Association between Oral *Candida* and Bacteriome in Children with Severe ECC

J. Xiao, A. Grier, R.C. Faustoferri, S. Alzoubi, A.L. Gill, C. Feng, Y. Liu, R.G. Quivey, D.T. Kopycka-Kedzierawski, H. Koo, and S.R. Gill

Appendix

Detailed Methods for *C. albicans* Identification and Bacterial Community Profiling

Isolation and Identification of *C. albicans*

Salivary and plaque samples of 100ul (resupended in 1ml of 0.89% saline, sonicated 3×10sec) were plated onto BBL™ CHROMagar™ Candida (BD, Sparks, MD, USA) within 2 hours of collection and incubated for 48h (37°C) to isolate *C. albicans*. *C. albicans* was identified by colony color and morphology (Odds and Bernaerts, 1994); additional germ tubing tests were conducted from three randomly selected colonies in each sample. Moreover, *C*. *albicans* strains isolated from S-ECC children were typed using an MLST scheme (Bougnoux et al., 2003) for molecular identification. The internal regions of seven housekeeping genes (*AAT1a*, *ACC1*, *ADP1*, *MPIb*, *SYA1*, *VPS13*, and *ZWF1b*) were sequenced. Each strain was characterized by a diploid sequence type (DST) that resulted from the combination of the genotypes obtained at the seven loci. Sanger sequence analysis and merger was performed using CAP3. The reference database used to build the MLST scheme was the *C*. *albicans* library from PubMLST.

Genomic DNA extraction, 16S rRNA sequencing and anslysis

Previously established methods (Grier et al., 2017; Merkley et al., 2015) were used to perform oral microbiome sequencing and related bioinformatics analysis. Total genomic DNA from clinical samples (100ul saliva and 200ul plaque suspension) was extracted using Quick-DNA™ Fecal/Soil Microbe Miniprep Kit (ZymoResearch, Irvine, CA) (Dardas et al., 2014; Merkley et al., 2015; Zhu et al., 2013). 16S ribosomal DNA (rDNA) was amplified with Phusion

High-Fidelity polymerase (Thermo Scientific, Waltham, MA) and dual indexed primers specific to the V1-V3 hypervariable (Fadrosh et al., 2014). Amplicons were pooled and paired-end sequenced on an Illumina MiSeq (Illumina, San Diego, CA) in the University of Rochester Genomics Research Center. Absolute abundance of bacteria in each sample was determined by quantitative real-time PCR (qPCR) using 16s primers, probe and cloned plasmid standards. Sequence reads were assessed for quality using Quantitative Insights into Microbial Ecology (QIIME) software (Caporaso et al., 2010). Each sequencing run included: (1) positive controls consisting of a 1:5 mixture of *Staphylococcus aureus*, *Lactococcus lactis*, *Porphyromonas gingivalis*, *Streptococcus mutans*, and *Escherichia coli*; and (2) negative controls consisting of sterile saline.

Raw data from the Illumina MiSeq was first converted into FASTQ format 2x300 paired end sequence files using the bcl2fastq program, version 1.8.4, provided by Illumina. Format conversion was performed without de-multiplexing and the EAMMS algorithm was disabled. All other settings were default. Sequence processing and microbial composition analysis were performed with the Quantitative Insights into Microbial Ecology (QIIME) software package (Caporaso et al., 2010), version 1.9.1. Reads were multiplexed using a configuration described previously (Fadrosh et al., 2014). Briefly, for both reads in a pair, the first 12 bases were a barcode, which was followed by a primer, then a heterogeneity spacer, and then the target 16S rRNA sequence. Using a custom Python script, the barcodes from each read pair were removed, concatenated together, and stored in a separate file. Read pairs were assembled using fastq-join from the ea-utils package, requiring at least 40 bases of overlap and allowing a maximum of 10% mismatched bases. Read pairs that could not be assembled were discarded. The concatenated barcode sequences were prepended to the corresponding assembled reads, and the resulting sequences were converted from FASTQ to FASTA and QUAL files for QIIME analysis. Barcodes, forward primer, spacer, and reverse primer sequences were removed during de-multiplexing. Reads containing more than four mismatches to the known primer

sequences or more than three mismatches to all barcode sequences were excluded from subsequent processing and analysis. Assembled reads were truncated at the beginning of the first 30 base window with a mean Phred quality score of less than 20 or at the first ambiguous base, whichever came first. Resulting sequences shorter than 300 bases or containing a homopolymer longer than six bases were discarded. Operational taxonomic units (OTU) were picked using the default settings of the reference-based USEARCH (version 5.2) (Edgar et al., 2011) pipeline in QIIME, and the January 21st, 2014 release of the OSU CORE oral microbiome database as a closed reference (Griffen et al., 2011) and a 97% similarity threshold. Chimera detection was performed *de novo* with UCHIME, using default parameters (Edgar et al., 2011). OTU clusters with less than four sequences were removed, and representative sequences used to make taxonomic assignments for each cluster were selected on the basis of abundance. Taxonomic assignments were made using the QIIME implementation of BLAST (Altschul et al., 1990) with default settings and the OSU CORE reference database.

One-way ANOVA with Tukey-Kramer test (Evans et al., 2014; Mejia-Leon et al., 2014) was used to compare the proportions of the most populated genus/species relative abundance between three groups: CF, S-ECC *C. albicans* positive and negative conditions. DESeq2 negative binomial Wald test was used to compare the microbial abundance at species level between two groups at each time: S-ECC *C. albicans* positive and negative conditions; CF child and mother; S-ECC child and mother. Bacterial diversity within each sample (alpha diversity) was computed as Shannon Index (Shannon, 1948). Bacterial diversity between samples (beta diversity) was computed as the Bray-Curtis dissimilarity index (Bray and Curtis, 1957). The average dissimilarity between every individual in each pair of groups of interest was calculated. Principal Coordinators Analysis (PCoA) was used for visualization of the data present in the beta diversity distance matrix in 3-dimensional plots, which revealed further clustering patterns within the samples. Non-parametric t-test and two-sample t-test was used to compare the alpha and beta diversity between two groups at a time.

Glucosyltransferases (Gtfs) enzymatic activity in plaque

Plaque Gtf enzyme activity was examined using methods detailed previously (Hayacibara et al., 2004; Vacca Smith et al., 2007). Briefly, 200 µl of plaque sample was mixed with 200 µl of ([^{14}C]glucosyl)-sucrose substrate (0.2 µCi/ml; 200.0 mmol/L sucrose, 40 µmol/L dextran 9000, 0.02% sodium azide in adsorption buffer, pH6.5) to reach a final concentration of 100 mmol/L sucrose. The mixed solution was incubated at 37˚C with rocking overnight to allow glucan synthesis. Subsequently, insoluble glucans were collected by centrifugation (10,000g, 10 min, 4˚C) and washed using MilliQ water and their amounts determined by scintillation counting to assess Gtfs activity.

References:

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J Mol Biol* 215(3)*:*403-410.

Bougnoux ME, Tavanti A, Bouchier C, Gow NA, Magnier A, Davidson AD *et al.* (2003). Collaborative consensus for optimized multilocus sequence typing of Candida albicans. *Journal of clinical microbiology* 41(11)*:*5265-5266.

Bray JR, Curtis JT (1957). An ordination of upland forest communities of southern Wisconsin. *Ecological Monographs* 27(325-349.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7(5)*:*335-336.

Dardas M, Gill SR, Grier A, Pryhuber GS, Gill AL, Lee YH *et al.* (2014). The impact of postnatal antibiotics on the preterm intestinal microbiome. *Pediatr Res* 76(2)*:*150-158.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27(16)*:*2194-2200.

Evans CC, LePard KJ, Kwak JW, Stancukas MC, Laskowski S, Dougherty J *et al.* (2014). Exercise prevents weight gain and alters the gut microbiota in a mouse model of high fat diet-induced obesity. *PLoS One* 9(3)*:*e92193.

Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM *et al.* (2014). An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* 2(1)*:*6.

Grier A, Qiu X, Bandyopadhyay S, Holden-Wiltse J, Kessler HA, Gill AL *et al.* (2017). Impact of prematurity and nutrition on the developing gut microbiome and preterm infant growth. *Microbiome* 5(1)*:*158.

Griffen AL, Beall CJ, Firestone ND, Gross EL, Difranco JM, Hardman JH *et al.* (2011). CORE: a phylogenetically-curated 16S rDNA database of the core oral microbiome. *PLoS One* 6(4)*:*e19051.

Hayacibara MF, Koo H, Vacca-Smith AM, Kopec LK, Scott-Anne K, Cury JA *et al.* (2004). The influence of mutanase and dextranase on the production and structure of glucans synthesized by streptococcal glucosyltransferases. *Carbohydr Res* 339(12)*:*2127-2137.

Mejia-Leon ME, Petrosino JF, Ajami NJ, Dominguez-Bello MG, de la Barca AM (2014). Fecal microbiota imbalance in Mexican children with type 1 diabetes. *Sci Rep* 4(3814.

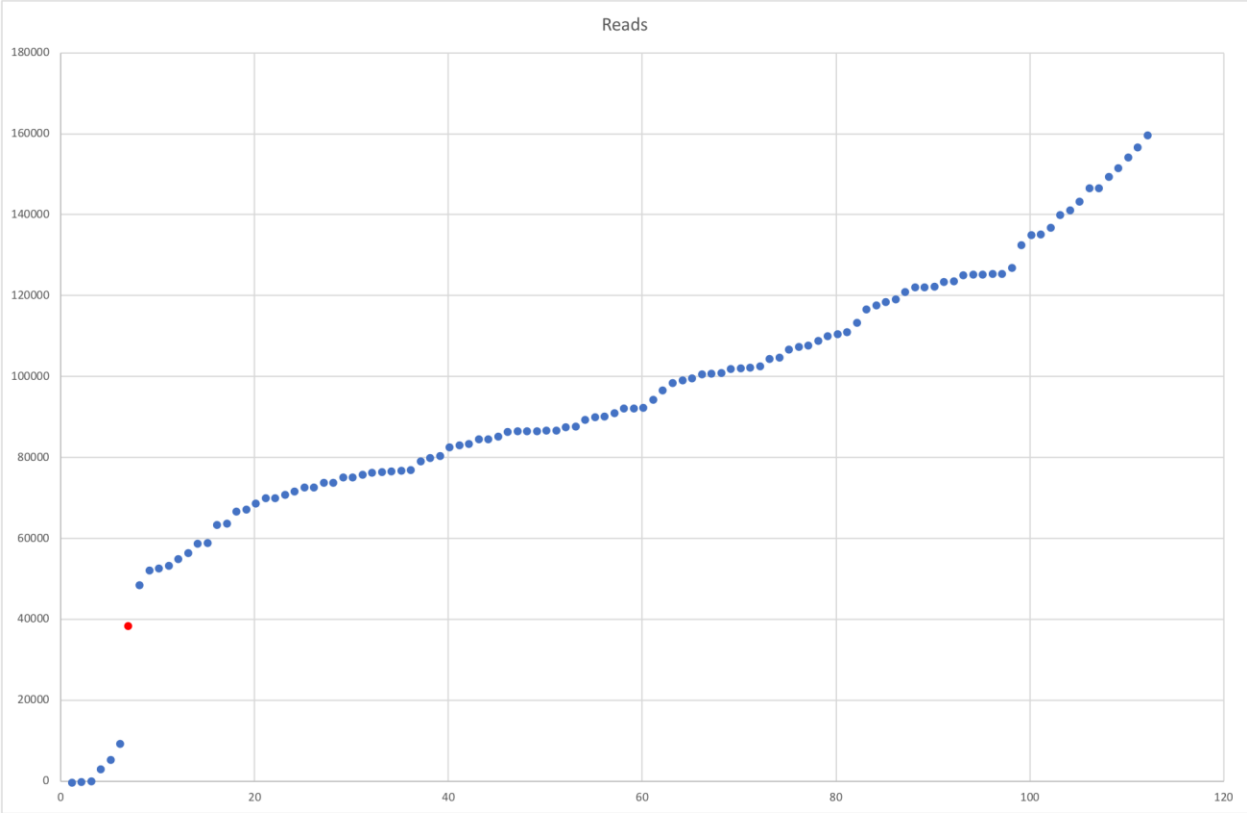
Merkley MA, Bice TC, Grier A, Strohl AM, Man LX, Gill SR (2015). The effect of antibiotics on the microbiome in acute exacerbations of chronic rhinosinusitis. *Int Forum Allergy Rhinol* 5(10)*:*884-893.

Odds FC, Bernaerts R (1994). CHROMagar Candida, a new differential isolation medium for presumptive identification of clinically important Candida species. *Journal of clinical microbiology* 32(8)*:*1923-1929.

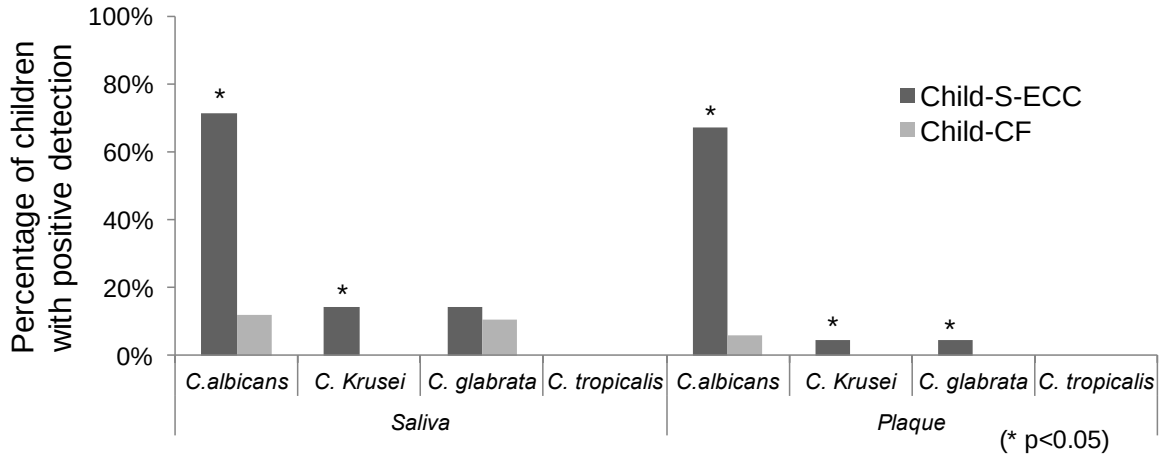
Shannon CE (1948). A mathematical theory of communication. *Bell System Technical Journal* 27(379-423, 623-656.

Vacca Smith AM, Scott-Anne KM, Whelehan MT, Berkowitz RJ, Feng C, Bowen WH (2007). Salivary glucosyltransferase B as a possible marker for caries activity. *Caries Res* 41(6)*:*445-450.

Zhu L, Baker SS, Gill C, Liu W, Alkhouri R, Baker RD *et al.* (2013). Characterization of gut microbiomes in nonalcoholic steatohepatitis (NASH) patients: a connection between endogenous alcohol and NASH. *Hepatology* 57(2)*:*601-609.

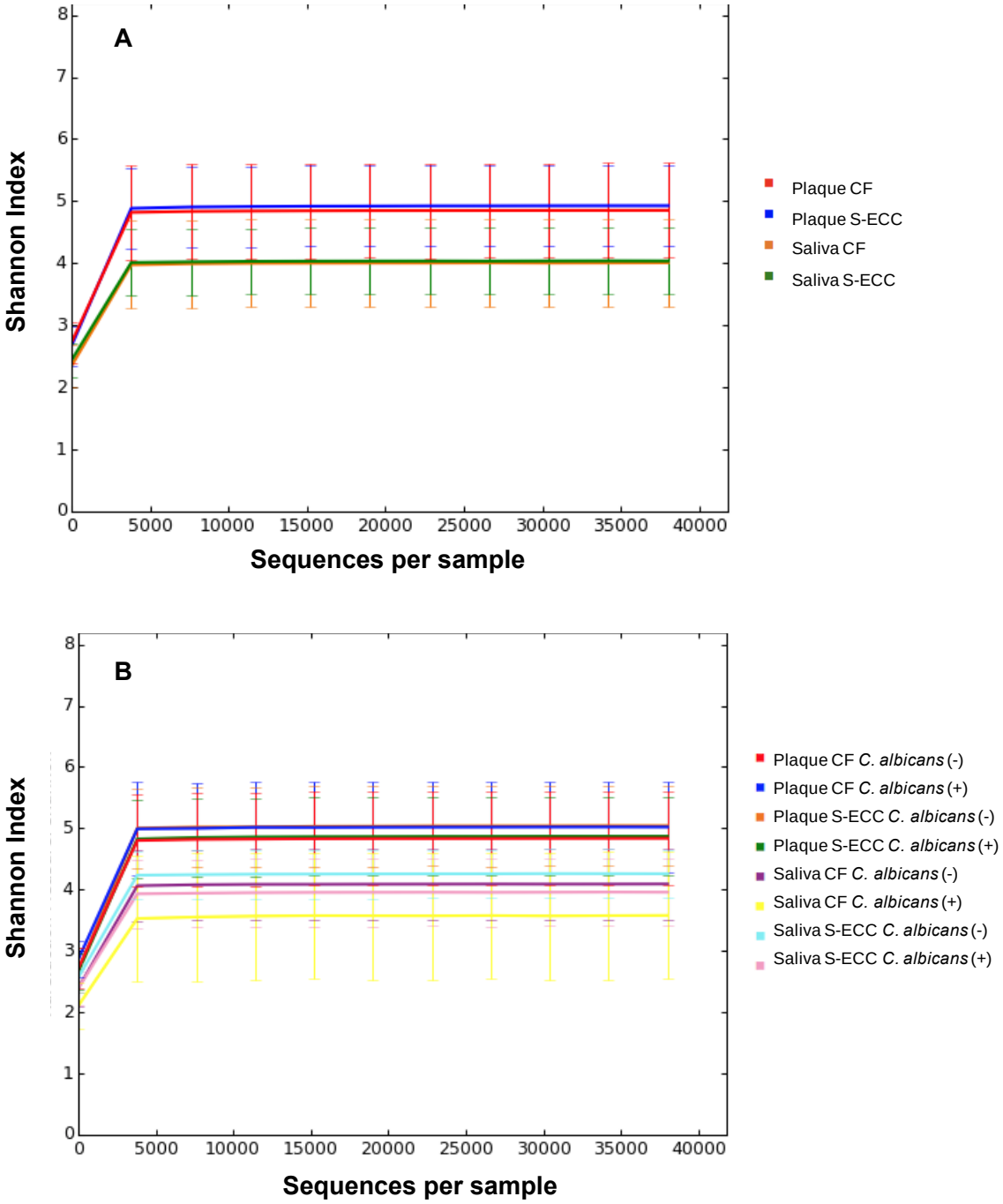


Appendix Figure 1. Reads per sequenced sample. Illumina sequencing produced an average of > 91,000 reads per sample after quality control and OTU identification among all sequenced samples. Five samples with markedly fewer reads than the threshold of 38,000 were excluded from all analyses.



Appendix Figure 2. *Candida* detection in saliva and plaque among S-ECC and CF children

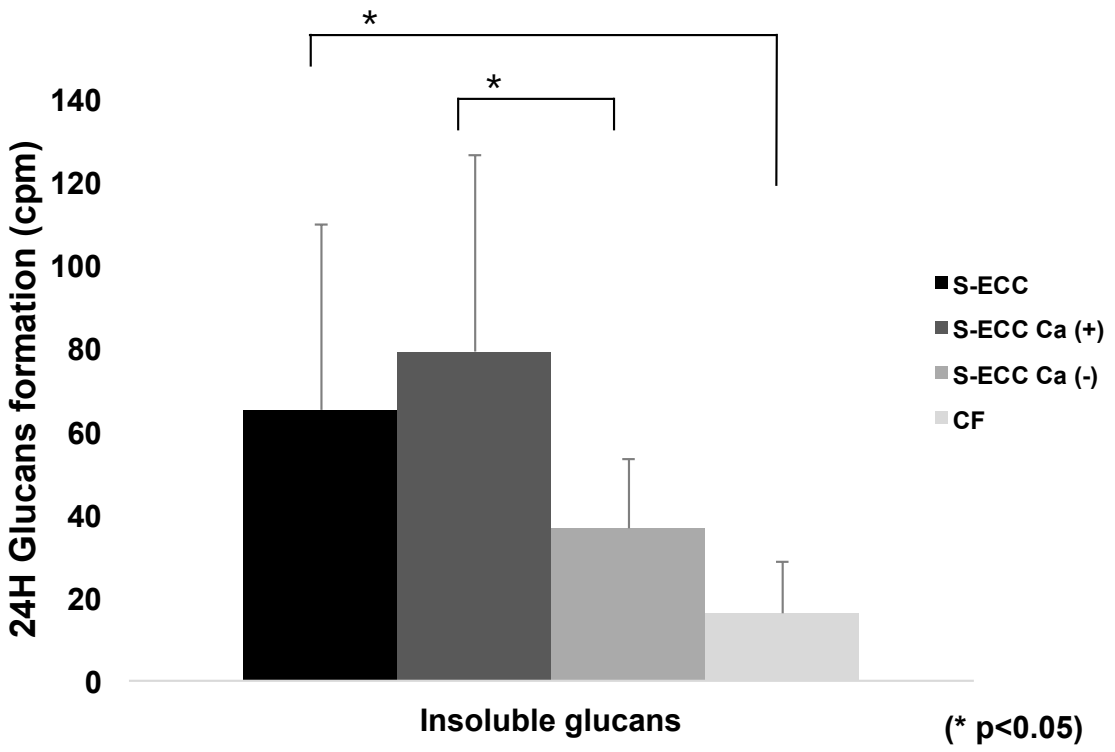
Percentage of children with positive detection of *C. albicans* and other *Candida* species are shown in this figure. *C. albicans* is the most prevalent isolate compared to other *Candida* strains in S-ECC children.



Appendix Figure 3. Rarefaction curves of alpha diversity measured from each sample groups.

Shannon curves were calculated at the 97% similarity level with bacteriome data from groups determined by caries status in (A) and *Candida albicans* detection status (B). CF - Caries free,

S-ECC - Severe early childhood caries.



Appendix Figure 4. Plaque Gtfs enzymatic activity compared between children with S-ECC, with and without *C. albicans*, and CF children.

200µl plaque suspensions were used for Gtfs activity test using the scintillation counting assay based on the 24h production of insoluble glucans. Gtfs activity was indicated by glucan formation shown in the figure.