Impact of enzymatic digestion on bacterial community composition in CF airway samples

Kayla M. Williamson [1], **Brandie D Wagner** [Corresp., 1, 2], **Charles E. Robertson** [3], **Emily J. Johnson** [2, 4], **Edith T. Zemanick** [2], **J. Kirk Harris** [2]

[1] Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado at Denver, Aurora, Colorado, United States
[2] Department of Pediatrics, School of Medicine, University of Colorado at Denver, Aurora, Colorado, United States
[3] Division of Infectious Diseases, School of Medicine, University of Colorado at Denver, Aurora, Colorado, United States
[4] Multicare Tacoma Family Medicine, Tacoma, Washington, United States

Corresponding Author: Brandie D Wagner
Email address: brandie.wagner@ucdenver.edu

Background. Previous studies have demonstrated the importance of DNA extraction methods for molecular detection of *Staphylococcus,* an important bacterial group in cystic fibrosis (CF). We sought to evaluate the effect of enzymatic digestion (EnzD) prior to DNA extraction on bacterial communities identified in sputum and oropharyngeal swab (OP) samples from patients with CF.

Methods. DNA from 81 samples (39 sputum and 42 OP) collected from 63 patients with CF was extracted in duplicate with and without EnzD. Bacterial communities were determined by rRNA gene sequencing, and measures of alpha and beta diversity were calculated. Principal Coordinate Analysis (PCoA) was used to assess differences at the community level and Wilcoxon Signed Rank tests were used to compare relative abundance (RA) of individual genera for paired samples with and without EnzD.

Results. Shannon Diversity Index (alpha-diversity) decreased in sputum and OP samples with the use of EnzD. Larger shifts in community composition were observed for OP samples (beta-diversity, measured by Morisita-Horn), whereas less change in communities was observed for sputum samples. The use of EnzD with OP swabs resulted in significant increase in RA for the
genera *Gemella* ($p < 0.01$), *Streptococcus* ($p < 0.01$), and *Rothia* ($p < 0.01$). *Staphylococcus (p<0.01)* was the only genus with a significant increase in RA from sputum, whereas the following genera decreased in RA with EnzD: *Veillonella (p<0.01)*, *Granulicatella (p<0.01)*, *Prevotella (p<0.01)*, and *Gemella (p<0.01).* In OP samples, higher RA of Gram-positive taxa was associated with larger changes in microbial community composition.

Discussion. We show that the application of EnzD to CF airway samples, particularly OP swabs, results in differences in microbial communities detected by sequencing. Use of

EnzD can result in large changes in bacterial community composition, and is particularly useful for detection of *Staphylococcus* in CF OP samples. The enhanced identification of *Staphylococcus aureus* is a strong indication to utilize EnzD in studies that use OP swabs to monitor CF airway communities.

Impact of enzymatic digestion on bacterial community composition in CF airway samples

Kayla M. Williamson[1], Brandie D. Wagner[1,2], Charles E. Robertson[3], Emily J. Johnson[2*], Edith T. Zemanick[2], and J. Kirk Harris[2]

[1]Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado School of Medicine, Aurora, Colorado, United States of America

[2]Department of Pediatrics, University of Colorado School of Medicine, Aurora, Colorado, United States of America

[3]Division of Infectious Diseases, University of Colorado, School of Medicine, Aurora, Colorado, United States of America

*Current Address: Multicare Tacoma Family Medicine, Tacoma, Washington, United States of America

Corresponding author:

Brandie Wagner

Email: brandie.wagner@ucdenver.edu

Abstract

Background. Previous studies have demonstrated the importance of DNA extraction methods for molecular detection of *Staphylococcus,* an important bacterial group in cystic fibrosis (CF). We sought to evaluate the effect of enzymatic digestion (EnzD) prior to DNA extraction on bacterial communities identified in sputum and oropharyngeal swab (OP) samples from patients with CF.

Methods. DNA from 81 samples (39 sputum and 42 OP) collected from 63 patients with CF was extracted in duplicate with and without EnzD. Bacterial communities were determined by rRNA gene sequencing, and measures of alpha and beta diversity were calculated. Principal Coordinate Analysis (PCoA) was used to assess differences at the community level and Wilcoxon Signed Rank tests were used to compare relative abundance (RA) of individual genera for paired samples with and without EnzD.

Results. Shannon Diversity Index (alpha-diversity) decreased in sputum and OP samples with the use of EnzD. Larger shifts in community composition were observed for OP samples (beta-diversity, measured by Morisita-Horn), whereas less change in communities was observed for sputum samples. The use of EnzD with OP swabs resulted in significant increase in RA for the genera *Gemella* ($p < 0.01$), *Streptococcus* ($p < 0.01$), and *Rothia* ($p < 0.01$). *Staphylococcus (p<0.01)* was the only genus with a significant increase in RA from sputum, whereas the following genera decreased in RA with EnzD: *Veillonella (p<0.01)*, *Granulicatella (p<0.01),* *Prevotella (p<0.01)*, and *Gemella (p<0.01).* In OP samples, higher RA of Gram-positive taxa was associated with larger changes in microbial community composition.

Discussion. We show that the application of EnzD to CF airway samples, particularly OP swabs, results in differences in microbial communities detected by sequencing. Use of EnzD can result

in large changes in bacterial community composition, and is particularly useful for detection of *Staphylococcus* in CF OP samples. The enhanced identification of *Staphylococcus aureus* is a strong indication to utilize EnzD in studies that use OP swabs to monitor CF airway communities.

Introduction

Cystic Fibrosis (CF) is an autosomal recessive disease, characterized by chronic airway infection, whose predominant pathogens include *Staphylococcus aureus, Pseudomonas aeruginosa* and other Gram-negative bacteria (Heijerman 2005; LiPuma 2010). S*taphylococcus aureus* in particular has a rigid cell wall that can be hard to rupture, hindering the ability to efficiently extract DNA for molecular detection (Zhao et al. 2012). Previous studies have indicated that use of enzymatic digestion (EnzD) may enhance the ability to detect *Staphylococcus* (Zhao et al. 2012).

Sequencing is becoming more widely used to evaluate bacterial communities in CF airway samples(Cummings et al. 2016). DNA preparation prior to sequencing can have a profound effect on microbial community composition especially in detecting certain organisms such as Sta*phylococcus aureus* (Lozupone et al. 2013; Pérez-Losada et al. 2016;Willner et al. 2012; Zhao et al. 2012). Given the differing results provided from various DNA extraction methods and the use of sequencing, it has become increasingly important to understand how these different methods impact sequencing results. We therefore sought to evaluate the effect of DNA extraction using enzymatic digestion (EnzD) on bacterial communities detected from oropharyngeal swab (OP) and sputum samples collected from patients with CF.

Materials & Methods

Patient demographics and samples. Sputum and OP samples were obtained from patients with CF as part of standard of care for monitoring bacterial infection during routine patient visits. Further explanation of selection of samples to be used for this study are included in a previous

publication(Johnson et al. 2016). Excess specimen was stored frozen at -80 ºC for molecular assessment of infection. Standard sputum processing protocol was performed for sample homogenization and standard CF culture was performed following CF Foundation guidelines(Burns et al. 1998). The Colorado Multiple Institutional Review Board approved the study (COMIRB 07-0835). Written informed consent was obtained from patients or guardians. Written informed assent was obtained for children 10-17 years of age.

DNA extraction methods. DNA was extracted using the Qiagen EZ1 Advanced automated extraction platform using the Tissue kit and bacterial card per manufacturer's instructions. One replicate of each sample was digested with Lysostaphin (0.18 mg/ml) and Lysozyme (3.6 mg/ml) for 30 min at 37 ºC, followed by treatment with Proteinase K (Qiagen) at 65 ºC for 10 min. Enzymes were heat inactivated at 95 ºC for 10 min (Johnson et al. 2016; Zhao et al. 2012).

16S Amplicon Library Construction. Bacterial profiles were determined by broad-range amplification and sequence analysis of 16S rRNA genes following our previously described methods (Hara et al. 2012; Markle et al. 2013). Amplicons were generated using primers that target approximately 300 base pairs of the V1/V2 variable region of the 16S rRNA gene. Each DNA was amplified in triplicate along with a barcode specific negative PCR control. PCR reactions contained 1X HotMaster Mix (5Prime), 150 nM each PCR primer and template in a reaction volume of 25µl. Cycling conditions were 94 ºC denaturation for 120 seconds followed by 30 cycles of 95 ºC 20 seconds, 52 ºC 20 seconds and 65 ºC 60 seconds. Amplification was confirmed using agarose gel electrophoresis. None of the negative PCR controls showed evidence of amplification. PCR products were normalized based on agarose gel densitometry, pooled, lyophilized, purified and concentrated using a DNA Clean and Concentrator Kit (Zymo, Irvine, CA). Pooled amplicons were quantified using Qubit Fluorometer *2.0* (Invitrogen, Carlsbad, CA). The pool was diluted to

4nM and denatured with 0.2 N NaOH at room temperature. The denatured DNA was diluted to 15pM and spiked with 10% of the Illumina PhiX control DNA prior to loading the sequencer. Illumina paired-end sequencing was performed on the MiSeq using a 500 cycle version 2 reagent kit.

Analysis of Illumina Paired-end Reads. As previously described, paired-end sequences were sorted by sample via barcodes in the paired reads with a python script (Markle et al. 2013) Sorted paired end sequence data were deposited in the NCBI Short Read Archive under accession number SRP043334. The sorted paired reads were assembled using phrap (Ewing and Green 1998; Ewing et al.1998.). Pairs that did not assemble were discarded. Assembled sequence ends were trimmed over a moving window of 5 nucleotides until average quality met or exceeded 20. Trimmed sequences with more than 1 ambiguity or shorter than 250 nt were discarded. Potential chimeras identified with Uchime (usearch6.0.203_i86linux32) (Edgar et al. 2011) using the Schloss Silva reference sequences (Schloss and Westcott 2011) were removed from subsequent analyses. Assembled sequences were aligned and classified with SINA (1.3.0-r23838)(Pruesse, Peplies, and Glöckner 2012) using the 479,726 bacterial sequences in Silva 115NR (Quast et al. 2013) as reference configured to yield the Silva taxonomy. Sequences with identical taxonomic assignments were grouped into Operational taxonomic units (OTUs). This process generated 4,302,223 sequences for 162 samples (average sequence length: 316 nt; average sample size: 26,557 sequences/sample; minimum sample size: 7,527; maximum samples size: 63,105). The median Good's coverage score was $\geq$ 99.70% at the rarefaction point of 7,527. The software package Explicet (v2.10.5, www.explicet.org) (Robertson et al. 2013) was used to calculate rarefied values for diversity measurements.

Statistical Analysis. Shannon-H alpha diversity and Morisita Horn (MH) beta diversity were calculated in Explicet. MH beta diversity is a measure of similarity between two communities and ranges from 0 (no similarity) to 1 (identical communities). Principal Coordinate Analysis (PCoA) was used to assess differences at the community level utilizing 1-MH. 1-MH distances were used to show the dissimilarity between with EnzD and without EnzD in PCoA plots instead of MH distances, which are a measure of similarity. Relative abundance (RA) for each genus was calculated by dividing the genera-specific sequence counts by the total number of sequences obtained for each sample. Differences in phyla and genera between EnzD samples versus non-EnzD samples with a median RA of at least 1% were evaluated using Wilcoxon Signed rank test for paired samples. Spearman correlations were used to assess associations between MH for paired samples and RA of specific genera from the non-digested sample. Analyses were calculated using R version 3.2.4 Revised (2016-03-16 r70336).

Results

Sample Collection

We analyzed 81 airway samples (39 sputum and 42 OP, Figure S1) collected from 63 patients with cystic fibrosis ranging in age from 1.5 to 24 years. Half of the subjects were female (51%), the majority of subjects were non-Hispanic white (89%) and 52% of subjects were homozygous F508. The median number of samples collected per subject was 1 and ranged between 1 and 3 samples. Seven samples (9%) were negative for standard CF pathogens by culture.

Changes in individual organisms

Individual taxa were assessed at the phyla and genera level by sample type (OP and Sputum). For OP samples, we observed a significantly lower RA for the phyla *Bacteroidetes, Fusobacteria* and *Proteobacteria* when EnzD was used. *Firmicutes* and *Actinobacteria* both were detected in higher RA with the use of EnzD. For sputum, the RA for *Bacteroidetes* and *Proteobacteria* were lower with EnzD (Table 1) consistent with findings in OP samples (Figure 1A).

Genera with a statistically significantly higher RA following EnzD for OP swabs consisted of: *Gemella, Streptococcus, Actinomyces, Johnsonella,* and *Rothia.* Genera with a statistically significantly lower RA for OP swabs consisted of: *Prevotella*, *Veillonella, Neisseria, Haemophilus, Leptotrichia*, *Campylobacter, Fusobacterium,* and *Granulicatella*. These genera, while statistically significant, may not be clinically meaningful. The largest changes in RA (>5%) are seen in *Streptococcus, Prevotella,* and *Veillonella* (Table 2). *Staphylococcus* was the only genus with a significantly higher RA in sputum following EnzD. Genera with a statistically significantly lower RA in sputum consisted of: *Veillonella*, *Granulicatella, Prevotella*, and *Gemella* (Table 2). The largest estimated change in RA for sputum samples is attributed to *Staphylococcus* (Table 2). *Gemella* was higher in OP samples with EnzD while it was lower in sputum samples with EnzD (Figure 1B).

Overall community changes.

The Shannon-H alpha Diversity Index was significantly lower ($p < 0.01$) with EnzD in sputum samples due to higher RA of *Staphylococcus* resulting in a less even community. For OP samples, the Shannon-H alpha Diversity Index was also lower likely due to higher RA of *Streptococcus* with EnzD, although this result did not reach statistical significance ($p = 0.06$).

Changes in alpha diversity due to use of EnzD do not necessarily correspond to changes in microbial community composition as measured using beta diversity (Figure S2).

Beta Diversity Changes

There was less impact of EnzD on community composition in sputum compared to OP samples as measured by MH beta diversity values for each sample pair. The majority of the MH values (79.5%) for sputum were greater than 0.8, whereas only 38.1% of OP samples were above that level. For the OP swabs the lowest MH was 0.4 for sputum samples the lowest MH was 0.48. The distributions between the two sample types were statistically different ($p < 0.01$, Figure 2). We further investigated these changes using the MH distance matrix in Principal Coordinate Analysis (PCoA). OP samples changed more consistently in ordination space with EnzD (Figure 3A). The consistent changes observed in the OP samples were largely due to the change in *Streptococcus*, *Gemella*, and *Rothia* (all Gram-positive) as well as *Neisseria*, *Leptotrichia*, *Prevotella*, and *Veillonella* (Gram-negative); with the samples undergoing EnzD primarily clustered around *Streptococcus, Gemella, and Rothia* (Figure 3A). Findings here align with previously discussed individual level changes in genera where Gram-positive bacteria were found in significantly higher RA while Gram-negative bacteria was found in significantly lower RA with EnzD (Table 2). As indicated in Figure 3B, sputum samples showed less impact from EnzD, with a large group of samples having limited changes in MH. A subset of the sputum samples demonstrated large changes in MH with EnzD, but with no consistent direction of change as was seen in the OP samples (Figure 3B). Sputum appears to be more heterogeneous compared to OP samples; Figure 3B shows three distinct clusters in sputum samples corresponding to communities primarily represented by *Staphylococcus*, *Streptococcus*, and a combination of *Veillonella*, and *Prevotella*. Sputum was primarily influenced by enhanced

179 extraction of *Staphylococcus, Veillonella,* and *Streptococcus* (Figure 3B). This result is
180 consistent with Table 2 where *Staphylococcus and Veillonella* are shown to have a larger change
181 than *Streptococcus.*

182 *Associations between community composition and MH.*

183 We found no striking evidence of a single phylum or genus whose relative abundance
184 was associated with MH (Figure S3). Combination of Gram-positive bacteria was calculated
185 using the summation of (Actinobacteria + Firmicutes (-*Veillonella*)). For the OP samples with
186 large changes in their communities with EnzD, the relative abundance of Gram-positive
187 organisms was lower compared to those samples with limited changes in their communities
188 (Figure 4). For all the OP samples, the relative abundance of Gram-positive organisms was
189 higher with EnzD. There was an association between change in relative abundance of Gram-
190 positive organisms with EnzD and MH values for the pairs ($r = -0.60$; $p < 0.01$; Figure 4B).

191 **Discussion**

192 In this work, we demonstrate that EnzD of airway samples changes the bacterial
193 community detected by sequencing primarily due to increased detection of organisms with a
194 Gram-positive cell wall structure. From the biplots (Figure 3A) we can see that for OP samples
195 there is a clear shift in microbial composition between those without EnzD and those with EnzD.
196 With EnzD the OP samples are highly clustered around *Gemella*, *Rothia*, and *Streptococcus*
197 whereas those without EnzD are dispersed around *Neisseria, Prevotella* and *Veillonella*. These
198 observations are consistent with cell wall structure; we would predict that given their rigid cell
199 wall Gram-positive bacteria would drive changes in community composition for those samples
200 with EnzD. *Streptococcus*, *Gemella,* and *Rothia* explain a large amount of the variability in the
201 OP samples. *Streptococcus* and *Staphylococcus* are Gram-positive taxa explaining a large

amount of the variability in the sputum samples. Among the Gram-negative taxa explaining a large amount of variability in OP and sputum samples are *Neisseria*, *Veillonella*, and *Prevotella* due to decreasing relative abundance with EnzD. Difference in OP and sputum samples can be due to differences in the microbial communities which may be related to multiple factors including anatomic location, inter individual variability and disease severity. The samples used in this study were collected for clinical evaluation of airway infection. The patients that were not capable of expectorating sputum tend to be younger with presumably less lung disease (i.e., fewer pathogens). There are also differences between upper and lower airways (Zemanick et al. 2015) that could also impact the community composition observed.

Willner et al. demonstrated that genera detection varied significantly between 5 commonly used extraction methods with *Staphylococcus* being detected with varying efficiency (Willner et al. 2012). Support for these findings showed the largest difference in microbial proportions between two DNA extraction methods (Norgen and Qiagen) was found in *Staphylococcus* (Pérez-Losada et al. 2016). Consistent with previous findings, EnzD revealed a higher RA of *Staphylococcus* in OP samples. Further, when evaluated in both OP swab and sputum samples from pediatric subjects with CF, variations in detection of *Staphylococcus* were observed (Johnson et al. 2016). Total RA for *Staphylococcus* in sputum samples was higher overall, but had less of a change before and after EnzD (Johnson et al. 2016). Zhao et al. looked at sputum and saw an increase in *Staphylococcus*, which we corroborate (Zhao et al. 2012). We also found a higher RA of *Streptococcus* in OP samples. Johnson et al. found that for OP samples DNA concentration consistently increased with EnzD. This indicates that in addition to observing changes in RA, EnzD probably better represents the absolute amounts of genera present in OP samples.

Limitations. There are certain limitations to our study. First, only two airway sample types collected from subjects with CF were evaluated, thus limiting more general inferences regarding the effects of EnzD in other samples types. However, this study does include OP samples, which have not previously been evaluated and represents an important sample type in early CF. Because only a single replicate with and without EnzD was evaluated, we are unable to compare effects due to EnzD versus technical variability. However, the observation of the consistent differences for the majority of the sample pairs provides some evidence that the change is due to EnzD rather than technical variability. In our study we considered beta-diversity values of 0.8 or greater to be within the limits of change due to biological variability.

The utility of EnzD is difficult to predict due to the strong influence of the community composition in any given sample on the results. There is a compelling argument to utilize EnzD in studies that rely on OP swabs due to the increased sensitivity for detection of *Staphylococcus*; while less compelling for sputum, EnzD should be applied consistently. Other sample types would require evaluation to determine how much impact is observed. However, increasing the RA of any group makes it harder to identify minor components in the community. EnzD could negatively impact the ability to identify Gram-negative organisms as seen through the decrease in Gram-negative genera shown in Table 2. This includes many important pathogens in CF (e.g. *Pseudomonas*) and should be considered carefully.

Conclusions

In summary, we found that the use of EnzD on airway samples prior to DNA extraction and sequencing alters microbiome community composition results, likely due to improved detection of gram-positive bacterial taxa (e.g. *Streptococcus* and *Staphylococcus*). The use of EnzD appears particularly important for analysis of OP swab samples. EnzD use with sputum

248 samples may be less critical and has the potential to decrease sensitivity for low abundance taxa.
249 Our findings highlight the need for a consistent approach to airway sample processing and
250 analysis, and suggest that EnzD should be applied routinely in studies using OP swabs and
251 studies requiring sensitive detection of *Staphylococcus* and *Streptococcus*.

252

253 **References**

254 Burns, J. L. et al. 1998. "Microbiology of Sputum from Patients at Cystic Fibrosis Centers in the
255 United States." *Clinical Infectious Diseases : An Official Publication of the Infectious*
256 *Diseases Society of America* 27(1):158–63. Retrieved August 12, 2016
257 (http://www.ncbi.nlm.nih.gov/pubmed/9675470).

258 Cummings, Lisa A. et al. 2016. "Clinical Next Generation Sequencing Outperforms Standard
259 Microbiological Culture for Characterizing Polymicrobial Samples." *Clinical Chemistry*
260 62(11).

261 Edgar, Robert C., Brian J. Haas, Jose C. Clemente, Christopher Quince, and Rob Knight. 2011.
262 "UCHIME Improves Sensitivity and Speed of Chimera Detection." *Bioinformatics (Oxford,*
263 *England)* 27(16):2194–2200. Retrieved July 14, 2016
264 (http://www.ncbi.nlm.nih.gov/pubmed/21700674).

265 Ewing, Brent and Phil Green. 1998. "Base-Calling of Automated Sequencer Traces Using *Phred.*
266 II. Error Probabilities." *Genome Research* 8(3):186–94. Retrieved August 22, 2016
267 (http://genome.cshlp.org/lookup/doi/10.1101/gr.8.3.186).

268 Ewing, Brent, Ladeana Hillier, Michael C. Wendl, and Phil Green. n.d. "Base-Calling of
269 Automated Sequencer Traces Using Phred. I. Accuracy Assessment."

270 Hara, N. et al. 2012. "Prevention of Virus-Induced Type 1 Diabetes with Antibiotic Therapy."
271 *The Journal of Immunology* 189(8):3805–14. Retrieved July 14, 2016
272 (http://www.jimmunol.org/cgi/doi/10.4049/jimmunol.1201257).

273 Heijerman, Harry. 2005. "Infection and Inflammation in Cystic Fibrosis: A Short Review."

Journal of Cystic Fibrosis : Official Journal of the European Cystic Fibrosis Society 4 Suppl 2:3–5. Retrieved August 23, 2016 (http://www.ncbi.nlm.nih.gov/pubmed/15970469).

Johnson, Emily J. et al. 2016. "Molecular Identification of Staphylococcus Aureus in Airway Samples from Children with Cystic Fibrosis." *PLoS ONE* 11(1):1–12.

LiPuma, J. J. 2010. "The Changing Microbial Epidemiology in Cystic Fibrosis." *Clinical Microbiology Reviews* 23(2):299–323. Retrieved August 23, 2016 (http://cmr.asm.org/cgi/doi/10.1128/CMR.00068-09).

Lozupone, C. A. et al. 2013. "Meta-Analyses of Studies of the Human Microbiota." *Genome Research* 23(10):1704–14. Retrieved August 11, 2016 (http://genome.cshlp.org/cgi/doi/10.1101/gr.151803.112).

Markle, Janet G. M. et al. 2013. "Sex Differences in the Gut Microbiome Drive Hormone-Dependent Regulation of Autoimmunity." *Science (New York, N.Y.)* 339(6123):1084–88. Retrieved July 14, 2016 (http://www.ncbi.nlm.nih.gov/pubmed/23328391).

Pérez-Losada, Marcos, Keith Crandall, Robert J. Freishtat, Marcos Pé Rez-Losada, and Keith A. Crandall. 2016. "Comparison of Two Commercial DNA Extraction Kits for the Analysis of Nasopharyngeal Bacterial Communities APA Citation Comparison of Two Commercial DNA Extraction Kits for the Analysis of Nasopharyngeal Bacterial Communities." *AIMS Microbiology* 2(2):108–19. Retrieved July 19, 2016 (http://hsrc.himmelfarb.gwu.edu/smhs_peds_facpubs).

Pruesse, Elmar, Jörg Peplies, and Frank Oliver Glöckner. 2012. "SINA: Accurate High-Throughput Multiple Sequence Alignment of Ribosomal RNA Genes." *Bioinformatics (Oxford, England)* 28(14):1823–29. Retrieved July 14, 2016

(http://www.ncbi.nlm.nih.gov/pubmed/22556368).

Quast, Christian et al. 2013. "The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools." *Nucleic Acids Research* 41(Database issue):D590-6. Retrieved July 14, 2016 (http://www.ncbi.nlm.nih.gov/pubmed/23193283).

Robertson, Charles E. et al. 2013. "Explicet: Graphical User Interface Software for Metadata-Driven Management, Analysis and Visualization of Microbiome Data." *Bioinformatics (Oxford, England)* 29(23):3100–3101. Retrieved August 4, 2016 (http://www.ncbi.nlm.nih.gov/pubmed/24021386).

Schloss, P. D. and S. L. Westcott. 2011. "Assessing and Improving Methods Used in Operational Taxonomic Unit-Based Approaches for 16S rRNA Gene Sequence Analysis." *Applied and Environmental Microbiology* 77(10):3219–26. Retrieved July 14, 2016 (http://aem.asm.org/cgi/doi/10.1128/AEM.02810-10).

Willner, Dana et al. 2012. "Comparison of DNA Extraction Methods for Microbial Community Profiling with an Application to Pediatric Bronchoalveolar Lavage Samples" edited by R. K. Aziz. *PLoS ONE* 7(4):e34605. Retrieved August 22, 2016 (http://dx.plos.org/10.1371/journal.pone.0034605).

Zemanick, Edith T. et al. 2015. "Assessment of Airway Microbiota and Inflammation in Cystic Fibrosis Using Multiple Sampling Methods." *Annals of the American Thoracic Society* 12(2):221–29. Retrieved September 14, 2016 (http://www.atsjournals.org/doi/10.1513/AnnalsATS.201407-310OC).

Zhao, Jiangchao et al. 2012. "Impact of Enhanced Staphylococcus DNA Extraction on Microbial Community Measures in Cystic Fibrosis Sputum." *PloS One* 7(3):e33127. Retrieved August

22, 2016 (http://www.ncbi.nlm.nih.gov/pubmed/22412992).

Table 1(on next page)

Table 1

Median absolute relative abundance in Non-EnzD groups and EnzD groups; estimated change (EnzD- Non-EnzD); in Median RA by Phyla

OP				
Phyla	EnzD %	Non-EnzD %	Estimated Change in RA %	P-Value
Actinobacteria	5.60	1.25	3.79	<0.01
Bacteroidetes	7.80	19.35	-12.73	<0.01
Firmicutes	68.31	48.32	16.78	<0.01
Fusobacteria	3.50	8.83	-1.97	<0.01
Proteobacteria	2.88	7.71	-5.52	<0.01
Sputum				
Phyla	EnzD %	Non-EnzD %	Estimated Change in RA %	P-Value
Actinobacteria	0.03	0.01	1.12	0.06
Bacteroidetes	0.01	0.06	-3.52	<0.01
Firmicutes	0.84	0.72	4.77	0.08
Proteobacteria	0.01	0.02	-1.36	<0.01

Table 2(on next page)

Table 2

Median absolute relative abundance in Non-EnzD groups and EnzD groups; estimated change (EnzD- Non-EnzD); in Median RA by Genera

OP				
Genera	EnzD %	Non-EnzD %	Estimated Change in RA %	P-Value
Streptococcus	36.51	14.81	21.75	<0.01
Prevotella	6.62	16.20	-10.90	<0.01
Veillonella	9.27	19.22	-7.86	<0.01
Neisseria	0.72	1.94	-3.62	<0.01
Actinomyces	2.93	0.73	2.50	<0.01
Granulicatella	1.80	2.44	-0.62	<0.01
Gemella	1.89	0.64	1.23	<0.01
Leptotrichia	1.16	3.79	-1.44	<0.01
Fusobacterium	1.47	3.08	-0.56	<0.01
Rothia	1.70	0.27	1.25	<0.01
Haemophilus	0.45	1.69	-0.91	<0.01
Johnsonella	0.97	0.40	1.19	<0.01
Firmicutes	0.65	0.89	-0.14	0.16
Campylobacter	0.50	0.75	-0.25	<0.01
Sputum				
Genera	EnzD %	Non-EnzD %	Estimated Change in RA %	P-Value
Streptococcus	27.85	22.48	1.33	0.5
Staphylococcus	1.39	0.62	5.47	<0.01
Prevotella	0.88	3.54	-1.91	<0.01
Veillonella	1.09	6.01	-3.55	<0.01
Granulicatella	0.53	1.77	-1.30	<0.01
Gemella	0.53	1.68	-0.62	<0.01
Rothia	0.84	0.57	0.21	0.24

Figure 1(on next page)

Figure 1

Boxplots for those individual genera that change within paired samples. The distribution for differences in paired samples is shown for phyla (top) and genera (bottom). Genera with limited differences between the paired samples are tightly distributed around zero, and those with increased relative abundance after EnzD have positive distributions. Note the median RA for *Fusobacteria* in sputum is less than 1%.

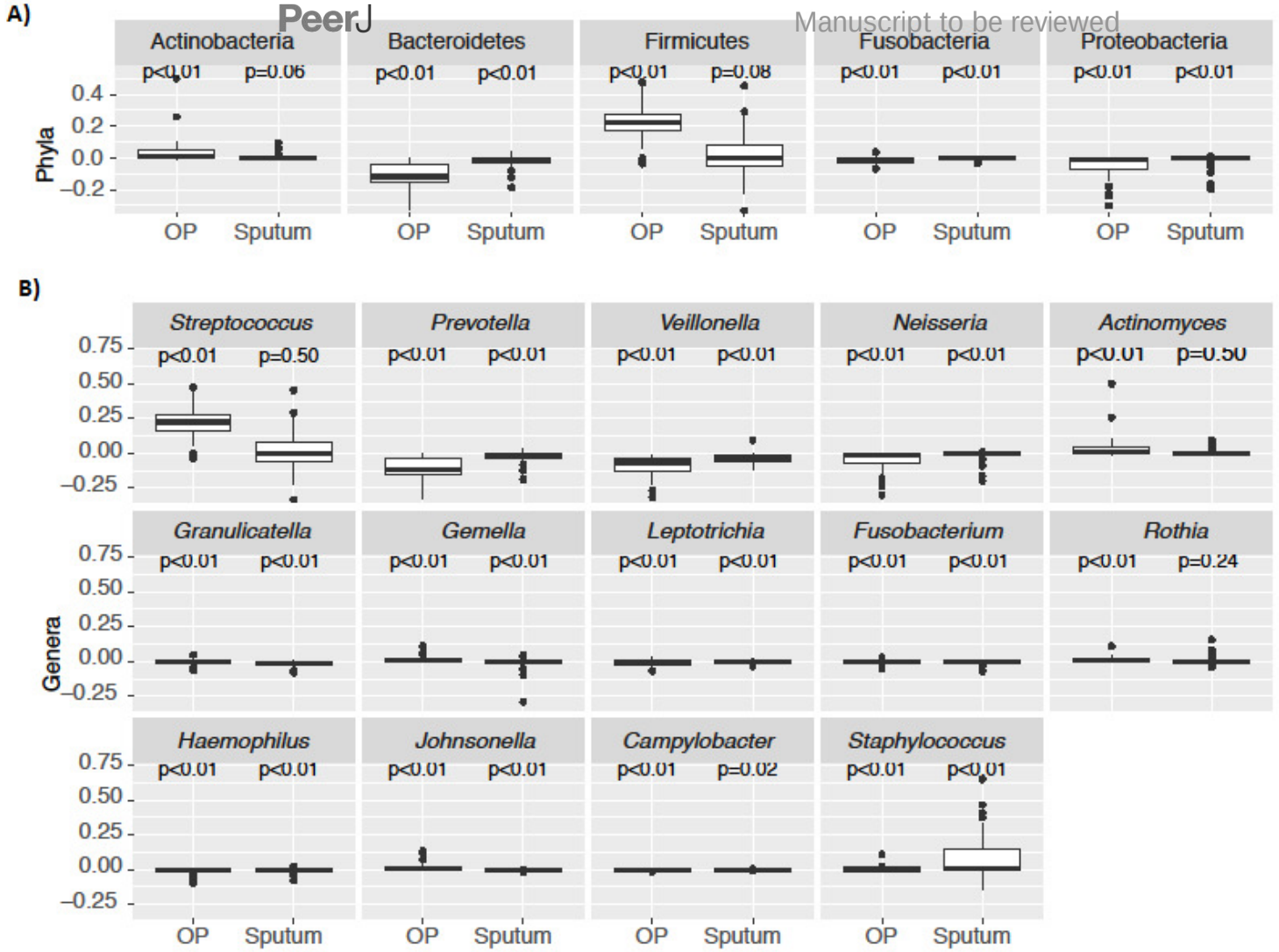


Figure 2(on next page)

Figure 2

Boxplots showing distribution of MH for paired samples in op and sputum

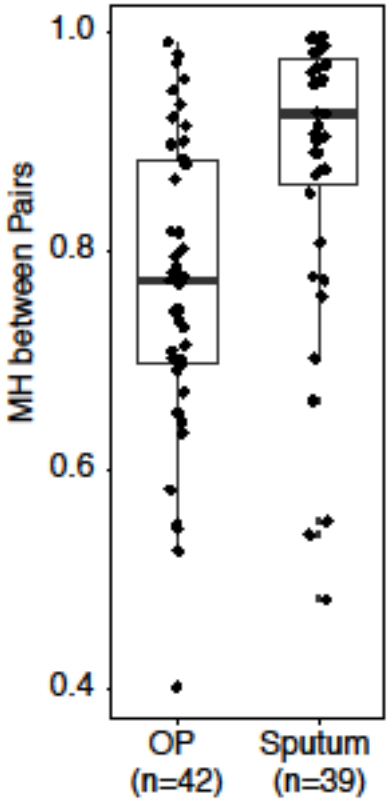

1.0
0.8
0.6
0.4
MH between Pairs
OP
(n=42)
Sputum
(n=39)

Figure 3(on next page)

Figure 3

Ordination biplot, using 1-MH beta diversity values for sputum (left) and OP (right). Paired samples are connected with a line, vectors for the genera with at least 1% RA and p<.005 of the loading.

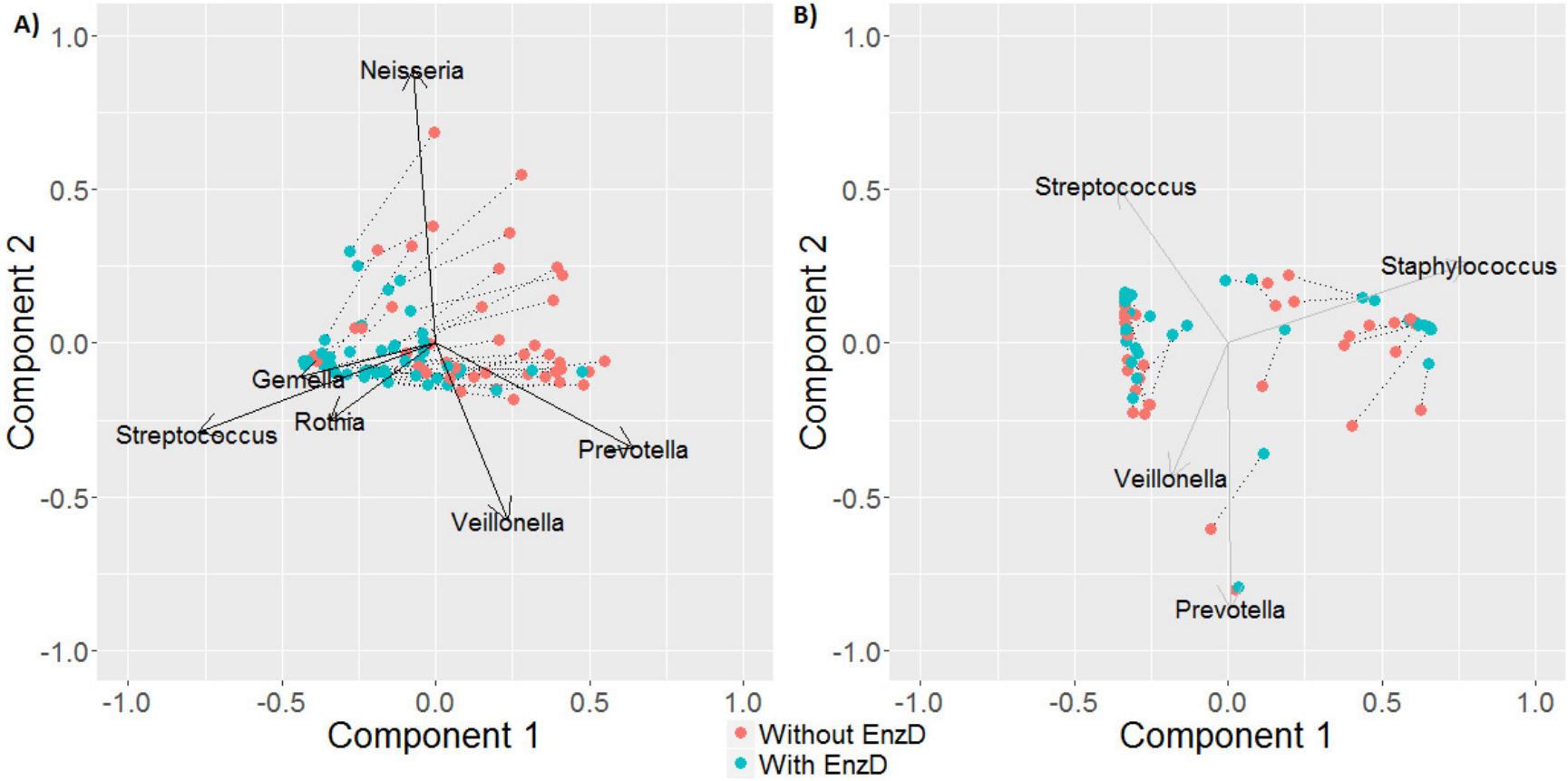
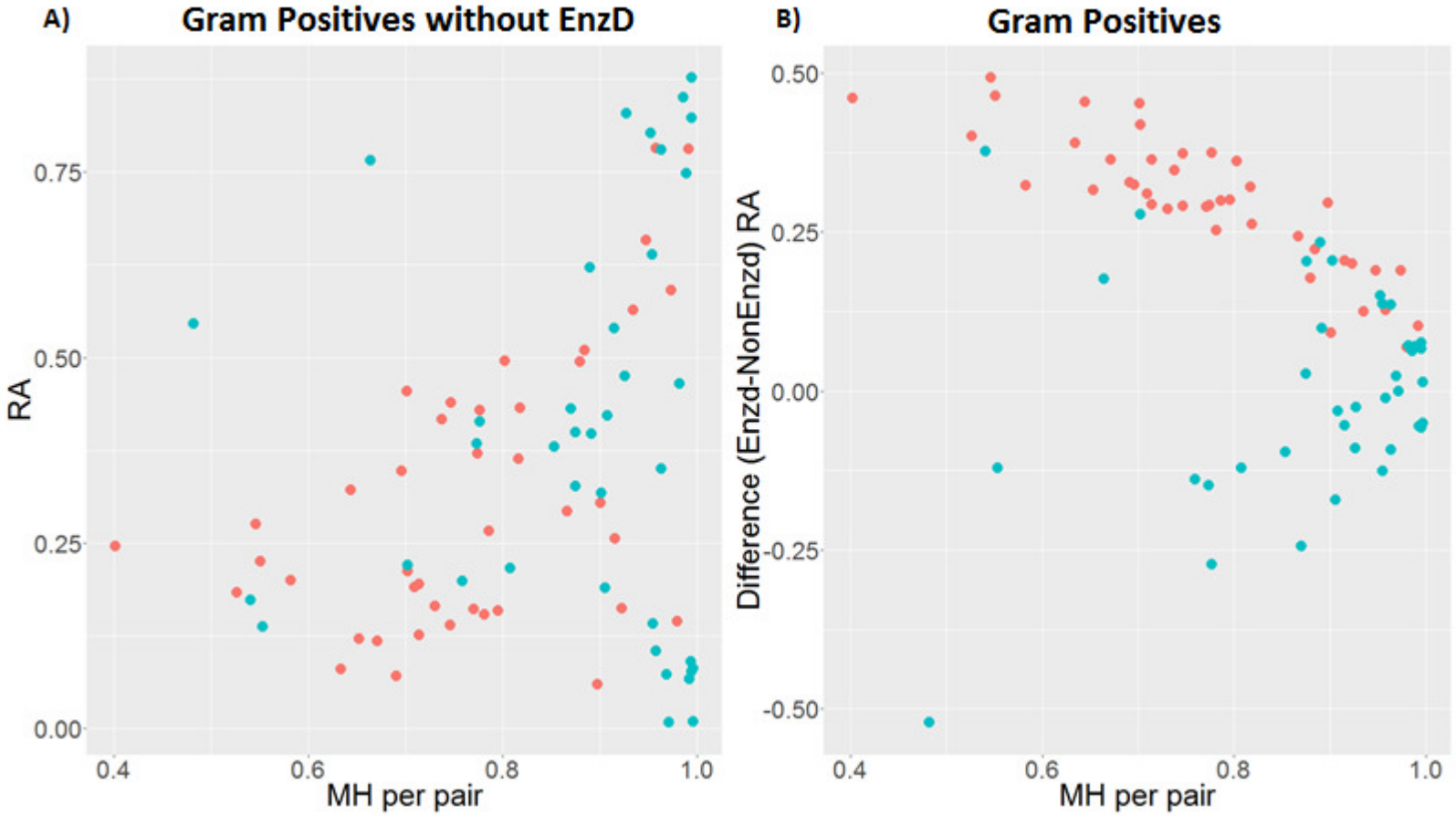

A)
B)
Component 1
Component 2
Neisseria
Gemella
Rothia
Streptococcus
Prevotella
Veillonella
Staphylococcus
Without EnzD
With EnzD

Figure 4(on next page)

Figure 4

Scatterplots of relative abundance of gram positive in non-EnzD sample versus MH of paired samples OP samples are shown in red and sputum in blue.


A)
Gram Positives without EnzD
RA
0.00
0.25
0.50
0.75
0.4
0.6
0.8
1.0
MH per pair
B)
Gram Positives
Difference (Enzd-NonEnzd) RA
-0.50
-0.25
0.00
0.25
0.50
0.4
0.6
0.8
1.0
MH per pair