Structural and Functional Characteristics of the Microbiome in Deep-Dentin Caries

G. Liu, C. Wu, W.R. Abrams, and Y. Li

APPENDICES

I. SUPPLEMENTARY MATERIALS AND METHODS:

(1) Contamination control for sample preparation. In this study, we followed the standard procedure protocols (Knight *et al.* 2018; National Institute of Health 2010) and took several measures to prevent potential sample contamination, including the following.

a. A rubber dam was used to isolate the tooth from saliva contamination.
b. Sterilized equipment and tools (tubes, excavators, and fresh reagents) were used each time when collecting samples.
c. After collection, all of the samples were placed in sterilized test tubes, immediately transferred on ice to a lab, and stored at -70°C.
d. The samples were packed with dry ice and transferred overnight to the sequencing facility.
e. TE buffer was used as a blank control for DNA isolation and quantitative and qualitative analyses. Successful purification and amplification of targeted DNA samples were examined and confirmed by agarose gel electrophoresis before setting up for sequencing.
f. In the process of sample library construction, a library of negative control samples (blank controls) was built in and then sequenced with the clinical samples. During the process of DNA extraction and building the database, the consumables and reagents used were inspected for possible contamination. No sample contamination was found in this study.
g. The study used second-generation metagenomic sequencing technology for human samples. The technology allows us to build a single library from a single sample. When constructing a library, a different linker index sequence is added to each library. The index is the basis for distinguishing between different samples.

(2) DNA library construction. Genomic DNA was isolated from each carious lesion sample using a NucleoSpin Soil Kit (Macherey-Nagel Inc. Bethlehem, PA) following a previously published protocol (Lazarevic *et al.* 2013). DNA libraries were constructed using the

Illumina TruSeq DNA Sample Prep Kit (Illumina Inc. San Diego, CA) according to the manufacturer's instructions with modifications. Briefly, DNA samples were sheared into 350 bp fragments by a Covaris ultrasonicator. The overhangs resulting from fragmentation were converted into blunt ends by using a T4 DNA polymerase, Klenow fragment, and T4 polynucleotide kinase. After adding an A (adenine) base to the 3' end of the blunt phosphorylated DNA fragments, adaptors were ligated to both ends of the DNA fragments, and the short fragments were removed with Ampure beads. The Agilent 2100 Bioanalyzer and ABI StepOnePlus Real-Time PCR System were used to qualify and quantify the sample libraries. The qualified libraries were then sequenced using an Illumina HiSeqTM platform at the Novogene Bioinformatics Institute, Tianjin, China.

(3) Dual assembly processing. The raw reads obtained from metagenome sequencing were processed and provided by Novogene using Readfq (v8, https://github.com/cjfields/readfq). Clean data were obtained by removing (a) low-quality reads (a threshold value ≤ 38 and more than a length of 40 bp), (b) reads with Ns (unknown bases) more than 10 bp, (c) reads overlapping with adaptor sequences (≥ 15 bp), and (d) filtering host-originated reads using the short oligonucleotide analysis software SOAPaligner/soap2 (Li *et al.* 2009). First, the qualified clean data were assembled using SOAPdenovo2 software (v2.04, http://soap.genomics.org.cn/soapdenovo.html). The parameters employed were as follows: -d 1, -M 3, -R, -u, -F, and -K-55 (Qin *et al.* 2014), and scaffold-level results were obtained. Subsequently, the scaffold results were interrupted from the N-connection to form scaftig-level results. For each sample, the clean data were BLASTed to the scaftigs set using SOAPaligner software again. The parameters used were as follows: % identity values greater than or equal to 90%, -m 200-x 400 (Qin *et al.* 2014). To explore the rare species in the samples, the second assembly was conducted with the mixed paired-end reads that were not used in the first assembly, using the same software SOAPdenovo2 and parameters. After filtering the fragments shorter than 500 bp, the scaftig sets generated from both assembly results were used for the subsequent statistical analysis and gene prediction.

(4) Software used in metagenomic analysis:

a. Readfq, v8: https://github.com/cjfields/readfq

b. SOAPaligner/soap2, v2.21: http://soap.genomics.org.cn/soapaligner.html

   * parameter settings for filtering host-origin data: identity ≥ 90%, -l 30, -v 7, -M 4, -m 200, -x 400

   * parameter settings for scaftig blasting: identity ≥ 90%, -m 200 and -x 400

   * parameter settings for IGC blasting: identity ≥95%, -m 200, and -x 400

c. SOAPdenovo2, v2.04: ftp://public.genomics.org.cn/BGI/SOAPdenovo2

   parameter settings: -d 1, -M 3, -R, -u, -F, and -K-55

d. MetaGeneMark, v2.10: http://topaz.gatech.edu/GeneMark/

   Parameter setting: default. (Zhu *et al.* 2010)

e. CD-HIT, v4.5.8: http://www.bioinformatics.org/cd-hit

   parameter settings: -c 0.95, -G 0, -aS 0.9, -g 1, -d 0. (Fu *et al.* 2012)

f. DIAMOND, v0.7.9: https://github.com/bbuchfink/diamond/

   parameters setting: blastp, evalue ≤ 1e-5. (Buchfink *et al.* 2015)

g The R Project for Statistical Computing: https://www.r-project.org/

II. SUPPLEMENTARY TABLES

Appendix Table 1. Correlation between the total number of genes acquired between superficial layers and deep layers obtained from the same carious lesions.

	1s	1d	2s	2d	3s	3d	4s	4d	5s	5d	6s	6d	7s	7d	8s	8d
1s	1	0.75	0.305	0.322	0.078	0.147	0.161	0.187	-0.001	0.039	0.109	0.139	0.157	0.163	0.236	0.312
1d	0.75	1	0.279	0.295	0.088	0.147	0.117	0.145	0.003	0.048	0.116	0.146	0.157	0.172	0.238	0.311
2s	0.305	0.279	1	0.711	0.298	0.322	0.191	0.167	0.216	0.254	0.326	0.326	-0.01	-0.004	0.328	0.282
2d	0.322	0.295	0.711	1	0.242	0.26	0.188	0.177	0.16	0.203	0.255	0.283	0.037	0.052	0.277	0.27
3s	0.078	0.088	0.298	0.242	1	0.846	-0.077	-0.124	0.456	0.376	0.435	0.334	0.039	0.012	0.454	0.337
3d	0.147	0.147	0.322	0.26	0.846	1	-0.054	-0.099	0.403	0.363	0.393	0.32	0.084	0.054	0.431	0.358
4s	0.161	0.117	0.191	0.188	-0.077	-0.054	1	0.782	-0.129	-0.093	-0.016	0.055	-0.165	-0.147	-0.041	-0.029
4d	0.187	0.145	0.167	0.177	-0.124	-0.099	0.782	1	-0.138	-0.071	-0.042	0.043	-0.178	-0.151	-0.069	-0.041
5s	-0.001	0.003	0.216	0.16	0.456	0.403	-0.129	-0.138	1	0.835	0.345	0.248	0.084	0.058	0.285	0.214
5d	0.039	0.048	0.254	0.203	0.376	0.363	-0.093	-0.071	0.835	1	0.354	0.285	0.097	0.088	0.254	0.237
6s	0.109	0.116	0.326	0.255	0.435	0.393	-0.016	-0.042	0.345	0.354	1	0.728	0.02	0.024	0.397	0.35
6d	0.139	0.146	0.326	0.283	0.334	0.32	0.055	0.043	0.248	0.285	0.728	1	-0.002	0.002	0.302	0.299
7s	0.157	0.157	-0.01	0.037	0.039	0.084	-0.165	-0.178	0.084	0.097	0.02	-0.002	1	0.818	0.19	0.292
7d	0.163	0.172	-0.004	0.052	0.012	0.054	-0.147	-0.151	0.058	0.088	0.024	0.002	0.818	1	0.179	0.299
8s	0.236	0.238	0.328	0.277	0.454	0.431	-0.041	-0.069	0.285	0.254	0.397	0.302	0.19	0.179	1	0.774
8d	0.312	0.311	0.282	0.27	0.337	0.358	-0.029	-0.041	0.214	0.237	0.35	0.299	0.292	0.299	0.774	1

The top row and left column represent the names of the caries samples. The numbers 1 to 8 represent the sample ID; “s” represents the superficial layer; “d” represents the deep layer. The pairwise comparisons between the samples showed higher degrees of similarities of gene profiles between the superficial layer and the deep layer of the same carious lesions.

Appendix Table 2. The number of annotated categories per sample.

Sample ID	Annotated		
	Phylum	Order	Genus
1s	27	81	399
1d	28	83	407
2s	47	92	573
2d	41	90	511
3s	91	133	825
3d	89	127	825
4s	21	54	261
4d	19	58	275
5s	88	139	959
5d	85	135	922
6s	87	131	802
6d	77	118	685
7s	94	143	907
7d	92	143	908
8s	88	134	853
8d	88	136	830

Appendix Table 3. A brief review of microbiota research of dentin caries in permanent teeth.

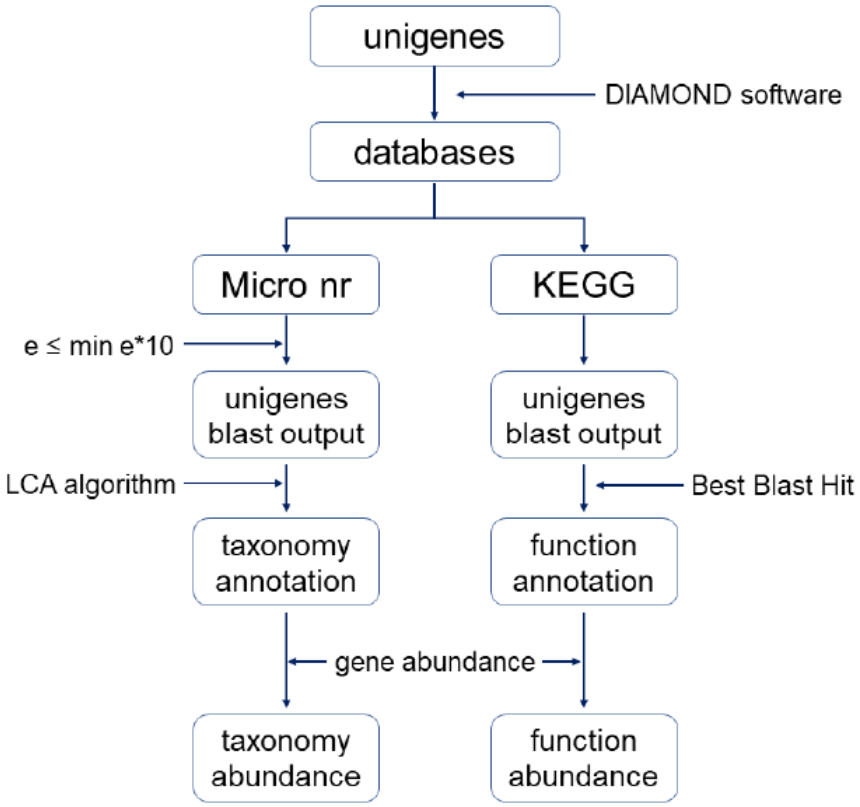
Published year & Authors	Subjects no.	Samples no.	Patient age (yr)	Methods	Taxa no.	Predominant taxa
2004 Munson *et al*. (Munson *et al.* 2004)	5	10	24~79	① 16S rRNA gene sequencing ② Anaerobic cultivation	95	The predominant taxa by anaerobic cultivation: *Propionibacterium sp*. 18%, Olsenella profusa 14%, and Lactobacillus rhamnosus 8%. The predominant taxa by molecular analysis: *Streptococcus mutans* 16%, *Lactobacillus gasseri/johnsonii* 13%, and *Lactobacillus rhamnosus* 8%.
2005 Chhour *et al*. (Chhour *et al.* 2005)	10	10	32~65	16S rRNA gene sequencing	75	*Lactobacill*: 50% of the species; *Prevotellae*: 15% of the species; Other predominant taxa included: *Selenomonas* spp., *Dialister* spp., *Fusobacterium nucleatum*, *Eubacterium* spp., members of the *Lachnospiraceae family*, *Olsenella* spp., *Bififidobacterium spp.*, *Propionibacterium* spp., and *Pseudoramibacter alactolyticus*.
2008 Aas *et al*. (Aas *et al.* 2008)	36	① 72 ② 10	2~21 (including primary teeth)	① Reverse-capture checkerboard assay ② Clone library analysis	① probes targeting 110 taxa ② 76 taxa	①: The microflora of deep-dentin lesions was dominated by *S. mutans*, *Lactobacillus* spp., *Propionibacterium* spp. strain FMA5, and *Atopobium genomospecies* C1. ②: With *S. mutans*, *Atopobium genomospecies* C1 or *Lactobacillus* spp., were present at significantly higher levels; without *S. mutans*, *Lactobacillus* spp., *Bifidobacterium dentium* and low-pH non-*S. mutans streptococci* were predominant.
2010 Gross *et al*. (Gross *et al.* 2010)	21	21	7~16	16S rRNA gene Sanger sequencing	< 144 species	The predominant genera were *Lactobacillus*, *Propionibacterium* FMA5.
2011 Lima *et al*. (Lima *et al.* 2011)	27	81	7~14	Reverse-capture checkerboard assay	28 bacterial taxa by using probes	The most frequently detected taxa were: *Atopobium genomospecies* C1, *F. nucleatum*, *Lactobacillus casei*, *Veillonella* spp. and *Lactobacillus fermentum*.
2014	32	44	4~76	① Pyrosequencing	79 genera	①The predominant genera were *Lactobacillus*, *Propionibacterium*, *Atopobium*, *Streptococcus*, *Actinomyces*, *Prevotella*, *Olsenella*,

Obata *et al*. (Obata *et al.* 2014)			(including primary teeth)	② Clone library analysis		*Veillonella*, *Pseudoramibacter*, *Mitsuokella*, *Scardovia.* ②The predominant genera were *Atopobium* and *Propionibacterium*.
2014 Schulze-Schweifiing *et al*. (Schulze-Schweifing *et al.* 2014)	6	6	22~35	① Anaerobic cultivation ② 16S rRNA gene sequencing ③ Pyrosequencing	①+② 229 genera ③ 264 genera	①+②: The predominant genera were *Prevotella*, *Lactobacillus*, *Selenomonas*, and *Streptococcus.* The five most abundant species: *Lactobacillus gasseri*, *Prevotella denticola*, *Alloprevotella tannerae*, *S. mutans* and *Streptococcus* spp. HOT 070. ③: The predominant genera were *Prevotella*, *Lactobacillus*, *Streptococcus*, *P. alactolyticus*, and *Fusobacteria.* The five most abundant species were *Lactobacillus gasseri*, a group of unclassified streptococci, *S. mutans*, *P. alactolyticus*, and *P. denticola*.
2015 Rôças *et al*. (Rocas *et al.* 2015)	30	30	12~33	Reverse-capture checkerboard assay	33 bacterial taxa by using probes	The most frequently detected taxa: *Atopobium genomospecie*s C1, *Pseudoramibacter alactolyticus*, *Streptococcus* spp., *Streptococcus mutan*, *Parvimonas micra*, *Fusobacterium nucleatum*, and *Veillonella* spp.
2016 Rôças *et al*. (Rocas *et al.* 2016)	10	10	16~60	16S rRNA gene sequencing	101 genera	The predominant genera were *Lactobacillus*, *Olsenella*, *Pseudoramibacter* and *Streptococcus.*
2019 Zheng *et al*. (Zheng *et al.* 2019)	75	75	12~60	16S rRNA gene sequencing	1132 genera	The top 10 genera were *Lactobacillus*, *Propionibacterium*, *Olsenella*, *Acinetobacter*, *Streptococcus*, *Actinomyces*, *Pseudoramibacter*, *Prevotella*-7, *Escherichia-Shigella* and *Neisseria*.

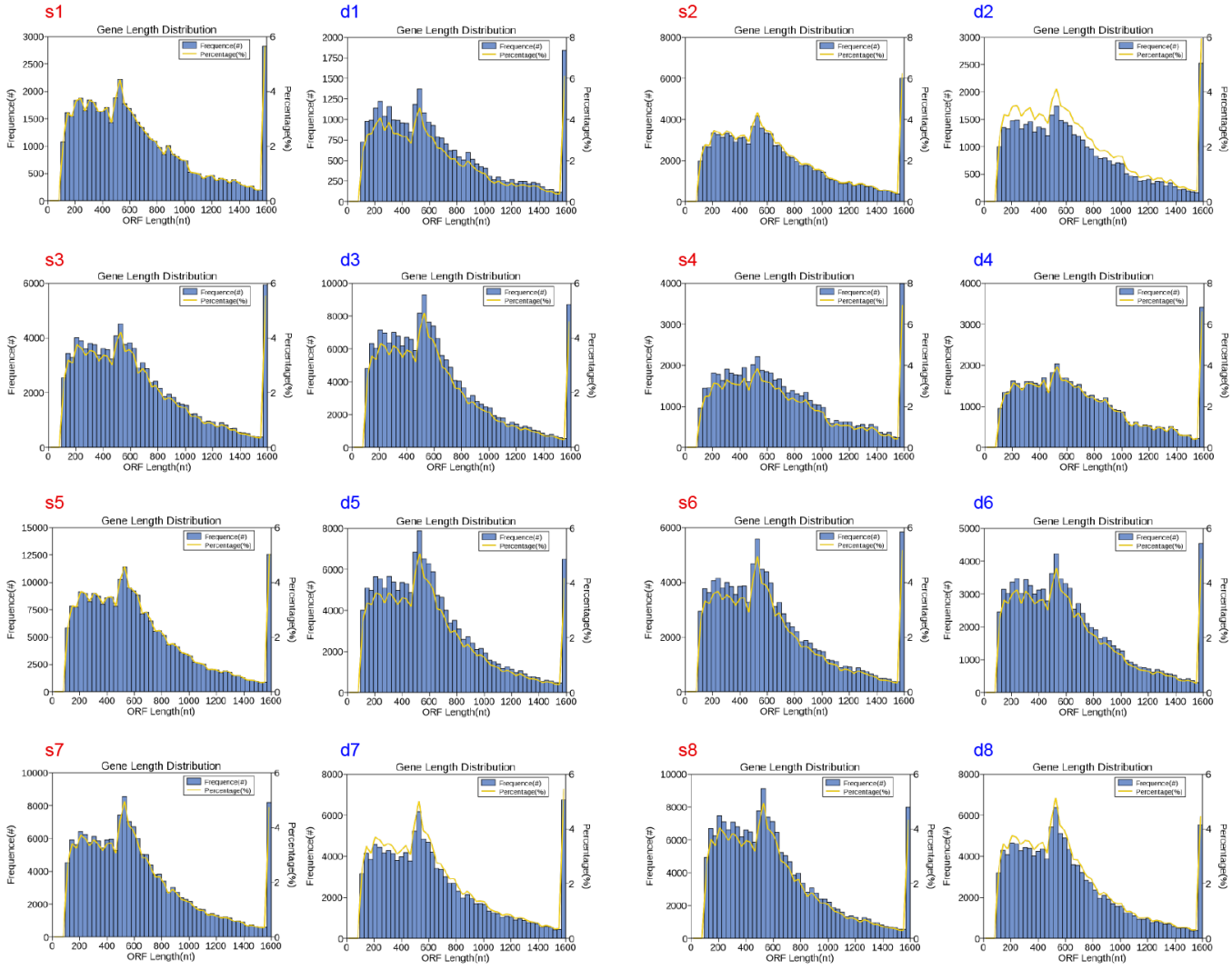
In preparation of our study of "The structural and functional characteristics of the microbiome in deep-dentin caries," we conducted a brief review on microbiota research of dentin caries in permanent teeth using different molecular methods. The main objective of this review was to gain current knowledge of this field. The table presents the summary information on the number and types of microorganisms identified by different investigators. Compared with these previously published reports, we found certain degrees of similarities and disagreements in terms of microbial diversity and abundance detected for deep-dentin caries using new metagenomics sequencing technology.

III. SUPPLEMENTARY FIGURES

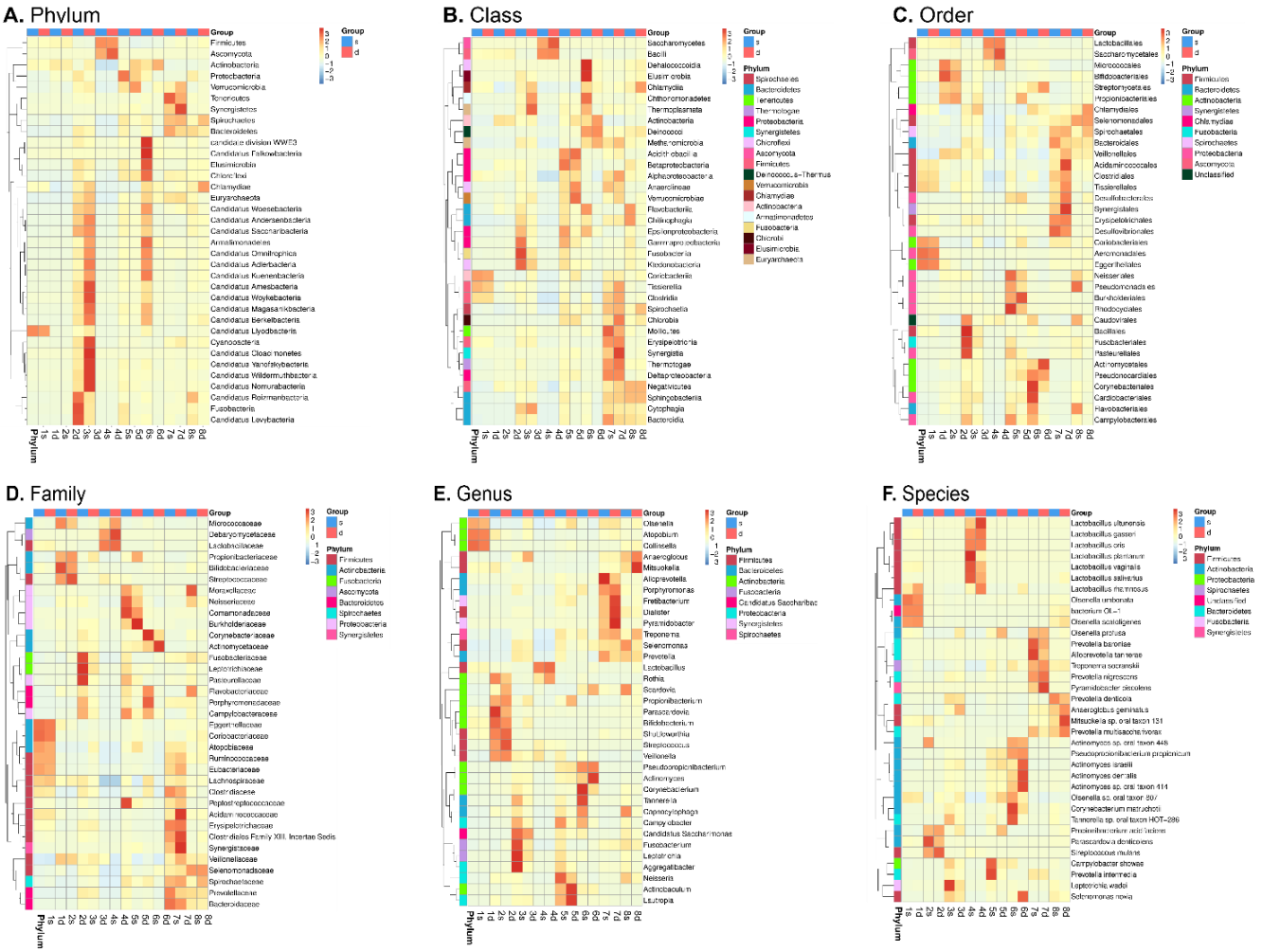
Appendix Figure 1. Deep carious lesion diagnosis. An intraoral image of the first molar with a carious lesion and a panoramic scanning dental X-ray of the upper and lower jaw of the patient. The dental panoramic X-ray shows a deep carious lesion in the first permanent molar (yellow arrow).



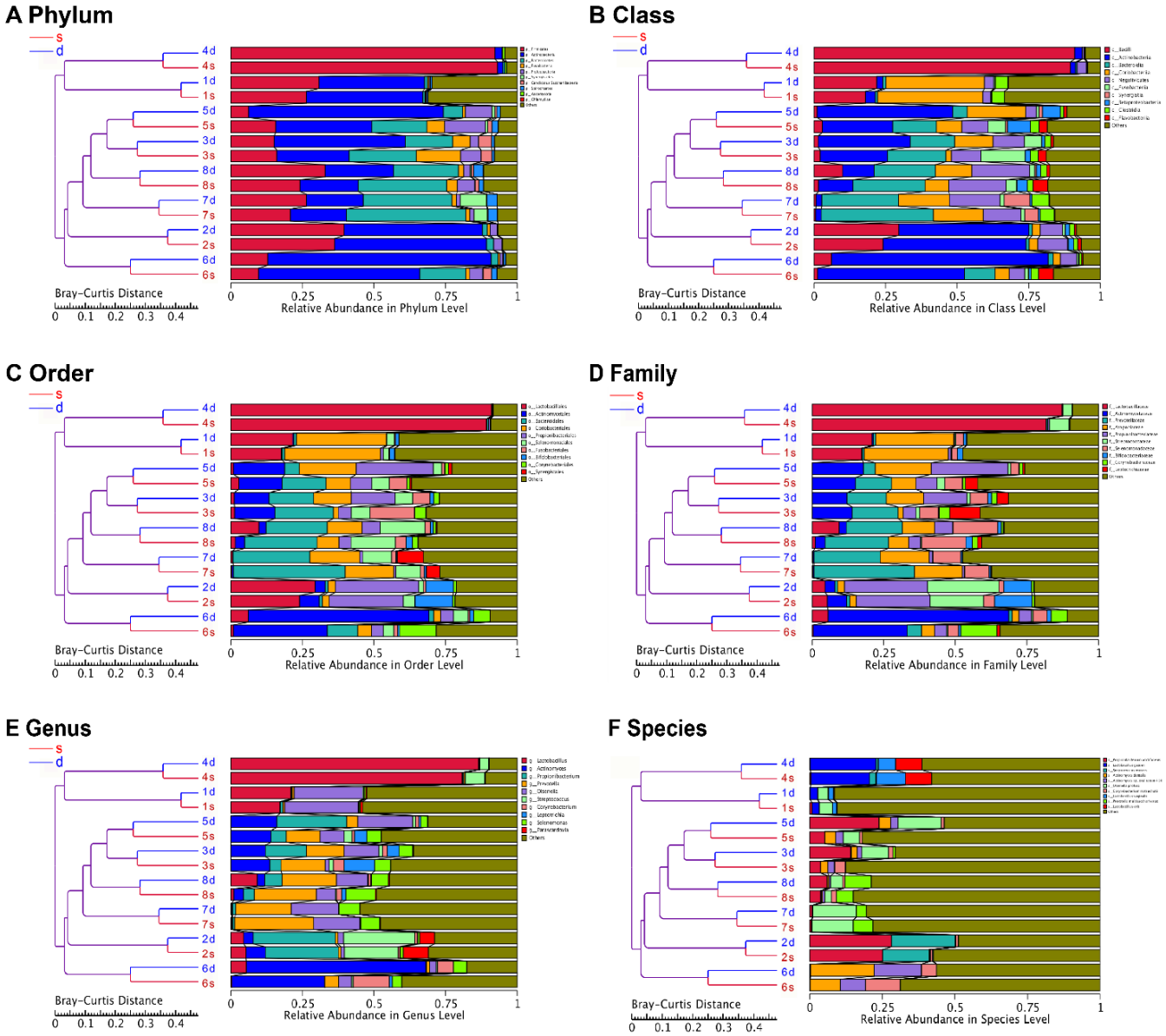
Appendix Figure 2. Workflow used for the taxonomic, functional annotation, and abundance analysis.



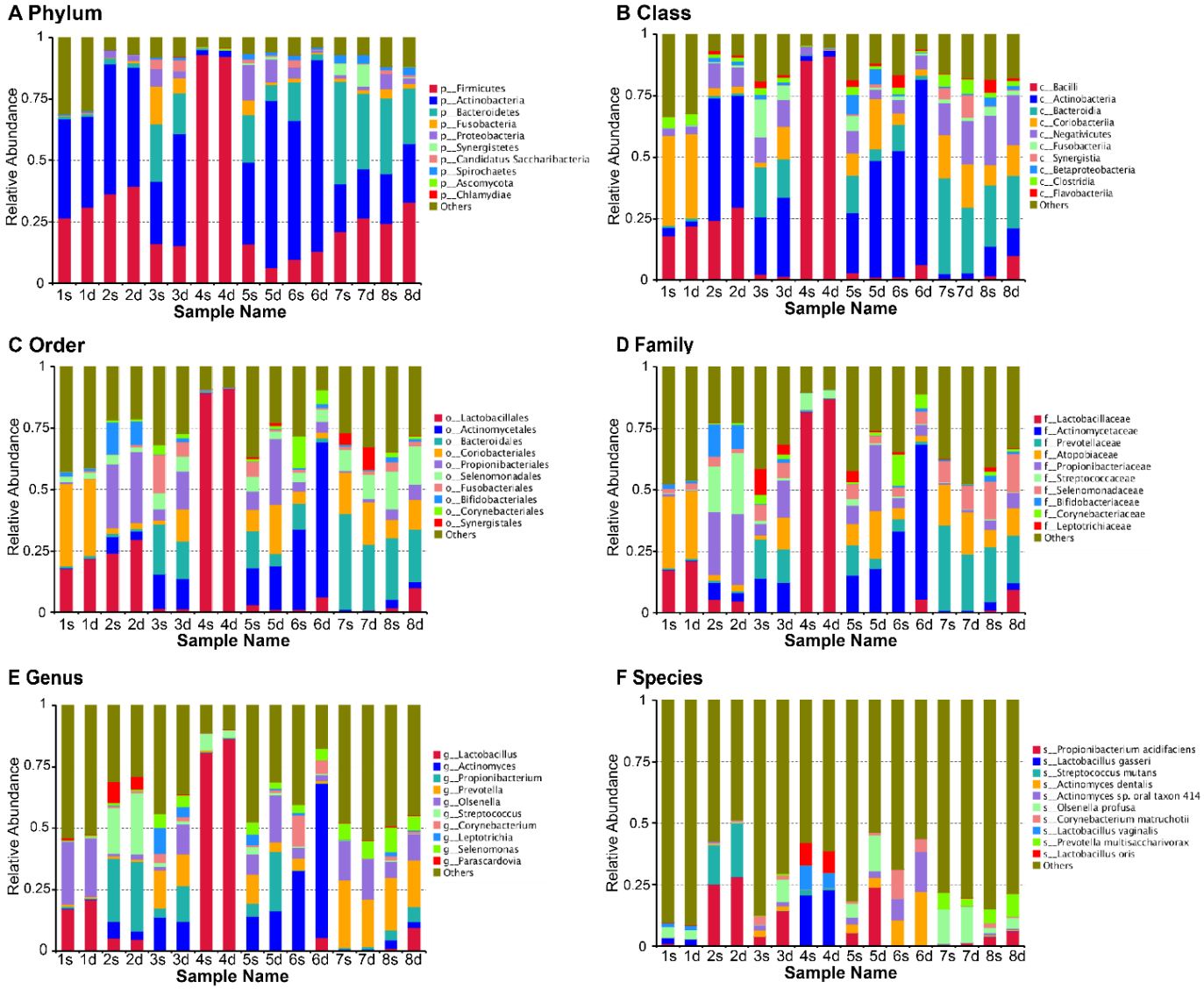
Appendix Figure 3. The distribution of the length of unigenes for each sample. The left axis represents the number of unigenes in the gene catalog, and the right axis represents the percentage (%) of the unigenes in the gene catalog. The *x*-axis represents the length of the unigenes in the gene catalog. Differences were found among the eight individuals but not between the two groups of carious samples. s = superficial layer; d = deeper layer.



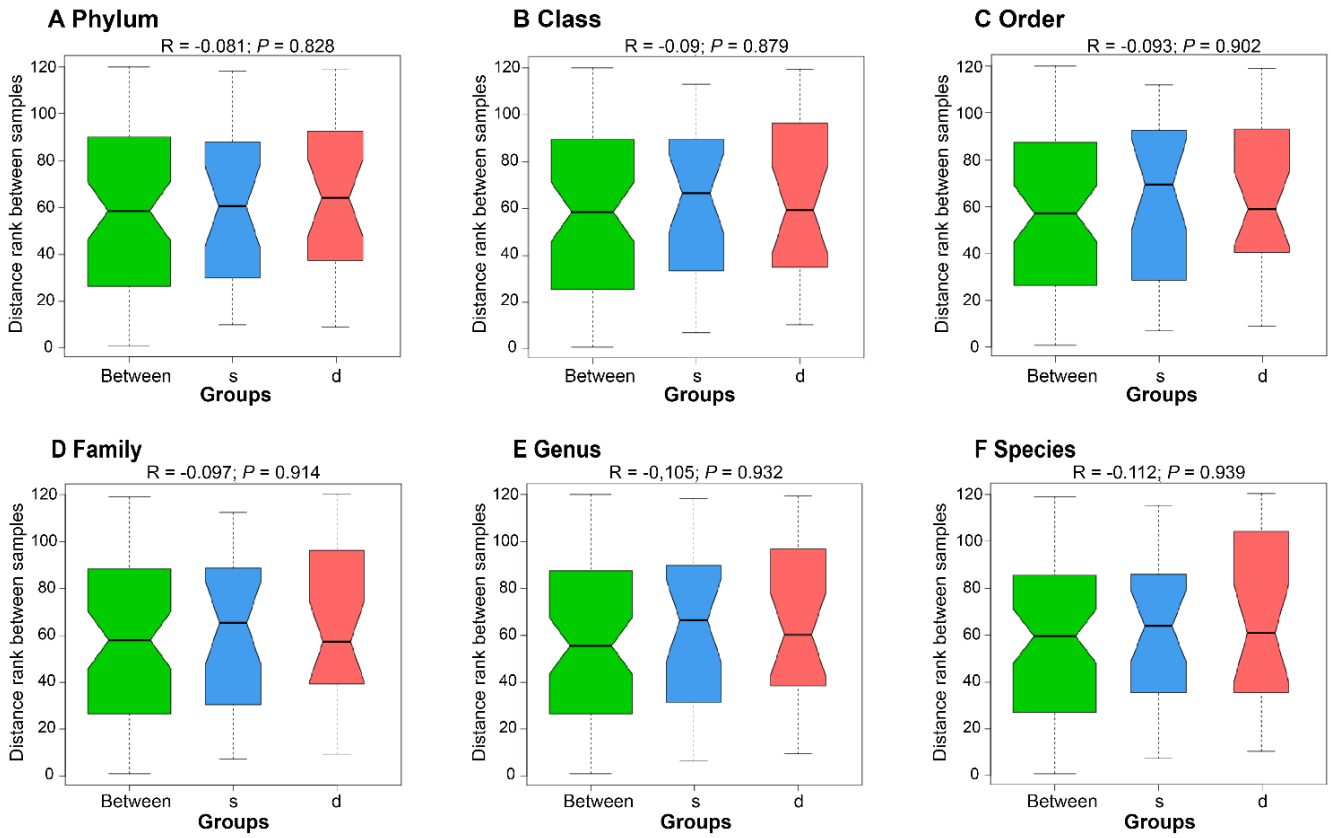
Appendix Figure 4. The relative abundance clustering heat maps of the top 35 microbial members at the six taxonomic levels. The selection was based on the number of unigenes in each sample. The horizontal axis is the sample name; the vertical axis on the right side is the microbial information; the different colors represent the relative abundance. On the left side of the figure is the clustering tree; the value corresponding to the intermediate heat map based on the *z* score (relative abundance of a sample - average relative abundance of all samples)/standard deviation of all samples) was obtained by normalizing the relative abundance of each row of microbial members at the phylum, class, order, family, genus, and species levels. The results show that the abundance of individual microbial members was closely related between the two layers of samples from the same carious lesion.



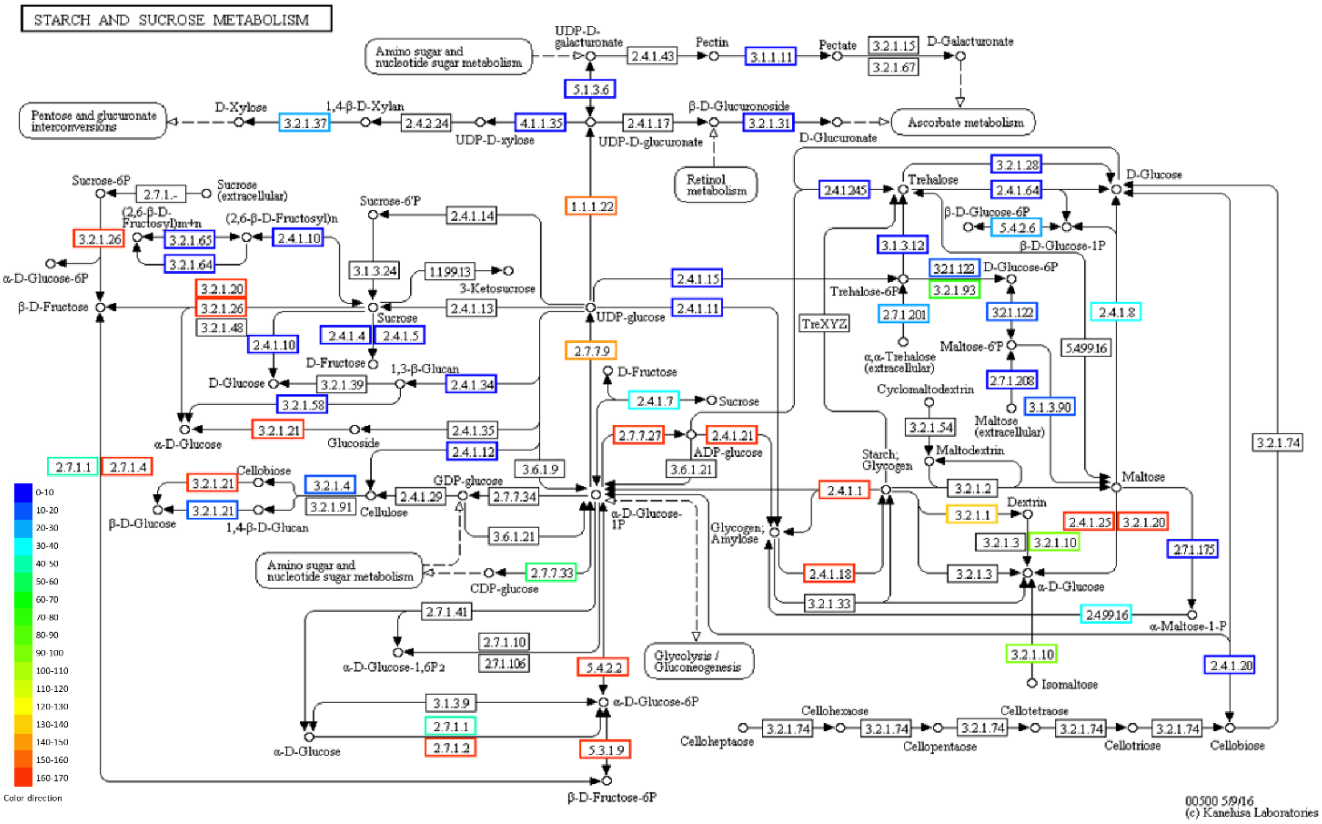
Appendix Figure 5. Cluster analysis of the top 10 microbial taxa at six taxonomic levels. A clustering tree of samples was constructed based on the Bray-Curtis distance to study the similarity between the samples. On the left is the Bray-Curtis distance cluster tree structure, and on the right is the relative abundance distribution of microbial members representing phylum, class, order, family, genus, and species level (A-F). The microbial profiles of the two related layers were consistently clustered together at all levels.



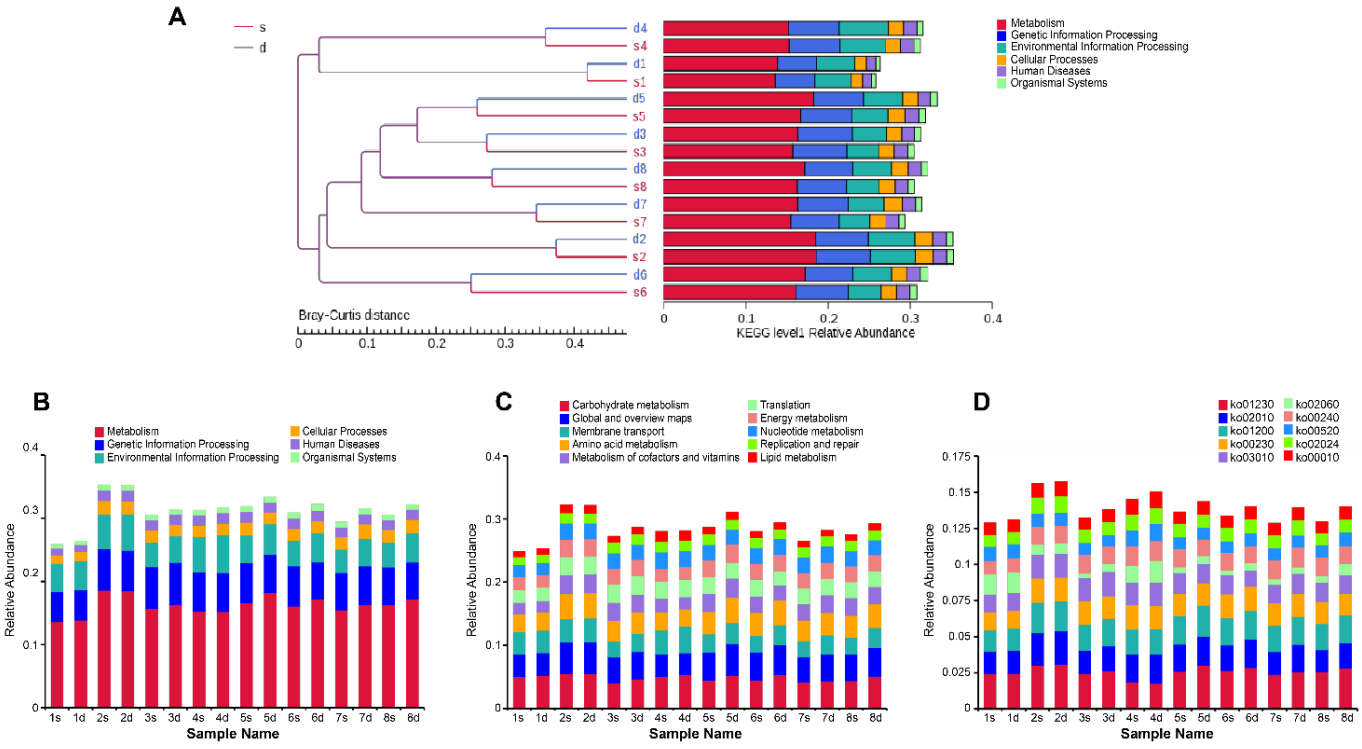
Appendix Figure 6. Microbial distribution of the top 10 bacteria among the 16 samples at six taxonomic levels.



Appendix Figure 7. Analysis of similarity (ANOSIM). The test for differences between the superficial layers and the deeper layers showed no significant differences at all taxonomic levels.



Appendix Figure 8. Schematic representation of the starch and sucrose metabolic pathway annotated from deep dentin carious lesions. The boxes represent enzyme-coding (EC) genes identified in these pathways. The number of annotated genes for an EC is represented by different colors from blue to red representing the gene numbers from 0 to 170. EC for the carbohydrate metabolism pathway, amino acid metabolism, and membrane transport were mostly identified in this study. However, the study observed discordance between the KEGG functional profiles and the abundance of EC genes that needs to be further elucidated.



Appendix Figure 9. Functional relative abundance profile. The microbial functional profile generated by the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa *et al.* 2006; Kanehisa *et al.* 2014). The Bray-Curtis distance matrix was used for cluster analysis between samples, and the cluster results were integrated with the relative abundance of each sample in the first level of the KEGG database (**A**). Comparison of the relative abundance of the top 10 functional genes at level 1 (**B**), level 2 (**C**), and level 3 (**D**) for each sample. The vertical axis represents the relative proportion of annotations to a functional class; the horizontal axis represents the sample name; the functional categories corresponding to each color block are shown in the legend on the right. No significant differences were found between the two layers.

IV. APPENDIX REFERENCES

Aas JA, Griffen AL, Dardis SR, Lee AM, Olsen I, Dewhirst FE, Leys EJ, Paster BJ. 2008. Bacteria of dental caries in primary and permanent teeth in children and young adults. J Clin Microbiol. 46(4):1407-1417.

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using diamond. Nat Methods. 12(1):59-60.

Chhour KL, Nadkarni MA, Byun R, Martin FE, Jacques NA, Hunter N. 2005. Molecular analysis of microbial diversity in advanced caries. J Clin Microbiol. 43(2):843-849.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. Cd-hit: Accelerated for clustering the next-generation sequencing data. Bioinformatics. 28(23):3150-3152.

Gross EL, Leys EJ, Gasparovich SR, Firestone ND, Schwartzbaum JA, Janies DA, Asnani K, Griffen AL. 2010. Bacterial 16s sequence analysis of severe caries in young permanent teeth. J Clin Microbiol. 48(11):4121-4128.

Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M. 2006. From genomics to chemical genomics: New developments in kegg. Nucleic Acids Research. 34:D354-D357.

Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2014. Data, information, knowledge and principle: Back to metabolism in kegg. Nucleic Acids Res. 42(Database issue):D199-205.

Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, Gonzalez A, Kosciolek T, McCall LI, McDonald D et al. 2018. Best practices for analysing microbiomes. Nature Reviews Microbiology. 16(7):410-422.

Lazarevic V, Gaia N, Girard M, Francois P, Schrenzel J. 2013. Comparison of DNA extraction methods in analysis of salivary bacterial communities. PLoS One. 8(7):e67699.

Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J. 2009. Soap2: An improved ultrafast tool for short read alignment. Bioinformatics. 25(15):1966-1967.

Lima KC, Coelho LT, Pinheiro IV, Rocas IN, Siqueira JF, Jr. 2011. Microbiota of dentinal caries as assessed by reverse-capture checkerboard analysis. Caries Res. 45(1):21-30.

Munson MA, Banerjee A, Watson TF, Wade WG. 2004. Molecular analysis of the microflora associated with dental caries. J Clin Microbiol. 42(7):3023-3029.

National Institute of Health. 2010. Manual of procedures for human microbiome project. Core microbiome sampling protocol a In: NIDCR NRaPTLf, editor.: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/GetPdf.cgi?id=phd003190.2.

Obata J, Takeshita T, Shibata Y, Yamanaka W, Unemori M, Akamine A, Yamashita Y. 2014. Identification of the microbiota in carious dentin lesions using 16s rrna gene sequencing. PLoS One. 9(8):e103712.

Qin N, Yang F, Li A, Prifti E, Chen Y, Shao L, Guo J, Le Chatelier E, Yao J, Wu L et al. 2014. Alterations of the human gut microbiome in liver cirrhosis. Nature. 513(7516):59-64.

Rocas IN, Alves FR, Rachid CT, Lima KC, Assuncao IV, Gomes PN, Siqueira JF, Jr. 2016. Microbiome of deep dentinal caries lesions in teeth with symptomatic irreversible pulpitis. PLoS One. 11(5):e0154653.

Rocas IN, Lima KC, Assuncao IV, Gomes PN, Bracks IV, Siqueira JF, Jr. 2015. Advanced caries microbiota in teeth with irreversible pulpitis. J Endod. 41(9):1450-1455.

Schulze-Schweifing K, Banerjee A, Wade WG. 2014. Comparison of bacterial culture and 16s rrna community profiling by clonal analysis and pyrosequencing for the characterization of the dentine caries-associated microbiome. Front Cell Infect Microbiol. 4:164.

Zheng J, Wu Z, Niu K, Xie Y, Hu X, Fu J, Tian D, Fu K, Zhao B, Kong W et al. 2019. Microbiome of deep dentinal caries from reversible pulpitis to irreversible pulpitis. J Endod. 45(3):302-309.

Zhu W, Lomsadze A, Borodovsky M. 2010. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res. 38(12):e132.