



Published in final edited form as:

Methods Mol Biol. 2019 ; 1922: 525–548. doi:10.1007/978-1-4939-9012-2_40.

The supragingival biofilm in early childhood caries: clinical and laboratory protocols and bioinformatics pipelines supporting

Correspondence: Kimon Divaris, DDS, PhD, 228 Brauer Hall, Department of Pediatric Dentistry, UNC School of Dentistry, CB#7450, University of North Carolina-Chapel Hill, Chapel Hill, NC 27599-7450, United States, T: (+1) 919-537-3556; F: (+1) 919-573-3950; Kimon_Divaris@unc.edu.

1. When feasible, obtaining tooth surface-specific biofilm samples is in principle preferable than collecting pooled samples—tooth surface-specific samples can be more informative as they can be linked to localized disease (e.g., a specific caries lesion) [5] or other anatomical or ecological features. Of note, the protocol does not describe biofilm collection from occlusal tooth surfaces, which are systematically different than facial/buccal surfaces of the maxillary dentition and perhaps most informative in some studies or for some analyses
2. Adherence to instructions (e.g., not brushing prior to biofilm collection) is an important factor affecting the amount (and potentially the quality) of biofilm collected for analysis
3. Efforts to collect a ‘full thickness’ biofilm sample are well-invested—arguably, the biofilm closest to the tooth surface is at least equally informative for ECC than the outmost layer. We recommend a sweeping movement with the toothpick from the distal to the mesial of each facial tooth surface, while staying supra-gingivally at all times.
4. If a sterile curette can be used instead of toothpicks for supragingival biofilm collection it may be preferable, as it would eliminate the need to separate biofilm material from the toothpick in later steps—in our case we chose sterile tooth picks for practical reasons, due to the conduct of the study in the field.
5. Quality assessment and validation studies are highly recommended to ensure that collection methods, cold chain, nucleic acid purification methods and all other processes are working as expected to produce the desired downstream ‘omics data.
6. During the manual nucleic acid purification process using the Mo Bio kit, we have used the Mo Bio vortex adapter and a Fisher horizontal vortex mixer (CAT #02215450) interchangeably, with no differences in NA yield or quality.
7. WGS shotgun is a versatile method that allows taxonomic profiling of microbial communities to species or, when utilizing additional tools, down to the strain level—it also enables functional profiling of metagenomic and metatranscriptomic sequence data, including aggregated whole-community level pathway reconstruction.
8. Metatranscriptomic sequence data are a logical and informative complement to metagenomics because they can provide information regarding the functional activity of the identified microbial communities at different taxonomic levels. Importantly, they may help illuminate ecological units with high transcriptional activity that might otherwise not be identified or appreciated from a taxonomic standpoint, due to rareness.
9. Segata and colleagues [37] offer an excellent review of technological and computational approaches available for microbiome research. The development of analytical tools for metagenomics and metatranscriptomic analyses is rapidly evolving—association analyses of microbial communities at different levels and respective functional profiles can be performed by stand-alone software specifically developed for this purpose (e.g., MaAsLin, LEfSe, ALDEx2) or using individualized approaches in standard statistical environments (e.g., R) and universal packages (e.g. lm, glm, glmm, etc.)
10. Handling and processing of WGS data is time and computationally intensive, involving numerous steps. We recommend that WGS analysis pipelines are well-documented and fully automatized or divided into broad processes (e.g., KneadData tool for quality control) to promote efficiency and reproducibility of multi-stage cascades of bioinformatics tools.
11. We recommend the consideration of cloud-based computing solutions as a means to accommodate the ever-increasing computational demands of WGS analyses in large-scale studies or in settings where advanced, resource-intensive statistical methods are used, while balancing cost.
12. In metabolomics analyses, there can be similarities between biochemical compounds based on RI index, mass match and MS/MS scores—the simultaneous use of all three estimates helps distinguish and differentiate these molecules. Currently, there are over 3,300 commercially purified standard compounds available into Metabolon’s LIMS. Additional mass spectral entries have been created for structurally unnamed biochemicals, which have been identified by virtue of their recurrent nature, both chromatographic and mass spectral. These compounds have the potential to be identified in the future.
13. We recommend including several blanks (e.g., 10 toothpick fragments without samples, handled according to study conditions) in metabolomics analyses, to detect and account for any resulting background noise due to this collection medium.
14. For metabolomics analyses that do not span more than one day, no normalization is necessary, other than for purposes of data visualization. Biochemical data can be normalized to an additional factor (e.g., cell counts, total protein as determined by Bradford assay, osmolality, etc.) to account for differences in metabolite levels due to differences in the amount of material present in each sample.

oral metagenomics, metatranscriptomics and metabolomics studies of the oral microbiome.

Kimon Divaris^{1,2}, Dmitry Shungin^{3,4}, Adaris Rodríguez-Cortés^{2,5}, Patricia V. Basta^{2,5}, Jeff Roach⁶, Hunyong Cho⁷, Di Wu⁸, Andrea G. Ferreira Zandoná⁹, Jeannie Ginnis¹, Sivapriya Ramamoorthy¹⁰, Jason M. Kinchen¹⁰, Jakub Kwintkiewicz^{11,12}, Natasha Butz^{11,12}, Apoena Aguiar Ribeiro¹³, and M. Andrea Azcarate-Peril^{11,12}

¹Department of Pediatric Dentistry, UNC School of Dentistry, University of North Carolina-Chapel Hill, Chapel Hill, NC, United States

²Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina-Chapel Hill, Chapel Hill, NC, United States

³Department of Odontology, Umeå University, Umeå, Sweden

⁴Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA, United States

⁵Biospecimen core processing facility, Gillings School of Global Public Health, University of North Carolina-Chapel Hill, Chapel Hill, NC, United States

⁶Research Computing, University of North Carolina-Chapel Hill, Chapel Hill, NC, United States

⁷Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina-Chapel Hill, Chapel Hill, NC, United States

⁸Department of Periodontology, UNC School of Dentistry, University of North Carolina-Chapel Hill, Chapel Hill, NC, United States

⁹Department of Comprehensive Dentistry, Tufts University School of Dental Medicine, Boston, MA, United States

¹⁰Metabolon, Inc., Durham, North Carolina 27560, United States

¹¹Microbiome Core Facility, Microbiome Core Facility, Department of Cell Biology and Physiology, School of Medicine, University of North Carolina-Chapel Hill, Chapel Hill, NC, United States

¹²Department of Medicine, Division of Gastroenterology and Hepatology, School of Medicine, University of North Carolina-Chapel Hill, Chapel Hill, NC, United States

¹³Department of Diagnostic Sciences, UNC School of Dentistry, University of North Carolina-Chapel Hill, Chapel Hill, NC, United States

Abstract

Early childhood caries (ECC) is a biofilm-mediated disease. Social, environmental and behavioral determinants as well as innate susceptibility are major influences on its incidence; however, from a pathogenetic standpoint, the disease is defined and driven by oral dysbiosis. In other words, the disease occurs when the natural equilibrium between the host and its oral microbiome shifts towards states that promote demineralization at the biofilm-tooth surface interface. Thus, a comprehensive understanding of dental caries as a disease requires the characterization of both the

composition and the function or metabolic activity of the supragingival biofilm according to well-defined clinical statuses. However, taxonomic and functional information of the supragingival biofilm is rarely available in clinical cohorts, and its collection presents unique challenges among very young children. This paper presents a protocol and pipelines available for the conduct of supragingival biofilm microbiome studies among children in the primary dentition that has been designed in the context of a large-scale population-based genetic epidemiologic study of ECC. The protocol has been developed for the collection of two supragingival biofilm samples from the maxillary primary dentition, enabling downstream taxonomic (e.g., metagenomics) and functional (e.g., transcriptomics and metabolomics) analyses. The protocol is being implemented in the assembly of a pediatric precision medicine cohort comprising over 5,000 participants to-date, contributing social, environmental, behavioral, clinical and biological data informing ECC and other oral health outcomes.

Keywords

dental caries; microbiome; transcriptome; metabolome; children; protocol

1. Introduction

Early childhood caries (ECC), as all other taxonomic entities of dental caries, is a biofilm-mediated, dysbiotic disease [1, 2]. Social, environmental and behavioral determinants, as well as innate susceptibility are known major influences on its incidence [2, 3]; however, from a pathogenetic standpoint, the disease is defined and driven by oral dysbiosis. In other words, the disease occurs when the natural equilibrium between the host and its oral microbiome shifts towards states that promote demineralization at the biofilm-tooth surface interface [4–6]. Conceptualizing ECC as an oral-ecological imbalance is a shift away from currently used clinical definitions [7]. The latter are undoubtedly useful for classifying and reporting clinical outcomes of the disease process in clinical practice, research, as well as dental public health and surveillance. However, for the realization of precision oral health care [2, 8], the biological processes underlying these clinical manifestations must be understood and defined.

Measures of the composition and activity of the supragingival biofilm can be regarded as endophenotypes [8], similar to that of cholesterol and systemic inflammatory markers in the context of cardiovascular health—proximal, sensitive traits of pathogenetic activity. A comprehensive understanding of ECC -and all presentations of dental caries- as a disease process (versus its clinical manifestation) requires the characterization of both the composition and the function or metabolic activity of the supragingival biofilm according to well-defined clinical statuses. However, taxonomic and functional information of the supragingival biofilm is rarely available in clinical cohorts, and its collection presents unique challenges among very young children. This paper presents a protocol and pipelines available for supragingival biofilm microbiome studies among children in the primary dentition, implemented in a large-scale population-based genetic epidemiologic study of early childhood caries (ECC). The protocol has been developed for the collection of two supragingival biofilm (plaque) samples from the maxillary primary dentition, enabling

downstream taxonomic (e.g., metagenomics [9]) and functional (e.g., transcriptomics [10] and metabolomics [11]) analyses. We are implementing this protocol in the assembly of a pediatric precision medicine cohort currently comprising over 5,000 participants, contributing a multitude of data informing ECC and other oral health outcomes. The project is the ZOE-G4S (Zero-out Early Childhood Caries- “Genes for Smiles”; ZOE 2.0) study—a NIH-funded genetic epidemiologic investigation among a large population-based sample of preschool-age children enrolled in Head Start centers in North Carolina. A key feature of the protocol is that it outlines biofilm collection and storage procedures that can be done under field conditions, hours or days away from a clinic or laboratory.

Supragingival biofilm is being collected from participating 3 to 5-year-old children by clinical examiners (registered dental hygienists or dentists) who are trained and calibrated for dental caries experience assessment. Study staff instruct families and school teachers not to brush the children’s teeth the morning of the study visit. After saliva (for genomic DNA extraction purposes) and biofilm collection, the examiners brush and floss children’s teeth and, besides dental caries experience, they record developmental defects of the enamel, anthropometric measures, extra-oral characteristics, dental occlusion classification, dental trauma prevalence and a summary of treatment needs.

2. Equipment, materials and software

The biofilm sample collection is done in the context of a main clinical protocol that entails measurement of dental caries experience and other clinical indices, requiring a full dental set-up, as detailed in §2.1. Conceivably, dental biofilm collection can be done without a full dental set-up (e.g., with the child sitting on a regular versus a dental chair)—however, in most instances, biofilm collection will be accompanied by measurement of clinical characteristics and, for this reason, a dental set-up will be necessary and available. The samples are frozen at the collection site and transferred to a core lab facility for long-term storage or nucleic acid (NA) extraction.

Several NA extraction methods from biofilms are available; our team has been testing and optimizing both manual and high-throughput (96-well plate) purification methods. After purification and quantitation, NA preps are transferred to a sequencing facility and carried forward to whole genome shotgun (WGS) and RNA sequencing. Our group has verified and reported that the described procedures lead to good quality and informative transcriptome (unpublished data) and metagenomics results [12] with respect to oral health and ECC statuses. Bioinformatics pipelines are then executed on a research computing cluster to carry out steps described in §3.6 as well as additional data management and statistical analyses.

Biofilm samples collected for metabolomics analyses are bio-banked in -80°C . Metabolon® uses multiple mass spectrometry methods, a large reference library of authenticated metabolite standards and a suite of patented informatics and quality-control software in the DiscoveryHD4™ platform [13–17]. A pilot study including metabolomics analysis of 300 ZOE 2.0 biofilm samples identified 503 compounds of known identity (named biochemicals) in these samples.

2.1 Clinical examination equipment and instruments

- Foldable dental chair (Aseptico® ADC-01P-RED)
- Foldable hydraulic dentist stool (Aseptico® ADC-10)
- Dental instrument tray (Aseptico® ATC-03CF)
- Examiner magnifying loupes with headlight (Orascoptic XV1™ Loupe + Light)

2.2 Clinical examination supplies

- Disposable gowns
- Gloves
- Masks
- Bibs
- Disposable dental examination mirrors
- Barrier tape
- Plastic chair and tray covers
- Surface disinfecting wipes
- Hand sanitizer

2.3 Supragingival biofilm collection equipment

- CoolBox™ XT (BioCision®; BCS-575) or 2XT (BioCision®; BCS-573) Workstations, with Freezing Cores (BioCision®; BCS-512), and M24 (BioCision®; BCS-535) and CFT24 (BioCision®; BCS-534) CoolRacks® for on-site immediate freezing and transportation of biofilm samples
- Fridge Freezer 50 qt. (ARB®; item #10800472) for long-distance transport of samples
- Power cord (ARB®; item #10910076) for in-car power supply of portable freezer
- Canvas travel bag (ARB®; item #10900013), optional for portable freezer
- Fridge remote monitor (ARB®; item #10900026), optional for portable freezer
- TruCool® Hinged Cryoboxes, pack of 50, green color (BioCision®; BCS-207G), for storage of RNA*later* tubes and cryovials (referenced below, in §2.4)

2.4 Supragingival biofilm collection supplies

- Sterile wooden toothpicks, 2 per pack (DeRoyal; product no. 30–413)
- RNA*later* 1.5ml TissueProtect tubes (QIAGEN; product no. 196594) for storage of biofilm samples to be carried forward to metagenomics and metatranscriptomics analyses

- TruCool™ Cryogenic Vials, sterile, 1ml, self-standing, external threads (BioCision®; BSC-2517) for downstream metabolomics analyses

2.5 Materials used for nucleic acid purification

- Total RNA Purification Kit, 96-Well Format (Norgen Biotek, Corp.; P/N 24304, L/N 591338) which includes:
 - Robust Lysis (RL) buffer (P/N 90055, L/N 589694)
 - Wash Solution A (P/N 90028, L/N 590111)
 - 96-Well Filter Plate (P/N 24304, L/N 591338)
 - 96-Well Collection Plate (P/N 24309)
 - 96-Well Elution Plate (P/N 24310)
- 96-Well Bead Plate (Norgen Biotek, Corp.; P/N 65700, L/N 591926) kit, which includes:
 - 96-Well Transfer Plate (P/N 65702)
 - 96-Well Bead Plate (P/N 65703, L/N 591924)
- Beta-Mercaptoethanol (B-ME), CAS [60–24-2] (MP Biomedicals, LLC; P/N 194834, L/N QR13543)
- Ethanol 200 Proof, Anhydrous, CAS [64–17-5] (Decon Labs, Inc.; P/N 2716, L/N DSP-MD.43)
- Molecular Biology Grade Water, Sterile, CAS [7732–18-5] (Corning; P/N 46–000-CM, L/N 14917003)

2.6 Software used for metagenomics and metatranscriptomics analyses

- Illumina bcl2fastq [18]
- FastQC [19]
- bowtie2 [20]
- vsearch [21]
- HUMAnN2 [22]
- Pathoscope2 [23]
- LEfSe [24]
- ALDEx2 R package [25]
- MetaPhlAn2 [26]
- KneadData (<http://huttenhower.sph.harvard.edu/kneaddata>)
- MaAsLin (<https://huttenhower.sph.harvard.edu/maaslin>)
- GraPhlAn (<https://huttenhower.sph.harvard.edu/graphlan>)

- ShotMAP [27]

Figure 1 illustrates basic small equipment (excluding CoolBoxes) and supplies needed for supragingival biofilm collection. Figure 2 illustrates an assembled CoolBox unit with CFT 24 CoolRack and a TruCool Cryogenic vial in place. Additional supplies and small equipment assemblies are shown in the supplementary material (supplementary material).

3. Methods

In the parent study, research examination teams use mobile dental equipment, instruments, and supplies to set up clinical examination stations where there is suitable space and a power outlet in community locations (e.g., classrooms). Examiners are trained in the use of basic child behavior guidance and management techniques routinely used in pediatric dentistry. The examination sequence including biofilm collection takes place before, or at least 30 (preferably 60) minutes after children have had breakfast or snack. The families (and teachers, where applicable) are instructed not to brush the children's teeth the morning of the examination-the examiner cleans the teeth with a toothbrush and dental floss after biospecimen collection. At the end of the day, the biofilm samples are stored in a -20°C freezer until they are transferred to a core facility that is equipped with -80°C freezers. The procedures are carried out as follows:

3.1 Pre-clinical procedures and assessments

1. Greet the child addressing them by their name and ideally kneeling down to their eye level
2. Explain the procedures that will follow in an age-appropriate manner (e.g., "counting teeth")

3.2 Supragingival biofilm collection and storage procedures

1. Collect supragingival biofilm (plaque) samples using the sterile toothpicks in a semi-reclined position, ideally a dental chair. Break off a small piece (~8–12mm, or 1/3''–1/2'') of each toothpick and use to collect the biofilm
2. One sample (biofilm sample A) from the upper right dentition (facial surfaces of #A, #B, #C, #D and #E according to the Universal nomenclature system; or #55, #54, #53, #52 and #51 according to the FDI system) and another sample (biofilm sample B) from the upper left dentition (facial surfaces of #F, #G, #H, #I and #J, or #61, #61, #63, #64 and #61)
3. Place biofilm sample A (part of toothpick with metagenomics/metatranscriptomics sample) in an RNA^{later} TissueProtect 1.5ml tube and biofilm sample B (part of toothpick with metabolomics sample) in a Cryovial
4. Add barcode or any other identifying labels on both vials
5. Store in CoolRacks/CoolBoxes and close lid hermetically
6. Add barcode or any other identifying label or notation on a sample manifest, that will accompany the stored vials

3.3 Nucleic acid purification protocols

Our group has optimized and tested both manual and high-throughput protocols for the processing of supragingival biofilm among child research participants. We have used protocol §3.3.1 below to conduct metagenomics and transcriptomics analyses among 118 study participants [12]—for that work, we prioritized samples yielding 1 µg of total Nucleic Acid (NA) and carried them forward to WGS and RNAseq analyses, obtaining on average more than 6 million high-quality reads per sample and informative results. The high-throughput protocol §3.3.2 is currently being optimized and tested in a similar fashion. In general, we strongly recommend that nucleic acid protocols are pilot-tested and validated, using clinical collection, biofilm specimen transport and storage, and analytical procedures identical to the study conditions.

3.3.1 Sample preparation for NA extraction.

1. Inspect the sample vials and ensure integrity of the sample vessel. If you notice cracks on the sample vial, transfer the sample to a DNase and RNase-free screw cap vial that has been properly labeled
2. Thaw plaque samples at 37°C (~5 min in water bath, ~10 min in heating block) and vortex for 5 minutes (use Mo Bio vortex adapter or horizontal vortex mixer)
3. Spin tubes for 15 min at $14000 \times g$ ($13200 \times g$)
4. Using clean forceps, transfer the toothpick fragment into the bead tube (pointed end first). Importantly, make note of plaque visibility on toothpick fragment. Always clean forceps between toothpick fragment transfers
5. Remove most, if not all, of the RNA*later* solution, without disturbing the pellet. A small amount of liquid, ~50–100 µL, can be left behind. Make notes as to the size of the pelleted material
6. Begin processing the pellet from the RNA*later* TissueProtect tube according to the manufacturer's instructions as described in section 3.3.2.

We have determined that, using the protocols detailed in this paper, the sequential processing of the nucleic acid sample is more efficient in terms of yield than splitting the sample into DNA and RNA using the AllPrep DNA/RNA Mini Kit (QIAGEN; catalogue no. 80204).

3.3.2 Manual total NA purification—We carry out the manual purification protocol (Mo Bio PowerBiofilm RNA kit) according to the manufacturer's protocol with minor modifications, as follows:

- a. Thaw samples at 37°C for 10 min
- b. Shake the samples at maximum speed for 5 min on a horizontal vortex mixer
- c. Centrifuge the plaque pellets and toothpick fragments (used for collection) at $14,000 \times g$ for 15 min
- d. Make note of plaque deposit on the toothpick fragment (not visible, barely visible, visible, or conspicuous; see supplemental material for examples)

- e. Transfer the toothpick fragment into the bead tube
- f. Follow the protocol according to the manufacturer's instructions through step 15
- g. Carry out steps involving disruption of the biofilm in the presence of both the dissolved pellet from the RNA*later* collection tube and the toothpick fragment to optimize yield and minimize material (biofilm) loss from what is remaining on the toothpick
- h. Omit all extraction steps dealing with DNA removal
- i. Conclude elution of a total NA preparation with steps 21– 26 of the manufacturer's protocol

3.3.3 High-throughput protocol—In this protocol, we use the Total RNA Purification Kit, 96-Well Format (Norgen Biotek, Corp.; catalogue no. 24304) and 96-Well Bead Plate Kit (Norgen Biotek, Corp.; catalogue no. 65700). We designed the protocol in a collaborative effort with the kit supplier. The high throughput total nucleic acid isolation method was optimized and validated. The resulting procedure is as follows:

1. Thaw samples at 37°C for 10 min
2. Shake the samples at maximum speed for 5 min on a horizontal vortex mixer
3. Centrifuge the plaque pellets and toothpick fragments (used for collection) at $14,000 \times g$ for 15 min
4. Make note of plaque deposit on the toothpick fragment (not visible, barely visible, visible, or conspicuous; see supplemental material for examples)
5. Transfer the toothpick fragment into the bead tube
6. Make note of plaque pellet in the sample plate (not visible, barely visible, visible, or conspicuous)
7. Remove RNA*later* solution.
 - a. If pellet is not loose, decant the RNA*later* solution
 - b. If the pellet is loose, remove most of the RNA*later* solution by pipetting the liquid out of the tube, leaving up to 100 μL of solution
8. Prepare Robust Lysis (RL) buffer with Beta-Mercaptoethanol (B-ME), CAS [60–24–2] (MP Biomedicals, LLC; catalogue no. 194834) (RL/BME) at a ratio of 10 μL of BME per 1 mL of RL buffer (1:100 ratio)
 - a. Prepare enough RL/BME solution to treat all samples been processed and an additional sample ($n + 1$). For example, to process 8 samples, prepare $8 + 1 = 9$, as follows:

Prepare RL/BME | $9 \times 400 \mu\text{L} = 3,600 \mu\text{L}; 36 \mu\text{L } \beta\text{-ME} + 3.6 \text{ mL RL Buffer}$
 - b. Use 400 μL of RL/BME solution per sample

- c. If processing a small number of samples, $n < 5$, prepare enough solution for $n + 0.5$ reactions
- 9. Resuspend each plaque pellet in 400 μ L RL/ME solution
- 10. Transfer the sample in RL/ME solution to a well of a bead plate and secure the silicone mat. Make note of the wells used and sample order
- 11. Vortex briefly to mix
- 12. Incubate the sample at 65°C for 5 min, inverting every two minutes to mix. Secure or the silicone mat with a piece of reinforcement foam, an assembled rounded-bottom 96-well plate and lid, and clamps on all four corners
 - a. Clamps and reinforcement material are not required if a vertical plate homogenizer is used.
- 13. Vortex the sample at maximum speed for 5 min on a horizontal vortex mixer (Fisher Scientific multi-tube vortexer). Disassemble the reinforcement foam, 96-well plate and lid, and clamps
- 14. Centrifuge the bead plate at $3000 \times g$ for 3 min at room temperature
- 15. Dispense 200 μ L of absolute (100%) ethanol into each well of a clean Ribonuclease-free collection microplate (96-Well Bead Plate Kit)
- 16. Collect the supernatants and transfer into the wells of the ethanol-containing collection microplate. Make note of the wells used and sample order
- 17. Mix each lysate and ethanol by gently pipetting up and down 3–5 times
- 18. Bind the nucleic acids to the filter plate as follows:
 - a. Place the 96-well filter plate on top of a provided 96-well collection plate
 - b. Apply up to 500 μ L of the lysate with the ethanol into each well of the 96-well filter plate.
 - c. Centrifuge the assembly at maximum speed or $3,000 \times g$ (~3,000 rpm) for 2 min
 - d. Discard the flow-through. Reassemble the 96-well filter plate and the bottom plate. Ensure that the lysate from each well has passed through into the bottom plate. If the entire lysate volume has not passed, centrifuge for an additional 2 min
 - e. Run all the samples through the spin column
- 19. Wash the nucleic acids thrice (3X) as follows:
 - a. Apply 400 μ L of Wash Buffer A
 - b. Centrifuge the assembly at maximum speed or $3,000 \times g$ (~3,000 rpm) for 2 min

- c. Discard the flow-through
20. Pat the bottom of the 96-Well Filter Plate dry with a clean lab wipe
21. Reassemble the 96-well filter plate and the bottom plate
22. Centrifuge the assembly at maximum speed or $3,000 \times g$ (~3,000 rpm) for 5 min to completely dry the plate
23. Elute the nucleic acids into a new nuclease-free tube by adding 60 μL of pre-warmed nuclease-free water (65°C) directly onto the center of the spin column filter to ensure complete elution of nucleic acids
24. Incubate at room temperature for 1 min
25. Centrifuge the assembly at maximum speed or $3,000 \times g$ (~3,000 rpm) for 2 min
26. Record the volume of the elute (usually ~50 μL)
27. Transfer sample into a labeled nuclease-free screw-cap tube with cap
28. Proceed to nucleic acid quantitation and QA

3.4 Whole genome sequencing shotgun (WGS): DNA requirements, library preparation and sequencing

3.4.1 Several factors can influence DNA quality and thus, read length and quality of sequencing.

To optimize results, each DNA sample should generally meet the following standard requirements:

- a. Has not undergone multiple freeze-thaw cycles as they can lead to DNA damage
- b. Has not been exposed to high temperatures (e.g.: > 65°C for 1 hour can cause a detectable decrease in sequence quality), or pH extremes (< 6 or > 9)
- c. Has an OD260/OD280 ratio of 1.8 to 2.0
- d. Does not contain RNA contamination
- e. Does not contain denaturants (e.g., guanidinium salts or phenol) or detergents (e.g., SDS or Triton X100)
- f. Does not contain carryover contamination

Dilute DNA in DNase-free water or low-salt, weakly buffered solutions containing little or no metal ion chelating agents such as EDTA (example: 10mM Tris). The minimum amount of DNA per sample required for WGS library preparation as described in §3.4.1 is 1 ng.

3.4.2 Process 1 ng of intact genomic DNA using the Nextera XT DNA Sample Preparation Kit (Illumina, San Diego, CA):

1. Quantify the sample DNA using Picogreen reagent and optimize concentration to 0.2ng/ μL

2. Using Nextera XT transposome simultaneously fragment the input DNA and add platform-specific adapter sequences:
 - a. Label a new 96-well plate
 - b. Add 10 μL of Tagment DNA Buffer to each well to be used in this assay. Change tips between samples
 - c. Add 5 μL of input DNA at 0.2 ng/ μL (1 ng total) to each sample well of the plate
 - d. Add 5 μL of Amplicon Tagment Mix to the wells containing input DNA and Buffer. Change tips between samples
 - e. Using a multichannel pipette, gently pipette up and down 5 times to mix. Change tips between samples
 - f. Cover the NTA plate with Microseal
 - g. Centrifuge at $280 \times g$ at 20°C for 1 minute
 - h. Place the NTA plate in a thermocycler and run the following program:
 - a. 55°C for 5 minutes
 - b. Hold at 10°C
3. Once the sample reaches 10°C immediately add 5 μL of Neutralization Buffer to each well of the plate. Change tips between samples
4. Amplify tagmented DNA via a limited-cycle PCR program adding index 1(i7) and index 2(i5) (Illumina) in unique combination for each sample, as well as primer sequences required for cluster formation:
 - a. Add 15 μL of Nextera PCR Master Mix to each well of the plate containing index primers. Change tips between samples
 - b. Using a multichannel pipette, add 5 μL of index 2 primers (white caps) to each column of the plate. Changing tips between columns is required to avoid cross-contamination
 - c. Using a multichannel pipette, add 5 μL of index 1 primers (orange caps) to each row of the plate. Tips must be changed after each row to avoid index cross contamination
 - d. Using a multichannel pipette, gently pipette up and down 3 to 5 times to mix
 - e. Cover the plate with Microseal and seal with a rubber roller
 - f. Centrifuge at $280 \times g$ at 20°C for 1 minute
 - g. Perform PCR using the following program on a thermal cycler:
 - i. 72°C for 3 minutes
 - ii. 95°C for 30 seconds

- iii. 12 cycles of:
 1. 95°C for 10 seconds
 2. 55°C for 30 seconds
 3. 72°C for 30 seconds
- iv. 72°C for 5 minutes
- v. Hold at 10°C

3.4.3 Purify each library using Agencourt® AMPure® XP Reagent (Beckman Coulter, Brea, CA).

This step allows to purify the library DNA, and provides a size selection step that removes very short library fragments from the population:

1. Centrifuge the plate at $280 \times g$ for 1 min (20°C) to collect condensation
2. Vortex the AMPure XP beads for 30 seconds to ensure that the beads are evenly dispersed. Add an appropriate volume of beads to a trough
3. Using a multichannel pipette, add 30 μL of AMPure XP beads to each well of the plate and pipette mix up and down 10 times
4. Incubate at room temperature without shaking for 5 minutes
5. Place the plate on a magnetic stand for 2 minutes or until the supernatant has cleared and remove and discard the supernatant
6. With the plate on the magnetic stand, wash the beads twice with freshly prepared 80% ethanol and allow the beads to air-dry for 15 minutes.
7. Remove the plate from the magnetic stand. Using a multichannel pipette, add 25 μL of Resuspension Buffer to each well.
8. Gently pipette mix up and down 10 times, changing tips after each column and incubate at room temperature for 2 minutes.
9. Place the plate on the magnetic stand for 2 minutes or until the supernatant has cleared.
10. Using a multichannel pipette, carefully transfer 23 μL of the supernatant to a new plate. Change tips between samples to avoid cross contamination.

3.4.4 Quantify each library using Quant-iT™ PicoGreen® dsDNA Reagent (Molecular Probes, Thermo Fisher Scientific division, Waltham, MA) according to manufacturer instructions

3.4.5 Calculate and pool all libraries in equimolar amounts to one tube

3.4.6 Follow the HiSeq 2500 System Guide Protocol (Illumina) for preparation of the library to 10 pM for loading:

1. Heat denature the resulting pool before loading on the HiSeq reagent cartridge and on the HiSeq 2500 instrument (Illumina). Denatured PhiX library can be used as an internal control or to balance low diversity libraries. For libraries with low complexity, combine 30% PhiX with your diluted sample
2. Enter run parameters according to consumable ID, indexing option and flow cell used.
3. If edited properly, for dual-indexed PE250 runs, the instrument will indicate that the run is 250×8×8×250. The run takes approximately 1 week to complete

3.5 RNA sequencing (RNAseq): RNA preparation, requirements and library preparation

3.5.1 RNA preparation and requirements

- a. Suspend RNA samples in RNase-free water or 1X TE buffer prepared with RNase-free water
- b. Assess RNA integrity using the Agilent Bioanalyzer (or any similar system), with optimal RNA Quality Number (RQN) or RNA Integrity Number (RIN) being 8
- c. We recommend a DNase treatment step in the RNA isolation protocol.
- d. We recommend using fluorometric methods such as Quant-iT™ RiboGreen® for RNA quantification

500 ng of total RNA is required for RNA library preparation according to the procedures described in §3.5.2

3.5.2 Library preparation

1. From a sample of total RNA, remove non-coding rRNA using Illumina Ribo-Zero Epidemiology Kit (San Diego, CA).
 - a. Wash the rRNA-specific magnetic beads off the storage buffer
 - b. Mix with 500 ng of total sample RNA with rRNA removal solution and incubate for 10 minutes at 65°C
 - c. Add pre-washed rRNA-specific magnetic beads from step (a) and incubate for 5 minutes at 50°C
 - d. Place samples on magnetic stand for 15 minutes in room temperature
 - e. Aspirate the supernatant containing coding RNA and proceed to library preparation
2. mRNA library preparation using TruSeq® Stranded mRNA Sample Preparation Kit (San Diego, CA).
 - a. Fragment and prime mRNA using divalent cation priming solution at 94°C for 8 minutes

- b. Equilibrate first strand synthesis Actinomycin D mix (FSS) to room temperature
 - c. Add FSS to mRNA samples
 - d. Place samples on the thermal cycler and use the following thermal conditions:
 - i. 25°C for 10 minutes
 - ii. 42°C for 15 minutes
 - iii. 70°C for 15 minutes
 - iv. Hold at 4°C
 - e. Remove samples from thermal cycler
 - f. Add second strand marking master mix (SMM) to each sample
 - g. Incubate at 16°C for 1 hour
 - h. Purify double-stranded cDNA using Agencourt® AMPure® XP Reagent (Beckman Coulter, Brea, CA)
 - i. Adenylate 3' ends of cDNA by adding A-tailing mix to each sample and incubating in the following thermal conditions:
 - i. 37°C for 30 minutes
 - ii. 70°C for 5 minutes
 - iii. Hold at 4°C
 - j. Ligate adapters by adding ligation mix and adapters, each with a unique barcode
 - k. Incubate samples at 30°C for 10 minutes
 - l. Add stop solution to quench ligation reaction
 - m. Purify cDNA libraries using Agencourt® AMPure® XP Reagent
 - n. Enrich DNA fragments by adding PCR primer cocktail and master mix to each sample and incubating in the following thermal conditions for 15 cycles:
 - i. 98°C for 10 seconds
 - ii. 60°C for 30 seconds
 - iii. 72°C for 30 seconds
 - o. Carry out a final purification of cDNA libraries using Agencourt® AMPure® XP Reagent
3. Validate quality of cDNA libraries on Agilent TapeStation

4. Assess library concentration using Quant-iT PicoGreen dsDNA Reagent from Thermo Fisher Scientific (Eugene, OR)
5. Pool all libraries in equimolar amounts

3.6 Analysis methods for metagenomics and metatranscriptomics (bacterial transcriptomics)

1. De-multiplex and convert to FASTQ format paired-end sequencing run results using bcl2fastq [12]
2. Assess quality of sequencing results with FastQC [13]
3. Align paired-end reads against the human Hg19 reference using bowtie2 [14]. Retain reads that do not align.
4. Join paired-end reads that do not align to reference into single-end reads with vsearch [15] when possible. Retain both reads that join as well as reads that do not
5. Combine all reads, joined and un-joined, into a single single-end data set. Align against the human hg19 reference using bowtie2 [14]. Retain all reads that do not align.
6. Perform microbial community profiling using MetaPhlAn2.2 [20]
7. Classify all retained reads by taxonomy, gene family, path coverage, and path abundance using HUMAnN2 [16]
8. Classify all retained reads by taxonomy using Pathoscope2 [17]
9. Profile strain-level variation in microbial communities using StrainPhlAn, as referenced in §2.6
10. Identify differentially abundant taxa with LEfSe [18], ALDEx2 [19] or MaAsLin, as referenced in §2.6. Additional information regarding statistical methods for analyzing differences in taxonomic abundance (metagenomics) and gene expression (metatranscriptomics) is presented in §3.7
11. Produce taxonomic or phylogenetic representations (i.e., circular heat maps and bar plots) of analytical results using GraPhlAn, as referenced in §2.6

3.7 Outline of statistical methods used to identify differentially abundant taxa or differentially expressed bacterial genes

The choice of statistical methods used for association analyses between bacterial taxa or transcripts and health/disease statuses is driven by the distributional assumptions and characteristics of these data. These analyses are, by nature, subject to multiplicity (i.e., multiple testing) issues.

3.7.1 Modeling assumptions—Taxa or gene abundance or expression levels are represented as counts, ratios or proportions: normal, binomial and Beta are the most widely used distributions. In general, it is expected that models that can handle zero-inflation will

yield higher power compared to their non-zero-inflated counterparts, due to the high prevalence of excess zeros in microbiome data. In the context of moderately small numbers of zeros, models that do not explicitly address zero-inflation can also be used. The following models and procedures can be considered for testing such associations with disease status.

i. MAST [28]	zero-inflated log-normal distribution
ii. Two-part Logistic Beta model [29, 30]	Bernoulli and Beta distribution
iii. Two-part Wilcoxon Rank Sum test [31]	Bernoulli and normal (rank)
iv. Zero-inflated negative binomial model [32, 33]	zero-inflated negative binomial
v. LEfSe [24]	normal
vi. Kruskal-Wallis test [34]	normal (rank)

Models i, ii, iv and v can incorporate continuous and/or discrete covariates in the model, while Models iii and vi can be modified to handle discrete covariates (e.g., in the context of batch effects). When the disease status of interest is expressed as a multi-category variable, Models i, ii, iv and v are applicable, whereas Models iii and vi are not. Of note, Models i, ii, iii, and iv are zero-inflated models.

3.7.2 Multiple testing and control of type-I error—Association analyses between bacterial taxa (or transcripts) and health/disease statuses are usually performed at the individual taxon level. Due to the high number of taxa to be tested, a typical multiple comparisons problem arises. To control type-I error, false discovery rate (FDR) controlling procedures and familywise error rate (FWER) controlling procedures can be used to correct/adjust p-values. FWER procedures include, among others, the Bonferroni and Holm corrections; these provide relatively stringent control of type-I error and are less powered than FDR. Typical FDR procedures include the Benjamini-Hochberg (BH) and Benjamini-Yekutieli (BY) corrections. Both FWER and FDR can be implemented using the R function *p.adjust* (<https://stat.ethz.ch/R-manual/R-devel/library/stats/html/p.adjust.html>).

Although the Bonferroni correction and BH procedure are popular, they both assume that the individual tests are independent. In the context of phylogenetic structure, which is present *de facto* in microbiome analyses, the hierarchical Benjamini-Hochberg procedure (HBH) [35] can be used to construct a two-stage framework. At the first stage, screening tests are performed on higher-ranked taxonomic groups (e.g., families) and on the second stage on lower-ranked taxonomic groups (e.g., species), with BH procedures applied to each stage [36].

3.8 Sample processing and metabolomics analyses of the supragingival biofilm using the Metabolon DiscoveryHD4™ platform

Metabolon has designed and implemented a global metabolomics platform aimed to overcome known challenges associated with broad-range metabolite profiling. These advanced analytical methodologies allow for the detection of metabolites in all major metabolite classes; an analytical feat due to the vast differences in molecular size, physical and chemical properties, and physiological concentrations across classes. The procedures

described can efficiently detect molecules ranging from small highly polar compounds, like TCA cycle intermediates, to large hydrophobic complex lipids with a single sample aliquot. The Metabolon pipeline currently used is the proprietary DiscoveryHD4™ platform [13–17], including multiple mass spectrometry methods, a large reference library of authenticated metabolite standards and a suite of patented informatics and quality control software. We present laboratory steps to process the supragingival biofilm samples for metabolomics analyses in §3.8.1.

3.8.1 Sample preparation—Keep samples stored at -80°C until needed and then thaw them on ice just prior to extraction. Supragingival biofilm samples collected and stored with a toothpick are extracted using an intact Biopsy Sample extraction method as described below:

- a. Prepare the extraction solvent by adding 40 ml of HPLC grade Methanol (CAS 67–56-1) and 10 mL of ultrapure water to a 50 mL conical tube. Invert several times to mix. [Note: The methanol contains four recovery standards, DL-2-fluorophenylglycine, tridecanoic acid, d6-cholesterol and 4-chlorophenylalanine to allow confirmation of extraction efficiency.]
- b. Add 1 mL of extraction solvent to each pre-labeled sample vial for your sample set.
- c. Deposit each biofilm sample on a tooth pick into the corresponding prelabeled sample vial containing room temperature (RT) extraction solvent (one biopsy per vial). The sample will come off of the toothpick with gentle swirling in the MeOH.
- d. Cap vials and incubate samples in extraction solvent at room temperature, on the benchtop, for a minimum of 4 hours and up to 48 hours. No agitation is necessary. [Note: While samples may be incubated in extraction solvent for 4h to 48h, it is critical that all samples in a study are incubated for the same amount of time.]
- e. Following incubation, remove the toothpick from the extraction solvent using forceps. It is crucial to rinse forceps in a rinse solution (80% MeOH, 20% water) between each toothpick retrieval.
- f. Tighten the lids on the vials and the MeOH extracted metabolites are ready to be analyzed by UPLC.
- g. Take four aliquots of each sample from the methanol extract and dry them.
- h. Reconstitute two aliquots of each sample in 50 μL of 6.5 mM ammonium bicarbonate in water (pH=8) for the negative ion analysis.
- i. Reconstitute two more aliquots of each sample using 50 μL of 0.1% formic acid in water (pH ~3.5) for the positive ion analysis. [Note: Reconstitution solvents contain instrument internal standards to assess instrument performance and to serve as retention index markers for chromatographic alignment].

3.8.2 Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectroscopy (UPLC-MS/MS).

- a. Conduct UPLC separation using a Waters Acquity UPLC (Waters, Milford, MA).
- b. Use the mobile phase consisting of 0.1% formic acid in water (A) and 0.1% formic acid in methanol (B) for reverse-phase (RP) positive ion analysis.
- c. Use the mobile phase consisting of 6.5 mM ammonium bicarbonate in water, pH 8 (A) and 6.5 mM ammonium bicarbonate in 95% methanol/ 5% water for the reverse-phase negative ion analysis. Use a sample injection volume of 5 μ L and a 2 \times needle loop overfill. Use a separate acid and base-dedicated 2.1 mm \times 100 mm Waters BEH C18 1.7 μ m columns held at 40°C for separations.
- d. Use the mobile phase consisting of 10 mM ammonium formate in 15% water, 5% methanol, 80% acetonitrile (effective pH 10.16 with NH₄OH) (A) and 10 mM ammonium formate in 50% water, 50% acetonitrile (effective pH 10.60 with NH₄OH) (B) for hydrophilic interaction liquid chromatography (HILIC). Use the same sample injection volume as in the RP method. The stationary phase consists of a 2.1 mm \times 150 mm Waters BEH Amide 1.7 μ m column held at 40°C.
- e. Conduct MS analysis alternating between MS and data-dependent MS scans using dynamic exclusion. The scan range varies slightly between methods but covers 70–1000 m/z.
- f. Archive the raw data files before carrying forward to analysis. An informatics pipeline is described in §3.8.3.

3.8.3 The Metabolon® informatics pipeline for data extraction, compound identification, quality control (QC), normalization and interpretation.—

The informatics system comprises a Laboratory Information Management System (LIMS), data extraction and peak-identification software, data processing tools for QC and compound identification, and libraries for interpretation and visualization tools. Execute the informatics pipeline as follows:

- a. Extract, peak-identify and QC process the raw data using the Metabolon® software pipeline on a research computing server.
- b. Identify compounds by comparison to library entries of purified standards or recurrent unknown entities, based on authenticated standards of retention time/index (RI), mass to charge ratio (m/z), and chromatographic data (including MS/MS spectral data) on all molecules present in the library.
- c. Identify biochemical compounds based on three criteria: 1. retention index within a narrow RI window of the proposed identification, 2. accurate mass match to the library \pm 10 ppm, and 3. the MS/MS forward and reverse scores between the experimental data and authenticated standards. The MS/MS scores are derived based on a comparison of the ions present in the experimental spectrum to the ions present in the library spectrum.

- d. Perform QC procedures (e.g., checks of consistency of peak identification between samples and library matches for each compound) as a means of improving identification of true chemical entities and removing system artifacts, miss-assignments and background noise.
- e. Quantify peaks for each metabolite using area-under-the-curve.
- f. Include a data normalization step to correct variation resulting from instrument inter-day tuning differences for studies spanning multiple days. Essentially, correct each compound in run-day blocks by registering the medians to equal one (1.00) and normalizing each data point proportionately.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by a grant from the National Institutes of Health, National Institute of Dental and Craniofacial Research, U01-DE025046. DS is supported by the Swedish Research Council (4.1-2016-00416). The Microbiome Core is supported in part by the NIH/National Institute of Diabetes and Digestive and Kidney Diseases grant P30 DK34987.

References

1. Nascimento MM, Zaura E, Mira A, Takahashi N, Ten Cate JM. Second Era of OMICS in Caries Research: Moving Past the Phase of Disillusionment. *J Dent Res* 2017 7;96(7):733–740. [PubMed: 28384412]
2. Divaris K Predicting Dental Caries Outcomes in Children: A “Risky” Concept. *J Dent Res* 2016 3; 95(3):248–54. [PubMed: 26647391]
3. Ballantine J, Carlson JC, Zandona A, Agler CS, Zeldin LP, Rozier G, Roberts MW, Basta PV, Luo J, Antonio-Obese ME, McNeil D, Weyant R, Crout RJ, Slayton R, Levy S, Shaffer JR, Marazita ML, North KE, Divaris K. Exploring the genomic basis of early childhood caries: a pilot study. *Int J Paed Dentistry* doi: 10.1111/ipd.12344
4. Kilian M, Chapple IL, Hannig M, Marsh PD, Meuric V, Pedersen AM, Tonetti MS, Wade WG, Zaura E. The oral microbiome - an update for oral healthcare professionals. *Br Dent J* 2016 11 18;221(10):657–666. [PubMed: 27857087]
5. Nyvad B, Crielaard W, Mira A, Takahashi N, Beighton D. Dental caries from a molecular microbiological perspective. *Caries Res* 2013;47(2):89–102. [PubMed: 23207320]
6. Featherstone JD. The continuum of dental caries--evidence for a dynamic disease process. *J Dent Res* 2004;83 Spec No C:C39–42. [PubMed: 15286120]
7. Drury TF, Horowitz AM, Ismail AI, Maertens MP, Rozier RG, Selwitz RH. Diagnosing and reporting early childhood caries for research purposes. A report of a workshop sponsored by the National Institute of Dental and Craniofacial Research, the Health Resources and Services Administration, and the Health Care Financing Administration. *J Public Health Dent* 1999 Summer; 59(3):192–197. [PubMed: 10649591]
8. Divaris K Precision Dentistry in Early Childhood: The Central Role of Genomics. *Dent Clin North Am* 2017 7; 61(3):619–625. [PubMed: 28577641]
9. Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, Bork P, Hugenholtz P, Rubin EM. Comparative metagenomics of microbial communities. *Science* 2005 4 22;308(5721):554–7. [PubMed: 15845853]
10. Gilbert JA, Hughes M. Gene expression profiling: metatranscriptomics. *Methods Mol Biol* 2011;733:195–205. [PubMed: 21431772]

11. Holmes E, Wilson ID, Nicholson JK. Metabolic phenotyping in health and disease. *Cell* 2008 9 5;134(5):714–7. [PubMed: 18775301]
12. Divaris K, Roach J, Basta PV, Ferreira Zandona AG, Ginnis J, Meyer BD, Hu S, Simancas-Pallares MA, Butz N, Azcarate-Peril MA. Metagenomics of early childhood oral health and early childhood caries. *J Dent Res* 2018;97 (Spec Iss A): 2412231 (CADR/AADR).
13. Evans AM, Bridgewater BR, Liu Q, Mitchell MW, Robinson RJ, Dai H, Stewart SJ, DeHaven CD, Miller LA. High resolution mass spectrometry improves data quantity and quality as compared to unit mass resolution mass spectrometry in high-throughput profiling metabolomics. *Metabolomics* 2014 4 1;4(2):1.
14. Evans AM, Mitchell MW, Dai H, DeHaven CD. Categorizing Ion-Features in Liquid Chromatography. *Mass Spectrometry Metabolomics Data*. *Metabolomics* 2012;2(110):2153–0769.
15. DeHaven CD, Evans AM, Dai H, Lawton KA. Software techniques for enabling high-throughput analysis of metabolomic datasets. In *Metabolomics 2012 InTech*.
16. DeHaven CD, Evans AM, Dai H, Lawton KA. Organization of GC/MS and LC/MS metabolomics data into chemical libraries. *Journal of cheminformatics* 2010 12 1;2(1):9. [PubMed: 20955607]
17. Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Analytical chemistry* 2009 7 22;81(16):6656–67. [PubMed: 19624122]
18. Bcl2Fastq 2.20 (2017). Illumina, Inc San Diego, CA, USA.
19. FastQC 0.11.5 (2016). Babraham Institute Cambridge, UK.
20. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012 3 4;9(4): 357–9. [PubMed: 22388286]
21. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 2016 10 18;4:e2584. [PubMed: 27781170]
22. Abubucker S, Segata N, Goll J, Schubert AM, Izard J, Cantarel BL, Rodriguez-Mueller B, Zucker J, Thiagarajan M, Henrissat B, White O, Kelley ST, Methé B, Schloss PD, Gevers D, Mitreva M, Huttenhower C. Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput Biol* 2012;8(6):e1002358. [PubMed: 22719234]
23. Hong C, Manimaran S, Shen Y, Perez-Rogers JF, Byrd AL, Castro-Nallar E, Crandall KA, Johnson WE. PathoScope 2.0: a complete computational framework for strain identification in environmental or clinical sequencing samples. *Microbiome* 2014 9 5;2:33. [PubMed: 25225611]
24. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. Metagenomic biomarker discovery and explanation. *Genome Biol* 2011 6 24;12(6):R60. [PubMed: 21702898]
25. Fernandes AD, Macklaim JM, Linn TG, Reid G, Gloor GB. ANOVA-like differential expression (ALDEx) analysis for mixed population RNA-Seq. *PLoS One* 2013 7 2;8(7):e67019. [PubMed: 23843979]
26. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* 2012 6 10;9(8):811–4. [PubMed: 22688413]
27. Ghosh S, Chan CK. Analysis of RNA-Seq Data Using TopHat and Cufflinks. *Methods Mol Biol* 2016;1374:339–61. [PubMed: 26519415]
28. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, Slichter CK, Miller HW, McElrath MJ, Prlic M, Linsley PS. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome biology* 2015 12;16(1):278. [PubMed: 26653891]
29. Peng X, Li G, Liu Z. Zero-inflated beta regression for differential abundance analysis with metagenomics data. *Journal of Computational Biology* 2016 2 1;23(2):102–10. [PubMed: 26675626]
30. Chai H, Jiang H, Lin L, Liu L. A marginalized two-part Beta regression model for microbiome compositional data. *PLoS computational biology* 2018 7 23;14(7):e1006329. [PubMed: 30036363]
31. Wagner BD, Robertson CE, Harris JK. Application of two-part statistics for comparison of sequence variant counts. *PloS one* 2011 5 23;6(5):e20296. [PubMed: 21629788]

32. Hu MC, Pavlicova M, Nunes EV. Zero-inflated and hurdle models of count data with extra zeros: examples from an HIV-risk reduction intervention trial. *The American journal of drug and alcohol abuse* 2011 9 1;37(5):367–75. [PubMed: 21854279]
33. Preisser JS, Das K, Long DL, Divaris K. Marginalized zero-inflated negative binomial regression with application to dental caries. *Statistics in medicine* 2016 5 10;35(10):1722–35. [PubMed: 26568034]
34. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association* 1952 12 1;47(260):583–621.
35. Yekutieli D Hierarchical false discovery rate–controlling methodology. *Journal of the American Statistical Association* 2008 3 1;103(481):309–16.38 Hu et al. 2018
36. Hu J, Koh H, He L, Liu M, Blaser MJ, Li H. A two-stage microbial association mapping framework with advanced FDR control. *Microbiome* 2018 12;6(1):131. [PubMed: 30045760]
37. Segata N, Boernigen D, Tickle TL, Morgan XC, Garrett WS, Huttenhower C. Computational meta’omics for microbial community studies. *Mol Syst Biol* 2013 5 14;9:666. [PubMed: 23670539]

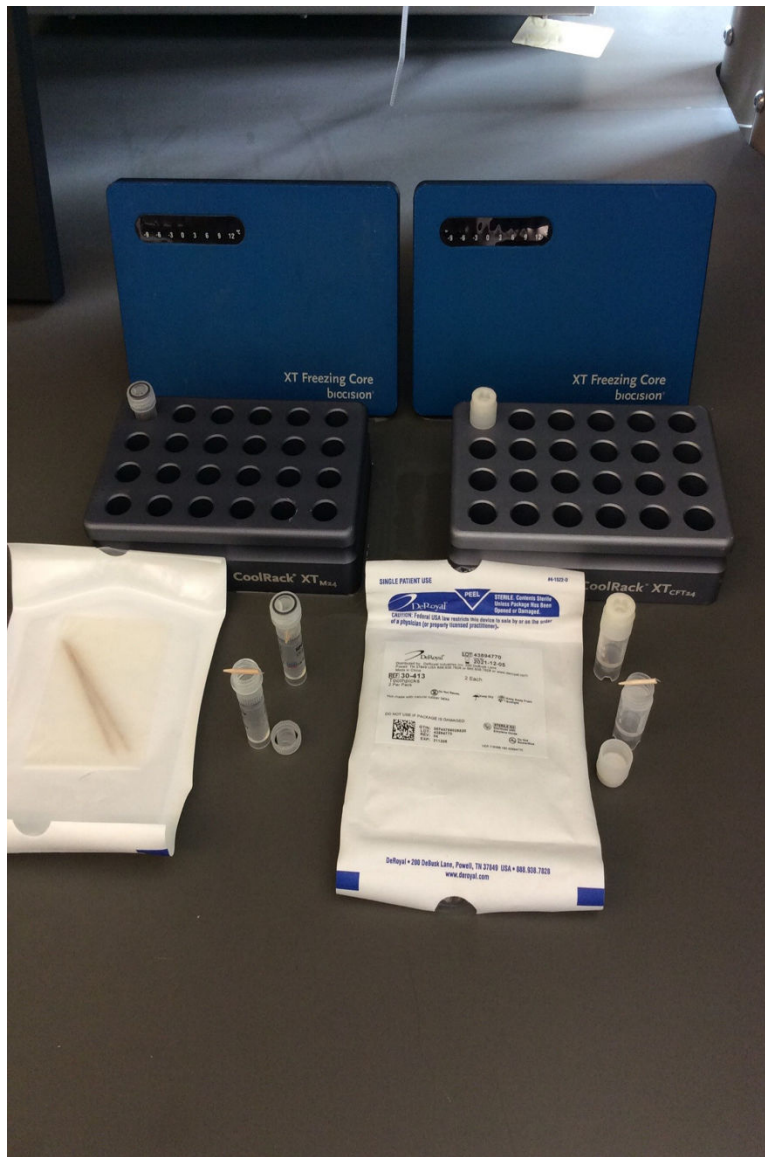


Figure 1. Illustration of the supragingival plaque collection equipment and supplies: Two packs of sterile toothpicks (pack of two toothpicks), presented from each other side (DeRoyal. Box of 50. Product No. 30–413); one toothpick is used for the collect.



Figure 2.
An illustration of a sample stored in portable freezing storage container (BioCision CoolBox XT CryoTube 24 Workstation, BioCision Catalogue No. BSC-575).