marine sediment samples for genomic studies through a preprocessing approach. *3 Biotech* **6**, ID 160.
- Roesch LF, Fulthorpe RR, Riva A, Casella G, Hadwin AK, Kent AD, Daroub SH, Camargo FA, et al (2007) Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J* **1**, 283–290.
- Multinu F, Harrington SC, Chen J, Jeraldo PR, Johnson S, Chia N, Walther-Antonio MR (2018) Systematic bias introduced by genomic DNA template dilution in 16S rRNA gene-targeted microbiota profiling in human stool homogenates. *mSphere* **3**, e00560-17.
- Cruaud P, Vigneron A, Lucchetti-Miganeh C, Ciron PE, Godfroy A, Cambon-Bonavita M-A (2014) Influence of DNA extraction method, 16S rRNA targeted hypervariable regions, and sample origin on microbial diversity detected by 454 pyrosequencing in marine chemosynthetic ecosystems. *Appl Environ Microbiol* **80**, 4626–4639.
- Fouhy P, Clooney AG, Stanton C, Claesson MJ, Cotter PD (2016) 16S rRNA gene sequencing of mock microbial populations-impact of DNA extraction method, primer choice and sequencing platform. *BMC Microbiol* **16**, ID 123.
- Wanga X, Chena Y, Shanga Y, Wua X, Weia Q, Chenb J, Yana J, Zhanga H (2019) Comparison of the gut microbiome in red deer (*Cervus elaphus*) and fallow deer (*Dama dama*) by high-throughput sequencing of the V3–V4 region of the 16S rRNA gene. *ScienceAsia* **45**, 515–524.
- Velásquez-Mejía EP, de la Cuesta-Zuluaga J, Escobar JS (2018) Impact of DNA extraction, sample dilution, and reagent contamination on 16S rRNA gene sequencing of human feces. *Appl Microbiol Biotechnol* **102**, 403–411.
- Simón-Soro A, Belda-Ferre P, Cabrera-Rubio R, Alcaraz L, Mira A (2013) A tissue-dependent hypothesis of dental caries. *Caries Res* **47**, 591–600.
- Sohrabi M, Nair RG, Samaranyake LP, Zhang L, Zulfiker AHM, Ahmetagic A, Good D, Wei MQ (2016) The yield and quality of cellular and bacterial DNA extracts from human oral rinse samples are variably affected by the cell lysis methodology. *J Microbiol Methods* **122**, 64–72.
- Almonacid DE, Kraal L, Ossandon FJ, Budovskaya YV, Cardenas JP, Bik EM, Goddard AD, Richman J, et al (2017) 16S rRNA gene sequencing and healthy reference ranges for 28 clinically relevant microbial

- taxa from the human gut microbiome. *PLoS One* **12**, e0176555.
26. Wen C, Wu L, Qin Y, Van Nostrand JD, Ning D, Sun B, Xue K, Liu F, et al (2017) Evaluation of the reproducibility of amplicon sequencing with Illumina MiSeq platform. *PLoS One* **12**, e0176716.
 27. Zheng W, Tsompana M, Ruscitto A, Sharma A, Genco R, Sun Y, Buck MJ (2015) An accurate and efficient experimental approach for characterization of the complex oral microbiota. *Microbiome* **3**, ID 48.
 28. Thomas AM, Segata N (2019) Multiple levels of the unknown in microbiome research. *BMC Biol* **17**, 48.
 29. Dong B, Yi J, Dai L, Dai X (2013) Evaluation of several DNA extraction methods for obtaining total community DNA from anaerobic digestion sludge. *Procedia Environ* **18**, 856–863.
 30. Liu Y, Chang Q, Guo X, Yi X (2017) Bacterial diversity and community structure in lettuce soil are shifted by cultivation time. *AIP Conf Proc*, ID 020094.
 31. Cao X, He L, Meng H, Li P, Chen Z (2015) Relationship between vitamin D receptor gene polymorphisms and chronic periodontitis. *J Peking Univ (Health Sci)* **47**, 697–702.
 32. Yuan S, Cohen DB, Ravel J, Abdo Z, Forney LJ (2012) Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One* **7**, e33865.
 33. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, et al (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37**, 852–857.
 34. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772–780.
 35. Griffen AL, Beall CJ, Firestone ND, Gross EL, DiFranco JM, Hardman JH, Vriesendorp B, Faust RA, et al (2011) CORE: a phylogenetically-curated 16S rDNA database of the core oral microbiome. *PLoS One* **6**, e19051.
 36. Brooks JP (2016) Challenges for case-control studies with microbiome data. *Ann Epidemiol* **26**, 336–341.
 37. Wagner Mackenzie B, Waite DW, Taylor MW (2015) Evaluating variation in human gut microbiota profiles due to DNA extraction method and inter-subject differences. *Front Microbiol* **6**, ID 130.
 38. Anukam KC, Onwuzor I, Olise N, Duru M, Agbakoba N (2018) Oral bacteriome compositions identified by 16S rRNA metagenomics in a randomly selected “healthy” nigerian male and female subjects. *IJS-RDMS* **1**, 1–11.
 39. Castelino M, Eyre S, Moat J, Fox G, Martin P, Ho P, Upton M, Barton A (2017) Optimisation of methods for bacterial skin microbiome investigation: primer selection and comparison of the 454 versus MiSeq platform. *BMC Microbiol* **17**, ID 23.
 40. Kim M, Morrison M, Yu Z (2011) Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *J Microbiol Methods* **84**, 81–87.
 41. Soriano-Lerma A, Pérez-Carrasco V, Sánchez-Marañón M, Ortiz-González M, Sánchez-Martín V, Gijón J, Navarro-Mari JM, García-Salcedo JA, et al (2020) Influence of 16S rRNA target region on the outcome of microbiome studies in soil and saliva samples. *Sci Rep* **10**, ID 13637.

Appendix A. Supplementary data

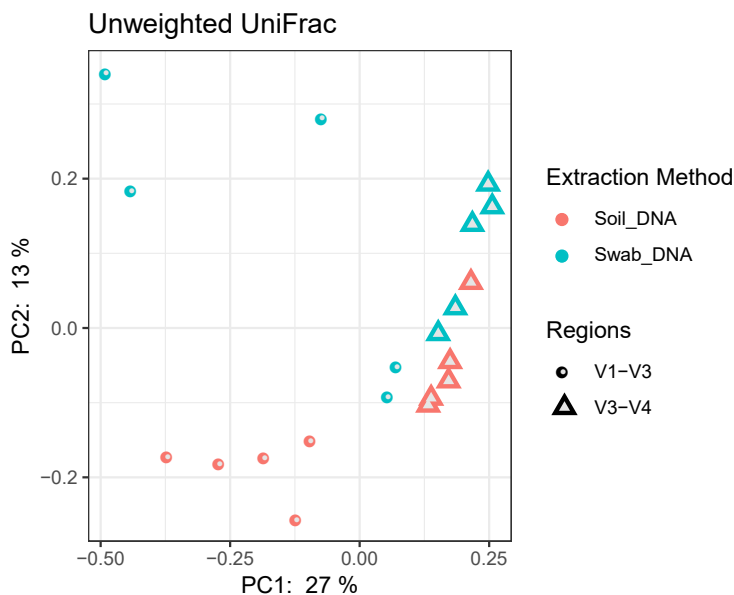


Fig. S1 PCoA plot using the UniFrac Unweighted distance matrix was created to show the variation among extraction methods and 16S hypervariable regions. DNA extraction methods and 16S hypervariable regions showed significant variance ($p < 0.01$, Wilcoxon signed rank test) among each other, whereas no variation was found between the extraction methods and hypervariable regions among different sample types ($p = 0.29$, Wilcoxon signed rank test) and sample subjects ($p = 0.30$, Wilcoxon signed rank test). PCoA showed that the distance between both extraction methods was higher and clustered to group separately, whereas the distance revealed by 16S regions were shorter and tended to group among themselves showing minimal difference between each other.

Table S1 Raw, filtered, denoised, merged, and non-chimeric reads.

Sample ID	Raw	Filtered	Denoised	Merged	Non-chimeric
GSX_4_V1_V3_Soil	223408	30722	30722	26265	23858
HYX_4_V1_V3_Soil	61349	13584	13584	12236	11224
S_4_V1_V3_Soil	64686	16245	16245	12649	11080
Z_4_V1_V3_Soil	70728	15953	15953	13513	12886
ZJ_4_V1_V3_Soil	68175	18025	18025	16427	15430
GSX_4_V3_V4_Soil	219596	154176	154176	149583	106425
HYX_4_V3_V4_Soil	245936	176116	176116	172443	127476
S_4_V3_V4_Soil	188207	134584	134584	124119	87811
Z_4_V3_V4_Soil	222335	158220	158220	154226	108398
ZJ_4_V3_V4_Soil	187699	125998	125998	121901	91452
GSX_1_V1V3a_Swab	56518	8355	8355	5440	4348
HYX_1_V1V3a_Swab	81837	35702	35702	28539	15735
S_1_V1V3a_Swab	80649	33869	33869	26077	13593
Z_1_V1V3a_Swab	41293	6450	6450	4231	3236
ZJ_1_V1V3a_Swab	42945	7062	7062	5083	4622
GSX_1_V3V4_Swab	55978	15903	15903	12703	7030
HYX_1_V3V4_Swab	48225	4939	4939	3455	2199
S_1_V3V4_Swab	61199	6705	6705	4598	2868
Z_1_V3V4_Swab	48345	15407	15407	11914	6944
ZJ_1_V3V4_Swab	50570	14234	14234	12023	6719

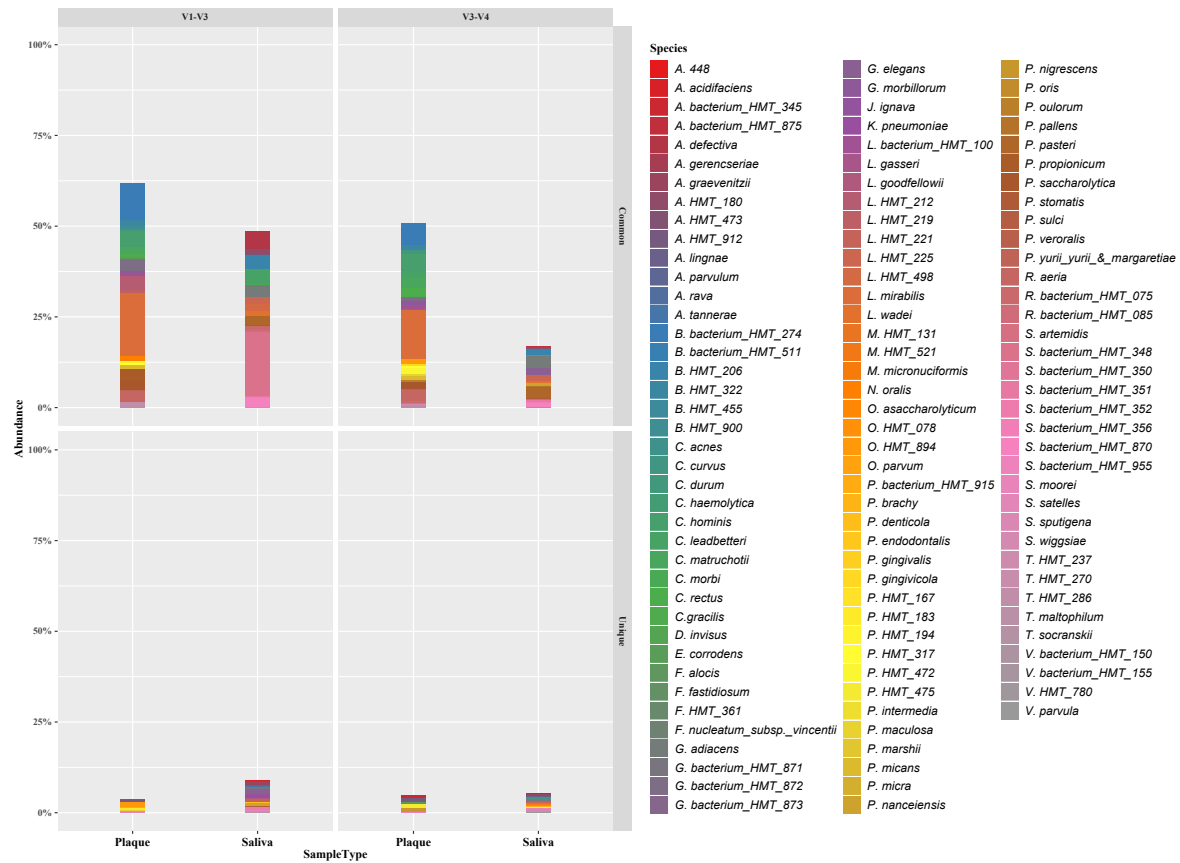


Fig. S2 Microbial relative abundance of all identified genera represented at species level by extraction methods. Soil extraction method was tested separately with the two hypervariable regions, and a total of 115 bacterial profiles were generated. V1–V3 region with soil extraction method reported the highest relative abundances of common bacterial species, while V3–V4 reported more unique species than V1–V3 (Wilcoxon test, $p < 0.05$).

Table S2 Permanova analyses for extraction methods, regions, sample types, and sample subjects.

PERMANOVA analysis	Df	SumOfSqs	R^2	F	Pr (>F)
Extraction methods	1	0.17570	0.0805	16.450	0.001***
Regions	1	0.08190	0.0375	7.327	0.011*
Sample type	1	0.01221	0.0056	1.057	0.290
Sample subjects	1	0.01221	0.0056	1.057	0.306