Reproducibility and repeatability of six high-throughput 16S rDNA sequencing protocols for microbiota profiling.

Authors:

Sajan C. Raju[1,2], sajan.raju@helsinki.fi

Sonja Lagström[3], sonja.lagstrom@kreftregisteret.no

Pekka Ellonen[3], pekka.ellonen@helsinki.fi

Willem M. de Vos [4,5], willem.devos@wur.nl

Johan G. Eriksson[1,6,7], johan.eriksson@helsinki.fi

Elisabete Weiderpass[1,2,8,9,10], Elisabete.Weiderpass@ki.se

Trine B. Rounge[*1,2,8], trine.rounge@kreftregisteret.no

Affiliations:

[1] Folkhälsan Research Center, Helsinki, Finland,

[2] Faculty of Medicine, University of Helsinki, Helsinki, Finland.

[3] Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland

[4] RPU Immunobiology, Department of Bacteriology and Immunology, University of Helsinki, Helsinki, Finland

[5] Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands

[6] Department of General Practice and Primary Health Care, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

[7] Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland

[8] Department of Research, Cancer Registry of Norway, Institute of Population-based Cancer Research, Oslo, Norway

[9] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

[10] Department of Community Medicine, Faculty of Health Sciences, University of Tromsø, The Arctic University of Norway, Tromsø, Norway.

Corresponding author:

Dr. Trine B. Rounge

Folkhälsan Research Center

Biomedicum 1

P.O. Box 63 (Haartmansgatan 8)

00014 University of Helsinki, Finland

E-mail address: trine.rounge@kreftregisteret.no

Phone: +47 99604304, Fax number: +358 2941 25382

Running title: Reproducibility of saliva microbiota

Abstract

Culture-independent molecular techniques and advances in next generation sequencing (NGS) technologies make large-scale epidemiological studies on microbiota feasible. A challenge using NGS is to obtain high reproducibility and repeatability, which is mostly attained through robust amplification. We aimed to assess the reproducibility of saliva microbiota by comparing triplicate samples. The microbiota was produced with simplified in-house 16S amplicon assays taking advantage of large number of barcodes. The assays included primers with Truseq (TS-tailed) or Nextera (NX-tailed) adapters and either with dual index or dual index plus a 6-nt internal index. All amplification protocols produced consistent microbial profiles for the same samples. Although, in our study, reproducibility was highest for the TS-tailed method. Five replicates of a single sample, prepared with the TS-tailed 1-step protocol without internal index sequenced on the HiSeq platform provided high alpha-diversity and low standard deviation (mean Shannon and Inverse Simpson diversity was 3.19 ± 0.097 and 13.56 ± 1.634 respectively). Large-scale profiling of microbiota can consistently be produced by all 16S amplicon assays. The TS-tailed-1S dual index protocol is preferred since it provides repeatable profiles on the HiSeq platform and are less labour intensive.

63 # Introduction

64 Presently, there is rising interest in studying human microbiota using high throughput approaches
65 based on 16S rRNA gene sequences. This gene is as a highly abundant, evolutionary conserved and
66 phylogenetically informative housekeeping genetic marker (Lane et al., 1985; Tringe and
67 Hugenholtz, 2008; Zheng et al., 2015). The composition and diversity of the human microbiota
68 have been correlated to health and disease, although only few cases of causal relationships have
69 been uncovered (Cho and Blaser, 2012; Human Microbiome Project Consortium, 2012; Nicholson
70 et al., 2012; Scheithauer et al., 2016; van Nood et al., 2013).
71 While attention has focused on the intestinal microbiota, it is well known that the oral cavity also
72 harbours a large microbial community that includes around 700 common bacterial species, out of
73 which 35% are still unculturable (Dewhirst et al., 2010). Cultivation-independent molecular
74 methods have validated these estimates, by identifying approximately 600 species or phylotypes
75 using 16S rRNA gene sequencing techniques (Dewhirst et al., 2010; Paster et al., 2001). Oral
76 bacteria have been linked to many oral diseases and non-oral diseases, testifying for their
77 importance (Krishnan et al., 2017). While metagenomic studies have provided insight in the large
78 coding capacity of the human microbiota (Li et al., 2014; Qin et al., 2010), taxonomic studies
79 mainly rely on amplifying and analysing hypervariable regions of 16S rRNA gene sequences.

80 It is known that a precise assessment of the microbiota depends heavily on the hypervariable
81 region selected, and primers used, whereas taxonomic resolution bias can arise with amplification
82 of non-representative genomic regions (Wen et al., 2017; Zheng et al., 2015). Next generation
83 sequencing (NGS) technology with application of barcode indexing are possible to achieve
84 thousands of sequences from large number of samples simultaneously (Andersson et al., 2008;
85 Hamady et al., 2008). However, reproducible identification and consistent quantification of
86 bacterial profiles remain challenging (Ding and Schloss, 2014). Studies have shown that β-diversity
87 metrics depicted significant correlation between oral bacterial composition for the V1–V3 and V3–

V4 regions (Zheng et al., 2015). The 16S rRNA V3-V4 hypervariable region is widely used for various microbiological studies (Belstrøm et al., 2016; Fadrosh et al., 2014; Janem et al., 2017). Protocols have been developed using the dual indexing strategy to yield the greater utilization of available sequencing capacity (Kozich et al., 2013). High throughput and cost effective sequencing approaches are continuous being developed, urging researchers to use the latest technologies while abandoning the old ones. However, evaluation of new methodologies is a crucial step in conducting rigorous scientific research (Sinclair et al., 2015). This specifically applies to generating representative libraries of 16S rRNA gene amplicons.

In this study, we aimed to simplify amplification procedure and investigate barcoding efficacy with internal indices, for sequencing 16S rRNA gene amplicons relative to sequencing quality, depth, reproducibility and repeatability. Specifically, we tested high-throughput workflows for amplicon library construction of the 16S rRNA V3–V4 hypervariable region using Truseq and Nextera adapters with dual index and dual index plus 6-nt internal index. We assessed the reproducibility of the saliva microbiota for four saliva samples in triplicates using the Illumina MiSeq platform and the repeatability using nine control samples, including five replicates from a single individual, with the 1-step TS-tailed dual index protocol on the Illumina HiSeq platform.

Materials and Methods

Saliva samples in triplicates from four volunteers were selected for this study (Fig. 1). The study was approved by the regional Ethics Committee of the Hospital District of Helsinki and Uusimaa (169/13/03/00/10). The saliva samples were collected in Oragene-DNA (OG-500) self-collection kits (DNA Genotek Inc, Canada) and mixed with stabilizing reagent within the collection tubes per manufacturer's instructions by participants, and stored at room temperature. A protocol with an intensive lysis step using a cocktail of lysozyme and mechanical disruption of microbial cells using bead-beating was employed. Fifty ml lysozyme (10 mg/ml, Sigma-Aldrich), 6 ml mutanolysin (25 KU/ml, Sigma-Aldrich), and 3 ml lysostaphin (4000 U/ml, Sigma-Aldrich) were added to a 500 ml

aliquot of cell suspension followed by incubation for 1 h at 37 ºC. Subsequently, 600 mg of 0.1-mm-diameter zirconia/silica beads (BioSpec, Bartlesville, OK) were added to the lysate and the microbial cells were mechanically disrupted using Mini-BeadBeater-96 (BioSpec, Bartlesville, OK) at 2100 rpm for 1 min (Yuan et al., 2012). After lysis, total DNA was extracted using cmg-1035 saliva kit, and Chemagic MSM1 nucleic acid extraction robot (PerkinElmer).

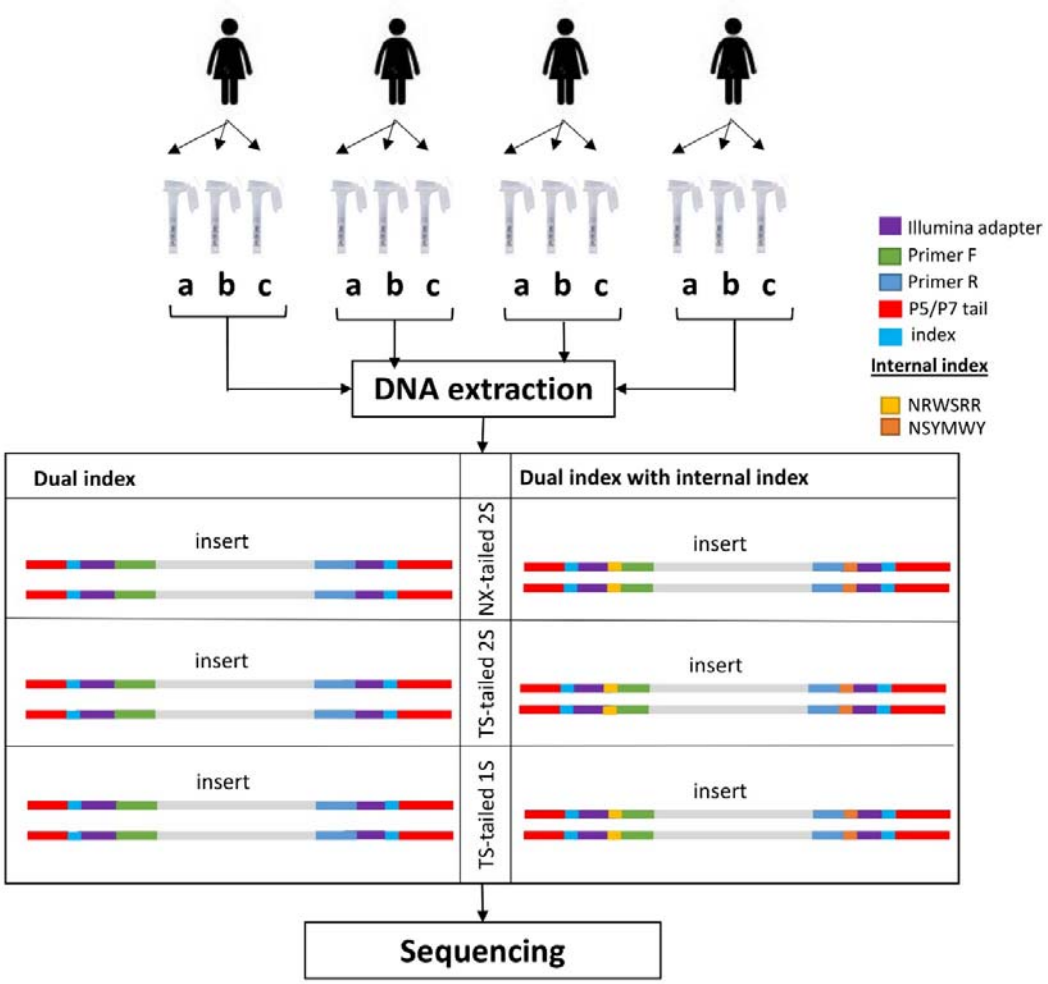


Fig. 1. Study Design

Schematic presentation of the study design showing the number of participants and different methods implemented.

PCR amplification

PCR amplification and sequencing libraries was prepared according to in-house 16S rRNA gene-based PCR amplification protocols. All protocols used 16S primers (S-D-Bact-0341-b-S-17: 5' CCTACGGGNGGCWGCAG '3 and S-D-Bact-0785-a-A-21: 5' GACTACHVGGGTATCTAATCC 3') targeting the V3-V4 region as reported previously (Klindworth et al., 2013). The 16S rRNA gene-based primers were modified by adding 5' tails corresponding to the Illumina Truseq and Nextera adapter sequences to the 5'-ends. Amplification was done using primers with and without incorporated internal index (Supplementary Table S1). Two sets of index primers carrying Illumina grafting P5/P7 sequence were used: in-house index primers with Truseq adapter sequence (Supplementary Table S2) and Illumina Nextera i5/i7 adapters. All oligonucleotides (except Illumina Nextera i5/i7 adapters) were synthesized by Sigma-Aldrich (St. Louis, MO, USA).

TS-tailed 1-step amplification

Amplification was performed in 20 µl containing 1 µl of template DNA, 10 µl of 2x Phusion High-Fidelity PCR Master Mix (Thermo Scientific Inc., Waltham, MA, USA), 0,25 µM of each 16S primer carrying Truseq adapter, 0.5 µM of each Truseq index primer. The cycling conditions were as follows: initial denaturation at 98 °C for 30 seconds; 27 cycles at 98 °C for 10 sec, at 62 °C for 30 sec and at 72 °C for 15 sec; final extension at 72 °C for 10 min, followed by a hold at 10 °C. Separate reactions were done using 16S rRNA gene-based primers with and without incorporated internal index (denoted as ii). Here after this protocol denoted as TS-tailed-1S.

TS-tailed 2-step amplification

Amplification was performed in 20 µl containing 1 µl of template DNA, 10 µl of 2x Phusion High-Fidelity PCR Master Mix (Thermo Scientific Inc., Waltham, MA, USA), 0,5 µM of each 16S primer carrying Truseq adapter. The cycling conditions were as follows: initial denaturation at 98 °C for 30 sec; 27 cycles at 98 °C for 10 sec, at 62 °C for 30 sec and at 72 °C for 15 sec; final

147 extension at 72 °C for 10 min, followed by a hold at 10 °C. Separate reactions were done using 16S
148 rRNA gene-based primers with and without incorporated internal index. Following PCR
149 amplification, samples were purified using a Performa V3 96-Well Short Plate (Edge BioSystems,
150 Gaithersburg, MD, USA) and QuickStep 2 SOPE Resin (Edge BioSystems, Gaithersburg, MD,
151 USA) according to the manufacturer's instructions. An additional PCR step was needed to add
152 index sequences to the PCR product. Amplification was performed in 20 µl containing 1 µl of
153 diluted (1:100) PCR product, 10 µl of 2x Phusion High-Fidelity PCR Master Mix (Thermo
154 Scientific Inc., Waltham, MA, USA), 0,5 µM of each Truseq index primer. The cycling conditions
155 were as follows: initial denaturation at 98 °C for 2 min; 12 cycles at 98 °C for 20 sec, at 65 °C for
156 30 sec and at 72 °C for 30 sec; final extension at 72 °C for 5 min, followed by a hold at 10 °C. Here
157 after this protocol denoted as TS-tailed-2S.

158 ***NX-tailed 2-step amplification***

159 Amplification was performed in 20 µl containing 1 µl of template DNA, 10 µl of 2x Phusion High-
160 Fidelity PCR Master Mix (Thermo Scientific Inc., Waltham, MA, USA), 0.5 µM of each of the 16S
161 rRNA gene-based primers carrying Nextera adapters. The cycling conditions were as follows: initial
162 denaturation at 98 °C for 30 sec; 27 cycles at 98 °C for 10 seconds, at 62 °C for 30 sec and at 72 °C
163 for 15 sec; final extension at 72 °C for 10 min, followed by a hold at 10 °C. Separate reactions were
164 done using 16S rRNA gene-based primers with and without incorporated internal index. Following
165 PCR amplification, samples were purified using a Performa V3 96-Well Short Plate (Edge
166 BioSystems, Gaithersburg, MD, USA) and QuickStep 2 SOPE Resin (Edge BioSystems,
167 Gaithersburg, MD, USA) according to the manufacturer's instructions. An additional PCR step was
168 needed to add index sequences to the PCR product. Amplification was performed according to
169 Illumina Nextera protocol to amplify tagmented DNA with following exceptions: i) reaction volume
170 was downscaled to 20 µl, ii) 1 µl of diluted (1:100) PCR product was used as template, and iii)

reaction mix was brought to the final volume with laboratory grade water. Here after this protocol denoted as NX-tailed-2S.

Pooling, purification and quantification

Following PCR amplifications, libraries were pooled in equal volumes. Library pool was purified twice with Agencourt® AMPure® XP (Beckman Coulter, Brea, CA, USA) according to the manufacturer's instructions using equal volumes of the Agencourt® AMPure® XP and the library pool. The purified library pool was analyzed on Agilent 2100 Bioanalyzer using Agilent High Sensitivity DNA Kit (Agilent Technologies Inc., Santa Clara, CA, USA) to quantify amplification performance and yield.

Sequencing

Sequencing of PCR amplicons was performed using the Illumina MiSeq instrument (Illumina, Inc., San Diego, CA, USA). Samples were sequenced as 251x 2 bp paired-end reads and two 8-bp index reads. DNA extracted from nine blank samples, two water samples and nine control saliva samples (in which 5 samples are replicates of sample 4c) using the above mentioned protocol and amplified with TS-tailed 1S protocol without internal index, and sequencing performed (271 x 2 bp) using the Illumina HiSeq instrument.

Phylogenetic Analysis.

Sequencing quality, index trimming and length filtering was carried out using Nesoni clip Version 0.130 (https://github.com/Victorian-Bioinformatics-Consortium/nesoni). Resulting sequences were processed using MiSeq_SOP in mothur (Version v.1.35.1) (Schloss et al., 2009) and sequences were aligned to ribosomal reference database arb-SILVA Version V119 (Quast et al., 2012). We used both SILVA database and the Human Oral Microbiome Database (HOMD) database for the alignment and classification of sequences but present here only the results from the SILVA database and taxonomy as it provides comprehensive, quality checked and regularly updated databases of

aligned small (16S / 18S, SSU) and large subunits (Quast et al., 2012). To obtain high quality data for analysis, sequence reads containing ambiguous bases, homopolymers > 8 bp, more than one mismatch in the primer sequence, or less than default quality score in mothur were removed. Assembled sequences with > 460 bp length and singletons were removed from the analysis. Chimeric sequences were also removed from the data set using the UCHIME algorithm within the mothur pipeline (Edgar et al., 2011). The high-quality sequence reads were aligned to the Silva 16S rRNA database (Version V119) and clustered into operational taxonomic units (OTUs) at a cut-off value > 98% sequence similarity. OTUs were classified using the Silva bacteria taxonomy reference. OTUs were calculated at distance 0.02 and alpha diversity (Shannon and inverse Simpson index) was calculated per sample. These diversity indexes are shown to be a robust estimation of microbial diversity (Haegeman et al., 2013).

Statistic procedures. Microbial diversity indices, both Shannon and Inverse Simpson, were used to summarize the diversity of a population. Simpson's index is more weighted on dominant species whereas Shannon index assumes all species are represented in a sample and that they are randomly sampled (Lozupone and Knight, 2008). Kruskal-Wallis (KW-test) test was performed on the alpha diversity indices to assess the statistical significance difference between microbial diversity and the methods used. Principal coordinate analysis (PCoA) plotted with bray-curtis distance without normalizing the data using biom formatted OTUs from mothur to the phyloseq R-package Ver 1.22.3 (McMurdie and Holmes, 2014). Intraclass correlation coefficients (ICC) to quantify the reproducibility, stability, and accuracy or neutrality of different protocol for six metrics included relative abundances of four major phyla (Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria) and two alpha diversity indices (Shannon & Inverse Simpson index). The ICCs were estimated using the SPSS (version 22) based on the mixed effects model (Sinha et al., 2016). All the graphics and plots were made in R using ggplot2 package (Wilkinson, 2011).

Results

220 ***16S rRNA sequencing***

221 Saliva microbiota sequence data of the 16S rRNA V3-V4 region for 4 individuals in triplicates
222 using TS-tailed and NX-tailed amplification, with and without internal index, were collected on the
223 Illumina MiSeq platform (Table 1**).**

224 **Table 1:** Sequencing statistics, quality check passed sequences and sequences classified for samples
225 (combined for methods used) sequenced in MiSeq and HiSeq platform.

#samples	Protocol	Total sequences	Reads/sample			QC passed sequences	#Sequences classified	% of sequences
			(Min)	(Max)	(Mean)			
12	NX-tailed-2S	56032	2273	6480	4669	52070	34332	61.27
12	NX-tailed-2S ii*	64791	2452	10192	5399	55407	24836	38.33
12	TS-tailed-2S	115736	6940	15161	9644	100788	69783	60.29
12	TS-tailed-2S ii	142880	8048	16271	11906	121134	66014	46.20
12	TS-tailed-1S	96252	733	10868	8021	79339	56761	58.97
12	TS-tailed-1S ii	123165	2187	14519	10263	89631	48479	39.36
9	TS-tailed-1S **	1228989	48100	398420	136554	974702	711088	57.86

226 * Internal indices, ** Control samples sequenced on the HiSeq platform
227

228 Two control water samples, nine saliva control samples (including 5 replicates) and blank samples
229 using TS-tailed 1S protocol without internal index, were sequenced on the Illumina HiSeq platform.
230 Samples sequenced using TS-tailed 1S and 2S protocol with and without internal index generated
231 comparatively higher amounts of sequence reads. This was true also after trimming of low-quality
232 sequences (Fig. 2 and Supplementary table S3**).** The sequences were clustered and assigned to 1086
233 OTUs. Sequence coverage and percentage of sequences passed quality check from each protocol
234 and qualified for taxonomic classification are summarized in Table 1.

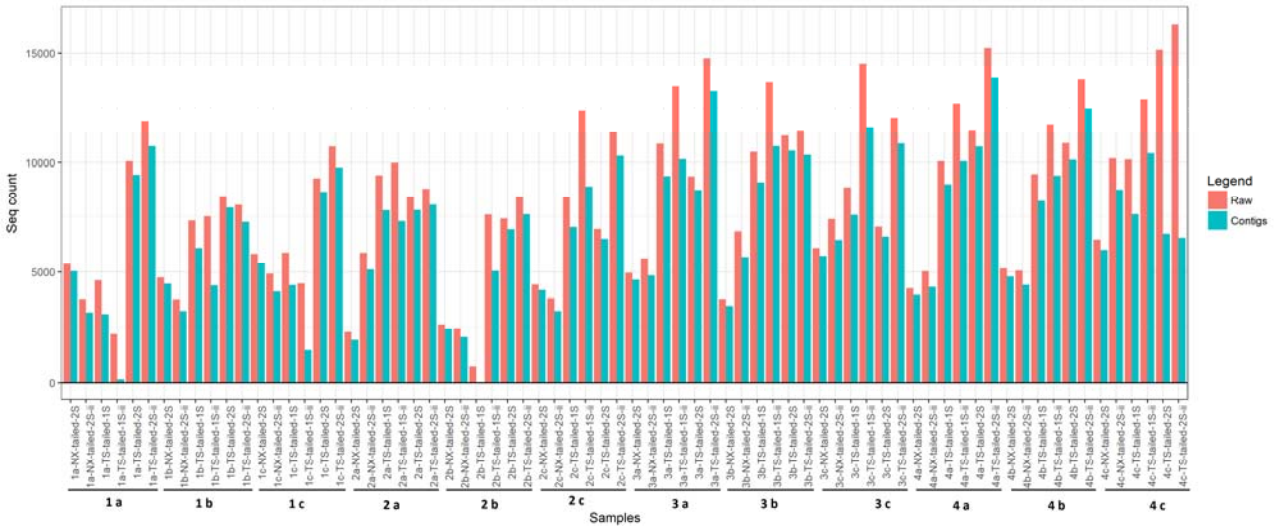


Fig. 2: Sequence data filtering.

Distribution of sequences before (Raw data count in blue colour) and after (contigs count in red colour) quality check and assembly for the saliva microbiota sequenced in MiSeq platform.

The protocols with the internal index approach showed consistently 14-23 % lower OTU's per sample for all protocol. About 61% of saliva microbiota sequences remained after quality check using NX-tailed protocol without internal index, while only 38% remained using NX-tailed protocol with internal index. In our study, the NX-tailed protocol produced slightly less sequences than the TS-tailed protocols, with 4669 and 5399 mean reads per sample respectively. About 60% of saliva microbiota passed the quality check in TS-tailed without internal index and produced in our protocol more than 8000 reads per sample. Principle coordinate analysis (PCoA) using Bray-Curtis distance, to visualize broad trends of how similar or different bacteria are between triplicate samples, shows that samples cluster according to the individual (Fig. 3).

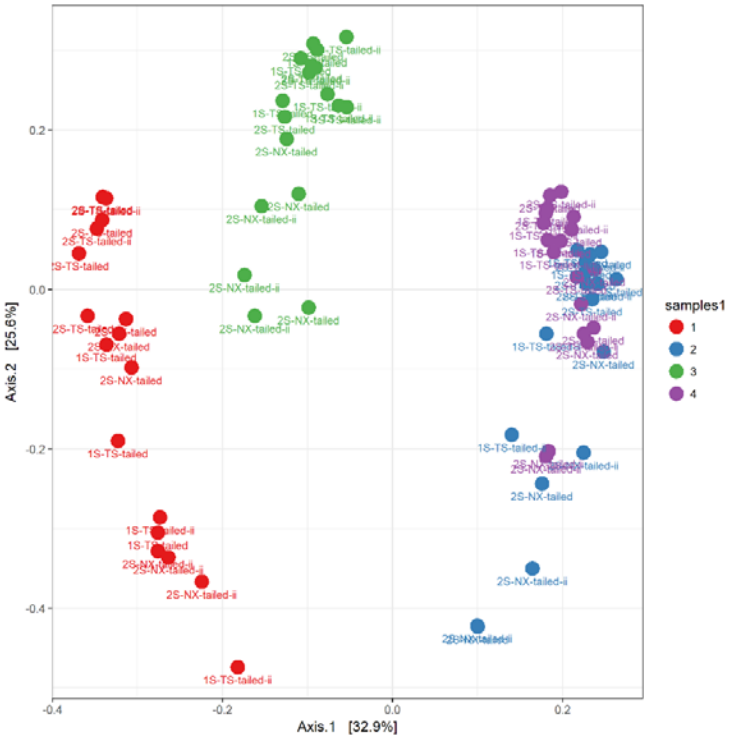
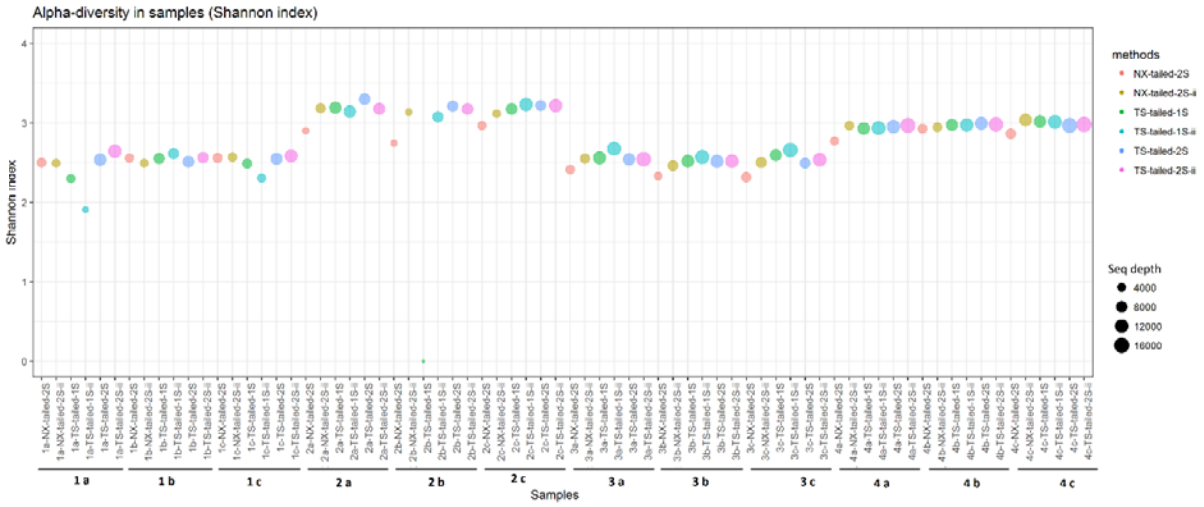


Fig. 3: Principal coordinate analysis (PCoA) plot

PCoA plot, based on the bray-curtis distances. The percentage of the total variance explained by each PC is indicated in parentheses.

Alpha diversity of saliva microbiota is similar for all the protocols

The Shannon diversity and inverse Simpson indices used to calculate the alpha diversity showed similar diversity for each sample irrespective of the protocols used with exception of few outliers (Fig. 4 a and b). The outliers are the samples with low diversity and low sequence depth, < 4000 sequences. Though Shannon diversity index showed less variation according to the sequence depth compared to inverse-Simpson index, we did not find any significant relationship (KW-test) between the diversity indices and the protocols used.



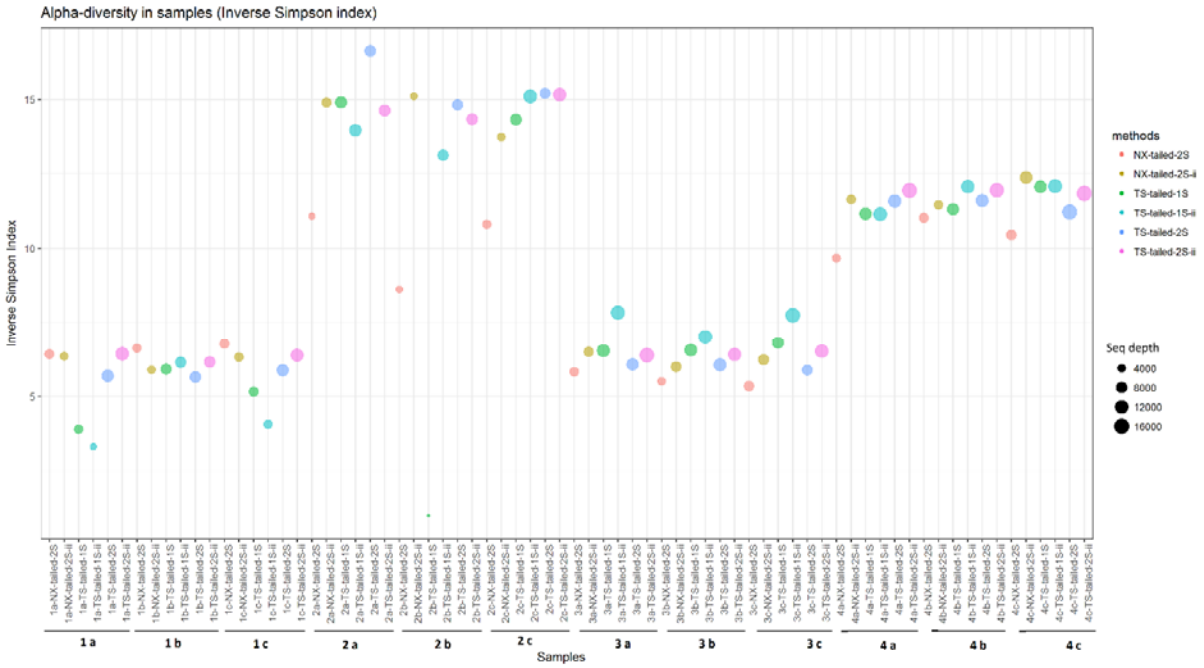


Fig. 4 a & b: Microbial diversity in saliva

Alpha diversity measured using Shannon (a) and Inverse Simpson (b) index in each replicates. Bubble size depicts the sequence depth and bubble colour is Illumina method used.

Consistent occurrence of bacterial abundance within the protocols

Taxonomic composition of saliva microbiota from four samples with different amplification protocols with and without internal index showed sample specific composition profile at two

taxonomic levels. The bacterial relative abundance at phylum level was measured using the top five abundant phyla; Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria and Proteobacteria (Fig. 5).

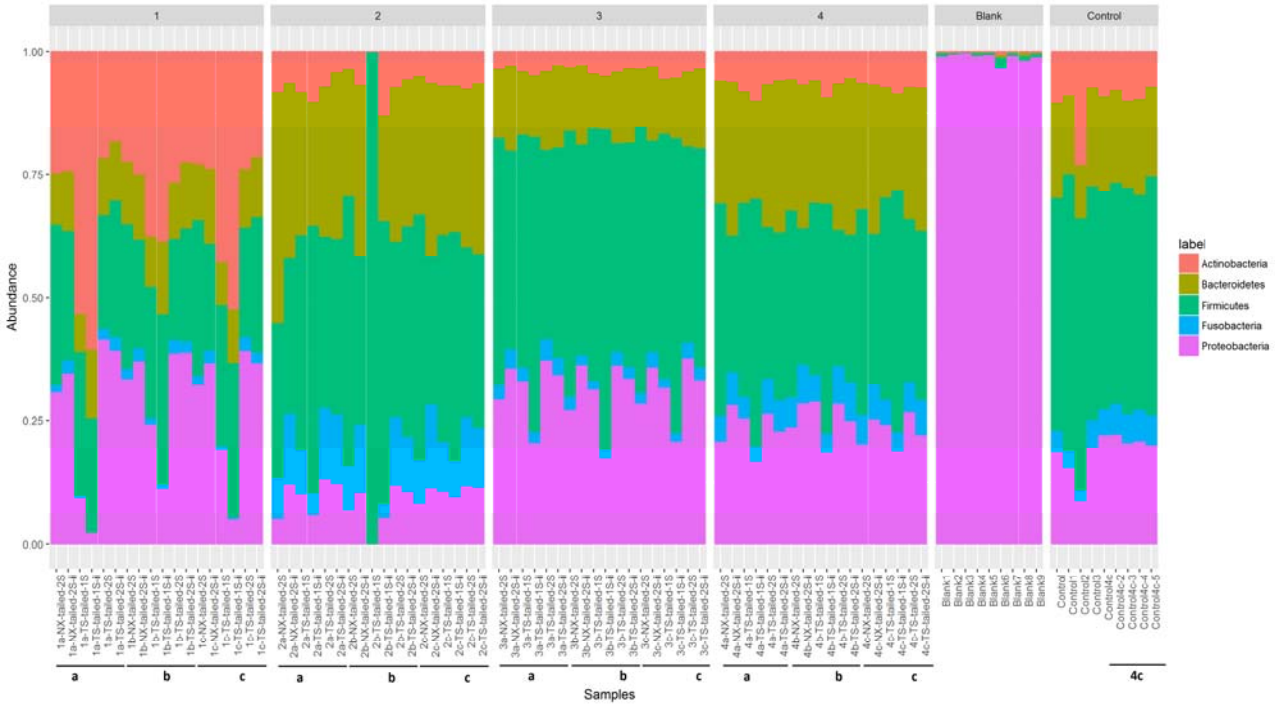


Fig. 5: Phylum abundance

Composition of abundant phylum in each sample separated by individuals. Phylum composition of blank and control samples were also included in the figure as separate samples. Samples with low sequence depth are marked in red coloured box.

Similar patterns of phyla abundance were observed for the samples from same individuals using the different protocols. However, detailed comparison of the phyla abundance showed that the oral microbiota of individual 1 included a high abundance of Actinobacteria, Proteobacteria and Firmicutes, that of individual 2 included mainly Firmicutes and Bacteroidetes, whereas that of individuals 3 and 4 included mainly Firmicutes, Bacteroidetes and Proteobacteria. Sample 2b from individual 2 which was sequenced using the TS-tailed 1S protocol without internal index was an outlier with only 733 sequences. The relative abundance at the bacterial genus level was measured using the top 30 abundant genera (Supplementary Figure. S4). Similar patterns of genus abundance

were also observed for the samples from same individuals using the different protocols. However, these compositions differed between the individuals in line with the differences at the phylum level (Fig. 5).

Reproducibility and stability of the protocols

Average Shannon diversity for sample 1 was comparatively similar except for the TS-tailed 1S protocol with internal index. In sample 2, NX-tailed 2S and TS-tailed 1S without internal indices protocol yielded comparatively less Shannon diversity. Where as in sample 3 and sample 4 Shannon diversity was comparatively similar for all the protocols. Average Inverse Simpson diversity was comparatively less, using NX-tailed 2S protocol for sample 2, 3 and 4, TS-tailed 1S protocol for sample 1, TS-tailed 1S protocol in sample 1 and 2, and, TS-tailed 1S protocol with internal index for sample 1 (Supplementary Table S5). Intra-class correlation coefficients (ICC) used to enumerate the reproducibility and stability of different protocols for six metrics included relative abundances of four top abundant phyla and two alpha diversity indices showed comparatively better reproducibility and stability with TS-tailed 2S protocol with and without internal index (Fig. 6).

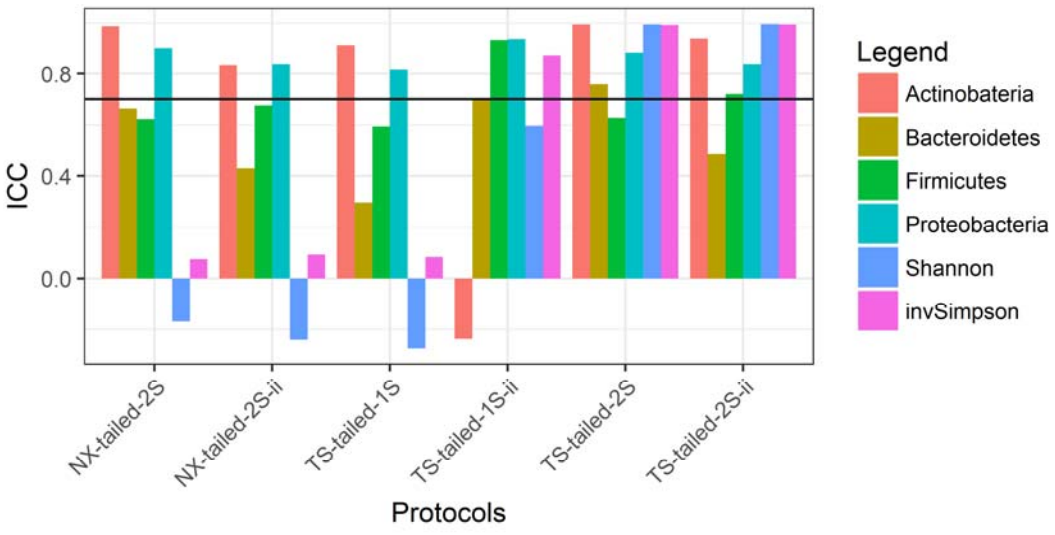


Fig. 6: Reproducibility and stability of the protocols

Intra-class correlation coefficient plotted for six metrics included relative abundances of four major phyla (Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria) and two alpha diversity indices (Shannon & Inverse Simpson index).

Actinobacteria from TS-tailed 1S protocol with internal indices, and Shannon index from TS-tailed 1S, NX-tailed protocol, and NX-tailed protocol with internal index showed negative ICC.

Repeatability of the saliva microbiota with TS-tailed 1S protocol

Repeatability of the saliva microbiota using the TS-tailed 1S protocol, which give the reproducibility and stability, were tested with nine control samples in HiSeq Illumina platform. We also amplified and sequenced negative controls; nine blank samples and two water samples to identify reagents and laboratory contamination. HiSeq platform provided 28936 mean sequences data for nine blank samples, 136554 sequences for nine control samples and 790 mean sequences for water samples. The result showed low diversity for blank samples sequenced and high diversity for the control samples (Fig. 7). None of the sequences from the water samples could be assembled. Mean Shannon diversity was 0.374 and 3.15 and standard deviation (SD) of 0.122 and 0.097, for blank and control samples respectively. Whereas mean inverse Simpson diversity was 1.177 and 13.460 and SD of 0.097 and 1.634, for blank and control samples respectively. Two abundant OTUs from the blank samples were explicitly assigned to two genera of the Proteobacteria phylum, *Pseudomonas* and *Achromobacter*. Bacterial relative abundance of control samples at phyla level shows high abundant of phyla Firmicutes, Bacteroidetes and Proteobacteria (Fig. 5). Relative abundance of bacteria at genus level showed that the control samples were enriched in *Veillonella*, *Prevotella*, *Rothia*, *Neisseria* and *Fusobacterium* spp. (Supplementary Figure. S4).

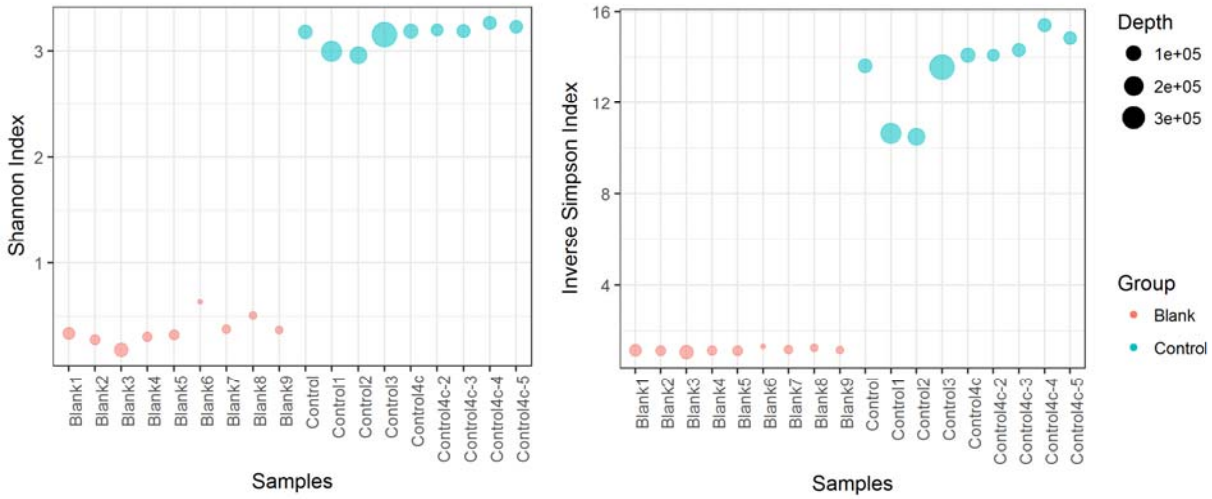


Fig. 7 a & b: Microbial diversity in blank and control samples

Alpha diversity measured using Shannon (7a) and Inverse Simpson (7b) index in blank and control samples data from HiSeq platform.

Discussion

Several studies have successfully used the Illumina technology approaches for 16S rRNA gene amplicon sequencing on diverse sample types (Bartram et al., 2011; Caporaso et al., 2012; Claesson et al., 2010; Degnan and Ochman, 2012; Fadrosh et al., 2014; Gloor et al., 2010; Kozich et al., 2013; Sinclair et al., 2015; Zhou et al., 2011). However, protocols differ in extraction methods, primers, chemistry and sequencing length between studies and a gold standard has not been established. In this study, we compared the reproducibility of six Illumina technology based amplification protocols on saliva samples with primers that were modified in-house (Klindworth et al., 2013; Yuan et al., 2012). We aimed to simplify amplification procedure, investigate barcoding efficacy and expand the number of available barcodes to make 16S assays feasible to run large smaple sets on the HiSeq platform.

Cells may vary in their susceptibility to lysing methods. Various studies have shown that mechanical lysis gives highest bacterial diversity in 16S rRNA gene based studies, notably when communities carry hard to lyse Gram-positive bacteria, such as in faecal samples (Robinson et al., 2016; Salonen et al., 2010; Santiago et al., 2014; Yuan et al., 2012). However, oral samples extracted using either mechanical or enzymatic lysis steps showed an overall similar microbiota profiles (Lazarevic et al., 2013). A recent study also showed that saliva sample collection, storage and genomic DNA preparation with enzymatic-mechanical lysis does not significantly influence the salivary microbiome profiles (Lim et al., 2017). All samples in this study were lysed with an identical protocol including both enzymatic and mechanical disruption of microbial cells using bead-beating to reduce the bias may arise due to the lysis step.

The four saliva samples in triplicates analysed in MiSeq using the different protocols provided comparatively high sequencing coverage for the TS-tailed protocols (>10 k) and less for all other protocols (< 10 k). With current read length of 251 x 2 bp, the V3-V4 region of the rRNA gene is a possible target for sequencing (Mizrahi-Man et al., 2013), although satisfactory quality of the overlap of the forward and reverse paired-end reads may be challenging. However, extending the sequencing cycle up to 271 bp on the Illumina HiSeq platform provided ample overlap, and assembling these reads increases the reliability and quality in the overlapping region. In MiSeq, overall, 39% - 68% of the reads were discarded due to the low-quality score, unassembled pairs, assembled pairs with mismatched barcodes, minimum overlap length and archaeal or eukaryotic sequences. Quality trimming of the NX-tailed protocol sequence data discarded a lower number (6% -8%) of data though it yielded fewer sequences than the TS-tailed protocols. Protocols without internal index pairs gave comparatively high percent (>59 %) qualified for the OTU classification per sequence. Saliva samples amplicons processed with internal index pairs had lower OTU classification per sequence. Several studies have shown that high incidence of mismatching barcodes is a main loss factor in the microbiota sequencing studies (Degnan and Ochman, 2012;

Sinclair et al., 2015). This suggests that the fragment length is at the borderline of what will yield high quality sequence for the overlap between the read pairs and adding only a few extra base pairs to the fragment will reduce output quality. Whereas protocols with and without internal index pairs produced different sequence depth and quality data, all the protocols provided similar bacterial profiles for each samples. Studies have reported that, with the dual-index approach large number of samples can be sequenced using a number of primers equal to only twice the square root of the number of samples (Kozich et al., 2013). Dual indexing protocol is modified by adding the heterogeneity spacers to increase nucleotide diversity at the start of sequencing reads (Fadrosh et al., 2014). Dual indexing strategy further modified by adding third Illumina compatible index with variable length heterogeneity spacers to minimizes the need for PhiX spike-in (de Muinck et al., 2017). However, the advantage of dual index with internal index is to reduce the PCR amplification artefacts in high multiplex amplicon sequencing (Peng et al., 2015) and to reduce the cost of sequencing when the study includes a large sample size.

Our results show that low amounts of sequences usually correlate with low diversity. Our sample size was not large enough to conclude that low amounts of sequences was due to the quality or quantity of DNA, technical issues in the lab or difference in robustness of the methods. However, differences in yields using the same DNA, for example seen in sample 1a and 2b, suggest that protocol robustness may cause differences in sequencing yield (Supplementary Figure S6). Laboratory protocol, sequencing platform or error rate and bioinformatics approach can be reason for the majority of variability detected in microbiota studies (Salter et al., 2014; Sinha et al., 2015) but in our study all the protocols delivered overall similar profile of the microbes in the given saliva samples in triplicates. Three OTUs were explicitly assigned only to blank samples in HiSeq run. Negative control samples often yield contaminating bacterial species which may be due to contamination of bacterial DNA in the kits used (Salter et al., 2014). This study also reported that

387 the presence of contaminating sequences is dependent on the amount of biomass in the samples;
388 however, we could not assess this in our samples.

389 Technical challenges have been reported in 16S rRNA amplicon sequencing, such as biases in
390 estimation of population abundance in microbial communities due to the PCR primer selection,
391 PCR template concentration and amplification conditions, pooling of multiple barcodes and
392 sequencing. Hence, it is important to carefully interpret the experimental results from the technical
393 replicates to validate the reproducibility of the methods (Wen et al., 2017). Average alpha-diversity
394 indices for each samples in different protocols yielded comparatively similar profiles with one or
395 two exception, which may due to the low sequence depth. We used the mixed-effect model–based
396 ICC to quantify the reproducibility and stability of the Illumina MiSeq sequencing of saliva
397 microbiome. ICC measures the variability among the multiple measurements for the same sample
398 and assumes that the errors from different measurements have exactly the same statistical
399 distributions and are indistinguishable from each other (Sinha et al., 2016). In our study, based on
400 ICC, sequencing protocols using TS-tailed 2S protocol with and without internal index performed
401 better than NX-tailed protocol and TS-tailed 1S protocols. The negative ICC values observed for all
402 the NX-tailed protocol and TS-tailed 1S protocols may be due to high variation within a subject.

403 Saliva samples sequenced on HiSeq platform yielded high sequence depth ie; 48k – 398k
404 sequences. Variation in technical replicates and low reproducibility, can be overcome by increasing
405 the sequencing depth (Wen et al., 2017), obtainable by the HiSeq platform. Repeatability of the TS-
406 tailed 1S method without internal index for nine control samples sequenced in HiSeq platform was
407 given comparatively high alpha diversity and low variation (SD) among the samples. Alpha
408 diversity was similar for the sample 4 sequencing repeated in MiSeq and HiSeq platform which
409 support the repeatability of method TS-tailed without internal index as good protocol for
410 microbiome studies. The major limitation of this study is the small number of samples tested for
411 each method. However, we believe that, the number of samples and the depth of the sequencing is

412 sufficient to identify method that should not be used, and also indicate the preferred method to use
413 in large scale studies.

414 In conclusion, NX-tailed 2S protocol and TS-tailed both 1S and 2S protocols were able to reproduce
415 bacterial profiles for the samples sequenced, however, in our hands the reproducibility was
416 comparatively highest for the TS-tailed 2S protocols without internal index on the MiSeq platform.
417 Repeatability of the TS-tailed 1S protocol without internal dual index for nine control samples
418 provided high alpha diversity and little variation among the samples. Considering the cost and time
419 efficiency of using this simplified protocol with numerous barcodes suitable for the HiSeq platform,
420 we suggest that the TS-tailed 1S method can be considered the most effective protocol for
421 consistent quantification of bacterial profiles in saliva. Reproducibility and repeatability should be
422 taken into consideration in design of a large-scale epidemiological study using saliva microbiota.

423 **Acknowledgements**

424 We thank the individuals who participated in this study, and the FIMM biobank and FIMM tech
425 centre. We also thank Timo Miettinen from FIMM tech centre for helping with the internal index
426 setup. We also thank our group members for assisting with the fieldwork of the study Nina Jokinen,
427 Jannina Viljakainen, Stephanie von Kraemer, and the scientific advisors' Dr Eva Roos and
428 Professor Anna Elina Lehesjoki.

429 **Ethics approval and consent to participate**

430 The study was approved by the regional Ethics Committee of the Hospital District of Helsinki and
431 Uusimaa (169/13/03/00/10).

432 **Availability of data and materials**

433 The datasets generated during the current study are available in the NCBI-SRA repository, with the
434 accession number SRP117317.

435 **Competing interests**

436 The authors have no potential conflicts of interest to declare.

437 **Funding**

438 This work was supported by Folkhälsan Research Foundation; Academy of Finland [grant number
439 250704]; Life and Health Medical Fund [grant number 1-23-28]; The Swedish Cultural Foundation
440 in Finland [grant number 15/0897]; Signe and Ane Gyllenberg Foundation [grant number 37-1977-
441 43]; and Yrjö Jahnsson Foundation [grant number 11486].

442 ## References

443 Andersson, A.F., Lindberg, M., Jakobsson, H., Backhed, F., Nyren, P., Engstrand, L., 2008.
444 Comparative analysis of human gut microbiota by barcoded pyrosequencing. PLoS One 3, e2836.
445 https://doi.org/10.1371/journal.pone.0002836

446 Bartram, A.K., Lynch, M.D.J., Stearns, J.C., Moreno-Hagelsieb, G., Neufeld, J.D., 2011.
447 Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial
448 communities by assembling paired-end Illumina reads. Appl. Environ. Microbiol. 77, 3846–3852.
449 https://doi.org/10.1128/AEM.02772-10

450 Belstrøm, D., Holmstrup, P., Bardow, A., Kokaras, A., Fiehn, N.E., Paster, B.J., 2016. Temporal
451 stability of the salivary microbiota in oral health. PLoS One 11, 1–9.
452 https://doi.org/10.1371/journal.pone.0147472

453 Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M.,
454 Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J.A., Smith, G., Knight, R., 2012. Ultra-
455 high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME
456 J. 6, 1621–1624. https://doi.org/10.1038/ismej.2012.8

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

459 Claesson, M.J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J.R., Ross, R.P., O'Toole, P.W.,
460 2010. Comparison of two next-generation sequencing technologies for resolving highly complex
461 microbiota composition using tandem variable 16S rRNA gene regions. Nucleic Acids Res. 38,
462 e200. https://doi.org/10.1093/nar/gkq873

463 de Muinck, E.J., Trosvik, P., Gilfillan, G.D., Hov, J.R., Sundaram, A.Y.M., 2017. A novel ultra
464 high-throughput 16S rRNA gene amplicon sequencing library preparation method for the Illumina
465 HiSeq platform. Microbiome 5, 68. https://doi.org/10.1186/s40168-017-0279-1

466 Degnan, P.H., Ochman, H., 2012. Illumina-based analysis of microbial community diversity. ISME
467 J. 6, 183–194. https://doi.org/10.1038/ismej.2011.74

468 Dewhirst, F.E., Chen, T., Izard, J., Paster, B.J., Tanner, A.C.R.R., Yu, W.H., Lakshmanan, A.,
469 Wade, W.G., 2010. The human oral microbiome. J. Bacteriol. 192, 5002–5017.
470 https://doi.org/10.1128/JB.00542-10

471 Ding, T., Schloss, P.D., 2014. Dynamics and associations of microbial community types across the
472 human body. Nature 509, 357–60. https://doi.org/10.1038/nature13178

473 Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., Knight, R., 2011. UCHIME improves
474 sensitivity and speed of chimera detection. Bioinformatics 27, 2194–2200.
475 https://doi.org/10.1093/bioinformatics/btr381

476 Fadrosh, D.W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R.M., Ravel, J., 2014. An
477 improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina
478 MiSeq platform. Microbiome 2, 6. https://doi.org/10.1186/2049-2618-2-6

479 Gloor, G.B., Hummelen, R., Macklaim, J.M., Dickson, R.J., Fernandes, A.D., MacPhee, R., Reid,
480 G., 2010. Microbiome profiling by illumina sequencing of combinatorial sequence-tagged PCR
481 products. PLoS One 5. https://doi.org/10.1371/journal.pone.0015406

482 Haegeman, B., Hamelin, J., Moriarty, J., Neal, P., Dushoff, J., Weitz, J.S., 2013. Robust estimation
483 of microbial diversity in theory and in practice. ISME J. 7, 1092–1101.
484 https://doi.org/10.1038/ismej.2013.10

485 Hamady, M., Walker, J.J., Harris, J.K., Gold, N.J., Knight, R., 2008. Error-correcting barcoded
486 primers for pyrosequencing hundreds of samples in multiplex. Nat. Methods 5, 235–237.
487 https://doi.org/10.1038/nmeth.1184

488 Human Microbiome Project Consortium, 2012. Structure, function and diversity of the healthy
489 human microbiome. Nature 486, 207–214. https://doi.org/10.1038/nature11234

490 Janem, W.F., Scannapieco, F.A., Sabharwal, A., Tsompana, M., Berman, H.A., Haase, E.M.,
491 Miecznikowski, J.C., Mastrandrea, L.D., 2017. Salivary inflammatory markers and microbiome in
492 normoglycemic lean and obese children compared to obese children with type 2 diabetes. PLoS One
493 12, e0172647. https://doi.org/10.1371/journal.pone.0172647

494 Klindworth, A., Pruesse, E., Schweer, T., Peplies, J.J., Quast, C., Horn, M., Gl??ckner, F.O.,
495 Glockner, F.O., 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical
496 and next-generation sequencing-based diversity studies. Nucleic Acids Res. 41, e1.
497 https://doi.org/10.1093/nar/gks808

498 Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 2013. Development of a
499 dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the
500 miseq illumina sequencing platform. Appl. Environ. Microbiol. 79, 5112–5120.
501 https://doi.org/10.1128/AEM.01043-13

502 Krishnan, K., Chen, T., Paster, B.J., 2017. A practical guide to the oral microbiome and its relation
503 to health and disease. Oral Dis. 23, 276–286. https://doi.org/10.1111/odi.12509

504 Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M.L., Pace, N.R., 1985. Rapid determination
505 of 16S ribosomal RNA sequences for phylogenetic analyses. Proc. Natl. Acad. Sci. U. S. A. 82,
506 6955–6959. https://doi.org/10.1073/pnas.82.20.6955

507 Lazarevic, V., Gaïa, N., Girard, M., François, P., Schrenzel, J., 2013. Comparison of DNA
508 Extraction Methods in Analysis of Salivary Bacterial Communities. PLoS One 8.
509 https://doi.org/10.1371/journal.pone.0067699

510 Li, J., Jia, H., Cai, X., Zhong, H., Feng, Q., Sunagawa, S., Arumugam, M., Kultima, J.R., Prifti, E.,
511 Nielsen, T., Juncker, A.S., Manichanh, C., Chen, B., Zhang, W., Levenez, F., Wang, J., Xu, X.,
512 Xiao, L., Liang, S., Zhang, D., Zhang, Z., Chen, W., Zhao, H., Al-Aama, J.Y., Edris, S., Yang, H.,
513 Wang, J., Hansen, T., Nielsen, H.B., Brunak, S., Kristiansen, K., Guarner, F., Pedersen, O., Doré,
514 J., Ehrlich, S.D., Pons, N., Le Chatelier, E., Batto, J.-M., Kennedy, S., Haimet, F., Winogradski, Y.,
515 Pelletier, E., LePaslier, D., Artiguenave, F., Bruls, T., Weissenbach, J., Turner, K., Parkhill, J.,
516 Antolin, M., Casellas, F., Borruel, N., Varela, E., Torrejon, A., Denariaz, G., Derrien, M., van
517 Hylckama Vlieg, J.E.T., Viega, P., Oozeer, R., Knoll, J., Rescigno, M., Brechot, C., M'Rini, C.,
518 Mérieux, A., Yamada, T., Tims, S., Zoetendal, E.G., Kleerebezem, M., de Vos, W.M., Cultrone, A.,
519 Leclerc, M., Juste, C., Guedon, E., Delorme, C., Layec, S., Khaci, G., van de Guchte, M.,
520 Vandemeulebrouck, G., Jamet, A., Dervyn, R., Sanchez, N., Blottière, H., Maguin, E., Renault, P.,
521 Tap, J., Mende, D.R., Bork, P., Wang, J., 2014. An integrated catalog of reference genes in the
522 human gut microbiome. Nat. Biotechnol. 32, 834–841. https://doi.org/10.1038/nbt.2942

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

526 Lozupone, C.A., Knight, R., 2008. Species divergence and the measurement of microbial diversity.
527 FEMS Microbiol. Rev. https://doi.org/10.1111/j.1574-6976.2008.00111.x

528 McMurdie, P.J., Holmes, S., 2014. Waste Not, Want Not: Why Rarefying Microbiome Data Is
529 Inadmissible. PLoS Comput. Biol. 10. https://doi.org/10.1371/journal.pcbi.1003531

530 Mizrahi-Man, O., Davenport, E.R., Gilad, Y., 2013. Taxonomic Classification of Bacterial 16S
531 rRNA Genes Using Short Sequencing Reads: Evaluation of Effective Study Designs. PLoS One 8,
532 e53608. https://doi.org/10.1371/journal.pone.0053608

533 Nicholson, J.K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., Pettersson, S., 2012.
534 Host-Gut Microbiota Metabolic Interactions. Science (80-.). 336, 1262–1267.
535 https://doi.org/10.1126/science.1223813

536 Paster, B.J., Boches, S.K., Galvin, J.L., Ericson, R.E., Lau, C.N., Levanos, V.A., Sahasrabudhe, A.,
537 Dewhirst, F.E., 2001. Bacterial diversity in human subgingival plaque. J. Bacteriol. 183, 3770–
538 3783. https://doi.org/10.1128/JB.183.12.3770-3783.2001

539 Peng, Q., Vijaya Satya, R., Lewis, M., Randad, P., Wang, Y., 2015. Reducing amplification
540 artifacts in high multiplex amplicon sequencing by using molecular barcodes. BMC Genomics 16,
541 589. https://doi.org/10.1186/s12864-015-1806-8

542 Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, S., Manichanh, C., Nielsen, T., Pons, N.,
543 Yamada, T., Mende, D.R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie,
544 Y., Tap, J., Lepage, P., Bertalan, M., Batto, J., Hansen, T., Paslier, D. Le, Linneberg, A., Nielsen,
545 H.B., Pelletier, E., Renault, P., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., 2010. A
546 human gut microbial gene catalog established by metagenomic sequencing. Nature 464, 59–65.
547 https://doi.org/10.1038/nature08821.A

548 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O.,
549 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based
550 tools. Nucleic Acids Res. 41, D590–D596. https://doi.org/10.1093/nar/gks1219

551 Robinson, C.K., Brotman, R.M., Ravel, J., 2016. Intricacies of assessing the human microbiome in
552 epidemiological studies. Ann. Epidemiol. 26, 311–321.
553 https://doi.org/10.1016/j.annepidem.2016.04.005

554 Salonen, A., Nikkila, J., Jalanka-Tuovinen, J., Immonen, O., Rajilic-Stojanovic, M., Kekkonen,
555 R.A., Palva, A., de Vos, W.M., 2010. Comparative analysis of fecal DNA extraction methods with
556 phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell
557 lysis. J. Microbiol. Methods 81, 127–134. https://doi.org/10.1016/j.mimet.2010.02.007

558 Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., Turner, P.,
559 Parkhill, J., Loman, N.J., Walker, A.W., 2014. Reagent and laboratory contamination can critically
560 impact sequence-based microbiome analyses. BMC Biol. 12, 87. https://doi.org/10.1186/s12915-
561 014-0087-z

562 Santiago, A., Panda, S., Mengels, G., Martinez, X., Azpiroz, F., Dore, J., Guarner, F., Manichanh,
563 C., 2014. Processing faecal samples: a step forward for standards in microbial community analysis.
564 BMC Microbiol. 14, 112. https://doi.org/10.1186/1471-2180-14-112

565 Scheithauer, T.P.M., Dallinga-Thie, G.M., de Vos, W.M., Nieuwdorp, M., van Raalte, D.H., 2016.
566 Causality of small and large intestinal microbiota in weight regulation and insulin resistance. Mol.
567 Metab. 5, 1–12. https://doi.org/10.1016/j.molmet.2016.06.002

568 Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski,
569 R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn,
570 D.J., Weber, C.F., 2009. Introducing mothur: Open-source, platform-independent, community-
571 supported software for describing and comparing microbial communities. Appl. Environ. Microbiol.
572 75, 7537–7541. https://doi.org/10.1128/AEM.01541-09

573 Sinclair, L., Osman, O.A., Bertilsson, S., Eiler, A., 2015. Microbial community composition and
574 diversity via 16S rRNA gene amplicons: Evaluating the illumina platform. PLoS One 10, 1–18.
575 https://doi.org/10.1371/journal.pone.0116955

576 Sinha, R., Abnet, C.C., White, O., Knight, R., Huttenhower, C., 2015. The microbiome quality
577 control project: baseline study design and future directions. Genome Biol. 16, 276.
578 https://doi.org/10.1186/s13059-015-0841-8

579 Sinha, R., Chen, J., Amir, A., Vogtmann, E., Shi, J., Inman, K.S., Flores, R., Sampson, J., Knight,
580 R., Chia, N., 2016. Collecting fecal samples for microbiome analyses in epidemiology studies.
581 Cancer Epidemiol. Biomarkers Prev. 25, 407–416. https://doi.org/10.1158/1055-9965.EPI-15-0951

582 Tringe, S.G., Hugenholtz, P., 2008. A renaissance for the pioneering 16S rRNA gene. Curr. Opin.
583 Microbiol. 11, 442–446. https://doi.org/10.1016/j.mib.2008.09.011

584 van Nood, E., Vrieze, A., Nieuwdorp, M., Fuentes, S., Zoetendal, E.G., de Vos, W.M., Visser, C.E.,
585 Kuijper, E.J., Bartelsman, J.F.W.M., Tijssen, J.G.P., Speelman, P., Dijkgraaf, M.G.W., Keller, J.J.,
586 2013. Duodenal Infusion of Donor Feces for Recurrent *Clostridium difficile*. N. Engl. J. Med. 368,
587 407–415. https://doi.org/10.1056/NEJMoa1205037

588 Wen, C., Wu, L., Qin, Y., Van Nostrand, J.D., Ning, D., Sun, B., Xue, K., Liu, F., Deng, Y., Liang,
589 Y., Zhou, J., 2017. Evaluation of the reproducibility of amplicon sequencing with Illumina MiSeq
590 platform. PLoS One 12, e0176716. https://doi.org/10.1371/journal.pone.0176716

591 Wilkinson, L., 2011. ggplot2: Elegant Graphics for Data Analysis by WICKHAM, H. Biometrics
592 67, 678–679. https://doi.org/10.1111/j.1541-0420.2011.01616.x

593 Yuan, S., Cohen, D.B., Ravel, J., Abdo, Z., Forney, L.J., 2012. Evaluation of Methods for the
594 Extraction and Purification of DNA from the Human Microbiome. PLoS One 7, e33865.
595 https://doi.org/10.1371/journal.pone.0033865

Zheng, W., Tsompana, M., Ruscitto, A., Sharma, A., Genco, R., Sun, Y., Buck, M.J., 2015. An accurate and efficient experimental approach for characterization of the complex oral microbiota. Microbiome 3, 48. https://doi.org/10.1186/s40168-015-0110-9

Zhou, H.-W., Li, D.-F., Tam, N.F.-Y., Jiang, X.-T., Zhang, H., Sheng, H.-F., Qin, J., Liu, X., Zou, F., 2011. BIPES, a cost-effective high-throughput method for assessing microbial diversity. ISME J. 5, 741–749. https://doi.org/10.1038/ismej.2010.160