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 thegenera *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=0.02).* 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.

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

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

3   T. Zemanick[2], and J. Kirk Harris[2]

5   [1]Department of Biostatistics and Informatics, Colorado School of Public Health, University of

6   Colorado School of Medicine, Aurora, Colorado, United States of America

7   [2]Department of Pediatrics, University of Colorado School of Medicine, Aurora, Colorado,

8   United States of America

9   [3]Division of Infectious Diseases, University of Colorado, School of Medicine, Aurora, Colorado,

10   United States of America

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

12   America

15   Corresponding author:

16   Brandie Wagner

17   Email: brandie.wagner@ucdenver.edu

19   **Abstract**

20   **Background.** Previous studies have demonstrated the importance of DNA extraction methods

21   for molecular detection of *Staphylococcus,* an important bacterial group in cystic fibrosis (CF).

22   We sought to evaluate the effect of enzymatic digestion (EnzD) prior to DNA extraction on

23   bacterial communities identified in sputum and oropharyngeal swab (OP) samples from patients

24   with CF.

25   **Methods.** DNA from 81 samples (39 sputum and 42 OP) collected from 63 patients with CF was

26   extracted in duplicate with and without EnzD. Bacterial communities were determined by rRNA

27   gene sequencing, and measures of alpha and beta diversity were calculated. Principal Coordinate

28   Analysis (PCoA) was used to assess differences at the community level and Wilcoxon Signed

29   Rank tests were used to compare relative abundance (RA) of individual genera for paired

30   samples with and without EnzD.

31   **Results.** Shannon Diversity Index (alpha-diversity) decreased in sputum and OP samples with

32   the use of EnzD. Larger shifts in community composition were observed for OP samples (beta-

33   diversity, measured by Morisita-Horn), whereas less change in communities was observed for

34   sputum samples. The use of EnzD with OP swabs resulted in significant increase in RA for the

35   genera *Gemella* ($p < 0.01$), *Streptococcus* ($p < 0.01$), and *Rothia* ($p < 0.01$). *Staphylococcus*

36   *($p<0.01$)* was the only genus with a significant increase in RA from sputum, whereas the

37   following genera decreased in RA with EnzD: *Veillonella ($p<0.01$)*, *Granulicatella ($p<0.01$)*,

38   *Prevotella ($p<0.01$)*, and *Gemella ($p=0.02$)*. In OP samples, higher RA of Gram-positive taxa

39   was associated with larger changes in microbial community composition.

40   **Discussion.** We show that the application of EnzD to CF airway samples, particularly OP swabs,

41   results in differences in microbial communities detected by sequencing. Use of EnzD can result

42   in large changes in bacterial community composition, and is particularly useful for detection of

43   *Staphylococcus* in CF OP samples. The enhanced identification of *Staphylococcus aureus* is a

44   strong indication to utilize EnzD in studies that use OP swabs to monitor CF airway

45   communities.

46 **Introduction**

47   Cystic Fibrosis (CF) is an autosomal recessive disease, characterized by chronic airway

48   infection, whose predominant pathogens include *Staphylococcus aureus, Pseudomonas*

49   *aeruginosa* and other Gram-negative bacteria (Heijerman, 2005; LiPuma, 2010). OP swabs are

50   commonly used when children are unable to expectorate (Zemanick et al., 2015). This is

51   particularly true for pediatric, non-expectorating subjects with CF in whom oropharyngeal (OP)

52   cultures are used as a surrogate for lower airway bacteria. (Zemanick et al., 2015). In the

53   pediatric studies, OP swabs have ranged from 30% to 68% of samples collected (Armstrong et

54   al., 1996; Zemanick et al., 2010; Wolter et al., 2013; Hoppe et al., 2015) and therefore represent

55   an important sample type in early CF.

56   Sequencing is becoming more widely used to evaluate bacterial communities in CF

57   airway samples (Cummings et al., 2016). DNA preparation prior to sequencing, including cell

58   lysis, can have a profound effect on microbial community composition especially in detecting

59   certain organisms such as *S. aureus*(Zhao et al., 2012; Yuan et al., 2012b; Willner et al., 2012;

60   Lozupone et al., 2013; Pérez-losada, Crandall & Freishtat, 2016). S. *aureus* in particular has a

61   rigid cell wall that can be hard to rupture, hindering the ability to efficiently extract DNA for

62   molecular detection (Zhao et al., 2012).

63   Previous studies have indicated that use of enzymatic digestion (EnzD) may enhance the

64   ability to detect *Staphylococcus* (Schindler & Schuhardt, 1964; Browder et al., 1965; Yuan et al.,

65   2012a; Zhao et al., 2012; Johnson et al., 2016) . Specifically, Yuan et al (2012) performed a

66   comprehensive experiment evaluating multiple DNA extraction methods utilizing human

67   associated bacterial species as well as a mock community. Zhao et al. (2012) found that EnzD

68   increased the yield of *Staphylococcus* in sputum samples. Similarly, Johnson et al (2016)

69   determined EnzD improved the sensitivity of sequencing in OP swabs while no differences were

70   observed in sputum samples. Increased sensitivity with EnzD was robust when compared to

71   clinical culture results and qPCR. Furthermore, the work showed that the majority of

72   *Staphylococcus* was *S. aureus*. It remains unclear, however, how the use of EnzD effects the

73   remaining bacterial community in clinical CF samples. In this work, we build on these two

74   previous CF studies to evaluate the effect of DNA extraction using enzymatic digestion (EnzD)

75   on bacterial communities detected from oropharyngeal swab (OP) and sputum samples collected

76   from patients with CF.

77

78   **Materials & Methods**

79   *Patient demographics and samples.* Sputum and OP samples were obtained from patients

80   with CF as part of standard of care for monitoring bacterial infection during routine patient visits.

81   Further explanation of selection of samples to be used for this study are included in a previous

82   publication(Johnson et al., 2016). Excess specimen was stored frozen at -80 ℃ for molecular

83   assessment of infection. Standard sputum processing protocol was performed for sample

84   homogenization utilizing sputalysin and standard CF culture was performed following CF

85   Foundation guidelines (Burns et al., 1998). The Colorado Multiple Institutional Review Board

86   approved the study (COMIRB 07-0835). Written informed consent was obtained from patients or

87   guardians. Written informed assent was obtained for children 10-17 years of age.

88   *DNA extraction methods*. DNA was extracted using the Qiagen EZ1 Advanced automated

89   extraction platform using the Tissue kit and bacterial card per manufacturer's instructions. EnzD

90   was performed on one replicate of each sample by mixing with lysostaphin (final concentration

91 0.18 mg/mL) and lysozyme (3.6 mg/mL) and incubated at 37°C for 30 minutes. Samples were

92 then digested with proteinase K (1.4 mg/mL) and incubated at 65°C for ten minutes, then

93 incubated at 95°C for 10 minutes. (Zhao et al., 2012; Johnson et al., 2016). Lysozyme and

94 lysostaphin target degradation of the bacterial cell wall by targeting peptidoglycan (lysozyme)

95 and pentaglycine bridges (lysostaphin) (Salazar & Asenjo, 2007). Lysostaphin is specific to the

96 subset of staphylococci that contain pentaglycine bridges including the human pathogen S.

97 aureus (Trayer & Buckley, 1970). Proteinase K is a broad specificity endopeptidase that is used

98 to digest proteins (Gradisar et al., 2005).

99   *16S* rRNA gene *Amplicon Library Construction.* Bacterial profiles were determined by

100 broad-range amplification and sequence analysis of 16S rRNA genes following our previously

101 described methods (Hara et al., 2012; Markle et al., 2013). Amplicons were generated using

102 primers that target approximately 300 base pairs of the V1/V2 variable region of the 16S rRNA

103 gene. Each DNA was amplified in triplicate along with a barcode specific negative PCR control.

104 PCR reactions contained 1X HotMaster Mix (5Prime), 150 nM each PCR primer and template in

105 a reaction volume of 25μl. Cycling conditions were 94 ºC denaturation for 120 seconds followed

106 by 30 cycles of 95 ºC 20 seconds, 52 ºC 20 seconds and 65 ºC 60 seconds. Amplification was

107 confirmed using agarose gel electrophoresis. None of the negative PCR controls showed evidence

108 of amplification. PCR products were normalized based on agarose gel densitometry, pooled,

109 lyophilized, purified and concentrated using a DNA Clean and Concentrator Kit (Zymo, Irvine,

110 CA). Pooled amplicons were quantified using Qubit Fluorometer *2.0* (Invitrogen, Carlsbad, CA).

111 The pool was diluted to 4nM and denatured with 0.2 N NaOH at room temperature. The denatured

112 DNA was diluted to 15pM and spiked with 10% of the Illumina PhiX control DNA prior to loading

113   the sequencer. Illumina paired-end sequencing was performed on the MiSeq using a 500 cycle

114   version 2 reagent kit.

115   *Analysis of Illumina Paired-end Reads.* As previously described, paired-end sequences

116   were sorted by sample via barcodes in the paired reads with a python script (Markle et al., 2013)

117   Sorted paired end sequence data were deposited in the NCBI Short Read Archive under

118   accession number SRP043334. The sorted paired reads were assembled using phrap (Ewing and

119   Green 1998; Ewing et al.1998.). Pairs that did not assemble were discarded. Assembled

120   sequence ends were trimmed over a moving window of 5 nucleotides until average quality met

121   or exceeded 20. Trimmed sequences with more than 1 ambiguity or shorter than 250 nt were

122   discarded. Potential chimeras identified with Uchime (usearch6.0.203_i86linux32) (Edgar et al.,

123   2011) using the Schloss Silva reference sequences (Schloss & Westcott, 2011) were removed

124   from subsequent analyses. Assembled sequences were aligned and classified with SINA (1.3.0-

125   r23838)(Pruesse, Peplies & Glöckner, 2012) using the 479,726 bacterial sequences in Silva

126   115NR (Quast et al., 2013) as reference configured to yield the Silva taxonomy. Sequences with

127   identical taxonomic assignments were grouped into Operational taxonomic units (OTUs). This

128   process generated 4,302,223 sequences for 162 samples (average sequence length: 316 nt;

129   average sample size: 26,557 sequences/sample; minimum sample size: 7,527; maximum samples

130   size: 63,105). The median Good's coverage score was $\geq$ 99.70% at the rarefaction point of 7,527.

131   The software package Explicet (v2.10.5, www.explicet.org) (Robertson et al., 2013) was used to

132   calculate rarefied values for diversity measurements.

133   *Statistical Analysis.* Shannon-H alpha diversity, evenness, richness and Morisita Horn

134   (MH) beta diversity were calculated in Explicet. MH beta diversity is a measure of similarity

135   between two communities and ranges from 0 (no similarity) to 1 (identical communities).

136   Principal Coordinate Analysis (PCoA) was used to assess differences at the community level

137   utilizing 1-MH. 1-MH distances were used to show the dissimilarity between with EnzD and

138   without EnzD in PCoA plots instead of MH distances, which are a measure of similarity.

139   Relative abundance (RA) for each genus was calculated by dividing the genera-specific sequence

140   counts by the total number of sequences obtained for each sample. Shannon-H diversity,

141   evenness and richness were compared using a Wilcoxon Signed Rank test. Differences in phyla

142   and genera between EnzD samples versus non-EnzD samples with a median RA of at least 1%

143   were evaluated using Wilcoxon Signed rank test for paired samples. Benjamini-Hochberg

144   corrections were used to account for multiple comparisons (Benjamini & Yekutieli, 2001).

145   Spearman correlations were used to assess associations between MH for paired samples and RA

146   of specific genera from the non-digested sample. Analyses were calculated using R version 3.2.4

147   Revised (2016-03-16 r70336).

148

149   **Results**

150   *Sample Collection.* We analyzed 81 airway samples (39 sputum and 42 OP, Figure S1)

151   collected from 63 patients with cystic fibrosis ranging in age from 1.5 to 24 years (9 to 24

152   sputum and 1.5 to 23 OP). Half of the subjects were female (51%), the majority of subjects were

153   non-Hispanic white (89%) and 52% of subjects were homozygous F508. The median number of

154   samples collected per subject was 1 and ranged between 1 and 3 samples. Seven samples (9%)

155   were negative for standard CF pathogens by culture (Figure S2).

156   *Changes in individual organisms.* Individual taxa were assessed at the phyla and genera

157   level by sample type (OP and Sputum). For OP samples, we observed a significantly lower RA

158   for the phyla *Bacteroidetes, Fusobacteria* and *Proteobacteria* when EnzD was used. *Firmicutes*

159   and *Actinobacteria* both were detected in higher RA with the use of EnzD. For sputum, the RA

160   for *Bacteroidetes* and *Proteobacteria* were lower with EnzD (Table 1) consistent with findings in

161   OP samples (Figure 1A).

162   Genera with a statistically significantly higher RA following EnzD for OP swabs

163   consisted of: *Gemella, Streptococcus, Actinomyces, Johnsonella,* and *Rothia.* Genera with a

164   statistically significantly lower RA for OP swabs consisted of: *Prevotella, Veillonella, Neisseria,*

165   *Haemophilus, Leptotrichia, Campylobacter, Fusobacterium,* and *Granulicatella*. For these

166   genera, while the changes are statistically significant, the changes may not be clinically

167   meaningful. The largest changes in RA (>5%) are seen in *Streptococcus, Prevotella,* and

168   *Veillonella* (Table 2). *Staphylococcus* was the only genus with a significantly higher RA in

169   sputum following EnzD. Genera with a statistically significantly lower RA in sputum consisted

170   of: *Veillonella, Granulicatella, Prevotella*, and *Gemella* (Table 2). The largest estimated change

171   in RA for sputum samples is attributed to *Staphylococcus* (Table 2). *Gemella* was higher in OP

172   samples with EnzD while it was lower in sputum samples with EnzD (Figure 1B).

173   Genera traditionally associated with CF include *Pseudomonas, Staphylococcus,*

174   *Haemophilus, Stenotrophomonas*, and *Achromobacter*(Zemanick et al., 2015). All of the samples

175   have low RA for *Achromobacter* and all of the OP samples had low RA for *Pseudomonas* and

176   *Stenotrophomonas*. In the sputum samples, 11 had low amounts of all CF pathogens, 11 were

177   dominated by *Staphylococcus*, all of which increased with EnzD, and 3 were dominated mostly

178   by *Haemophilus*. Of the remaining 14 sputum samples, 3 samples were dominated by

179   *Stenotrophomonas*, 1 increased, 1 decreased and 1 remained unchanged with EnzD. Eleven

180   sputum samples had > 5% RA *Pseudomonas* prior to EnzD, the RA increased after EnzD in 6 of

181 these samples (median 18.7 range 1.1% to 68.4%), in those where RA decreased, the median and

182 range was -6.1%, -1.2% to 42.0% (Table 3).

183 *Overall community changes.* The Shannon-H alpha Diversity Index was significantly

184 lower (p < 0.01) with EnzD in sputum samples (Figure S3) due to higher RA of *Staphylococcus*

185 resulting in a significantly less even community (Shannon-H, Wilcoxon Signed Rank, p <0.01).

186 For OP samples, the Shannon-H alpha Diversity Index was also lower likely due to higher RA of

187 *Streptococcus* with EnzD, although this result did not reach statistical significance (Shannon-H,

188 Wilcoxon Signed Rank, p = 0.06). Richness and evenness were also not significantly affected by

189 EnzD in OP samples. Changes in alpha diversity due to use of EnzD do not necessarily

190 correspond to changes in microbial community composition as measured using beta diversity

191 (Figure S4).

192 *Beta Diversity Changes.* There was less impact of EnzD on community composition in

193 sputum compared to OP samples as measured by MH beta diversity values for each sample pair.

194 The majority of the MH values (79.5%) for sputum were greater than 0.8, whereas only 38.1% of

195 OP samples were above that level. For the OP swabs the lowest MH was 0.4 for sputum samples

196 the lowest MH was 0.48. The distributions between the two sample types were statistically

197 different (MH, Wilcoxon Rank Sum, p < 0.01, Figure 2). We further investigated these changes

198 using the MH distance matrix in Principal Coordinate Analysis (PCoA). OP samples changed

199 more consistently in ordination space with EnzD (Figure 3A). The consistent changes observed

200 in the OP samples were largely due to the change in *Streptococcus*, *Gemella*, and *Rothia* (all

201 Gram-positive) as well as *Neisseria*, *Leptotrichia*, *Prevotella*, and *Veillonella* (Gram-negative);

202 with the samples undergoing EnzD primarily clustered around *Streptococcus, Gemella, and*

203 *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. Sputum was primarily influenced by enhanced extraction of *Staphylococcus, Veillonella,* and *Streptococcus* (Figure 3B). This result is consistent with Table 2 where *Staphylococcus and Veillonella* are shown to have a larger change than *Streptococcus.*

Associations between community composition and MH. We found no striking evidence of a single phylum or genus whose relative abundance was associated with MH (Figure S5). Combination of Gram-positive bacteria was calculated using the summation of (Actinobacteria + Firmicutes (-*Veillonella*)). For the OP samples with large changes in their communities with EnzD, the relative abundance of Gram-positive organisms was lower compared to those samples with limited changes in their communities (Figure 4). For all the OP samples, the relative abundance of Gram-positive organisms was higher with EnzD. There was an association between change in relative abundance of Gram-positive organisms with EnzD and MH values for the pairs (Spearman's correlation, r = -0.60; p < 0.01; Figure 4B).

Discussion

In this work, we demonstrate that EnzD of airway samples changes the bacterial community detected by sequencing primarily due to increased detection of organisms with a Gram-positive cell wall structure. From the biplots (Figure 3A) we can see that for OP samples there is a clear shift in microbial composition between those without EnzD and those with EnzD.

227 With EnzD the OP samples are highly clustered around *Gemella*, *Rothia*, and *Streptococcus*

228 whereas those without EnzD are dispersed around *Neisseria*, *Prevotella* and *Veillonella.* These

229 observations are consistent with cell wall structure; we would predict that given their rigid cell

230 wall Gram-positive bacteria would drive changes in community composition for those samples

231 with EnzD. *Streptococcus*, *Gemella,* and *Rothia* explain a large amount of the variability in the

232 OP samples. *Streptococcus* and *Staphylococcus* are Gram-positive taxa explaining a large

233 amount of the variability in the sputum samples. Among the Gram-negative taxa explaining a

234 large amount of variability in OP and sputum samples are *Neisseria*, *Veillonella*, and *Prevotella*

235 due to decreasing relative abundance with EnzD. *Granulicatella*, a member of the Firmicutes had

236 decreased RA with EnzD, which was not consistent with other Gram positive taxa.

237 *Granulicatella* is reported to have pleomorphic and variable Gram stain characteristics, which

238 may explain this observation(Bottone et al., 1995; Christensen & Facklam, 2001).

239   Difference in OP and sputum samples can be due to differences in the microbial

240 communities which may be related to multiple factors including anatomic location, inter

241 individual variability and disease severity. The samples used in this study were collected for

242 clinical evaluation of airway infection, although separate evaluation for sample adequacy was

243 not performed, these samples are reflective of what is being used to make clinical decisions. The

244 patients that were not capable of expectorating sputum tend to be younger with presumably less

245 lung disease (i.e., fewer pathogens). There are also differences between upper and lower airways

246 (Zemanick et al., 2015) that could also impact the community composition observed.

247   Willner et al. demonstrated that genera detection varied significantly between 5

248 commonly used extraction methods with *Staphylococcus* being detected with varying efficiency

249 (Willner et al., 2012). Support for these findings showed the largest difference in microbial

250   proportions between two DNA extraction methods (Norgen and Qiagen) was found in

251   *Staphylococcus* (Pérez-Losada et al., 2016). Consistent with previous findings, EnzD revealed a

252   higher RA of *Staphylococcus* in OP samples. Further, when evaluated in both OP swab and

253   sputum samples from pediatric subjects with CF, variations in detection of *Staphylococcus* were

254   observed (Johnson et al., 2016). Total RA for *Staphylococcus* in sputum samples was higher

255   overall, but had less of a change before and after EnzD (Johnson et al., 2016). Zhao et al. looked

256   at sputum and saw an increase in *Staphylococcus*, which we corroborate (Zhao et al., 2012). We

257   also found a higher RA of *Streptococcus* in OP samples. Johnson et al. found that for OP

258   samples DNA concentration consistently increased with EnzD. This indicates that in addition to

259   observing changes in RA, EnzD probably better represents the absolute amounts of genera

260   present in OP samples.

261   *Limitations.* There are certain limitations to our study. First, only two airway sample

262   types collected from subjects with CF were evaluated, thus limiting more general inferences

263   regarding the effects of EnzD in other samples types. However, this study does include OP

264   samples, which have not previously been evaluated and represents an important sample type in

265   early CF. The patient's clinical status at the time of sample collection is unknown, this hampers

266   our ability to determine whether these samples are reflective of both clinical stability or during a

267   pulmonary exacerbation. However, because paired samples were used to assess the effect of

268   EnzD, the difference for each pair in not confounded by clinical status. Because only a single

269   replicate with and without EnzD was evaluated, we are unable to compare effects due to EnzD

270   versus technical variability. However, the observation of the consistent differences for the

271   majority of the sample pairs provides some evidence that the change is due to EnzD rather than

272 technical variability. In our study we considered beta-diversity values of 0.8 or greater to be

273 within the limits of change due to biological variability.

274   The utility of EnzD is difficult to predict due to the strong influence of the community

275 composition in any given sample on the results. There is a compelling argument to utilize EnzD

276 in studies that rely on OP swabs as previous studies have shown increased sensitivity for

277 detection of Staphylococcus. While the argument is less compelling for sputum, EnzD should be

278 applied consistently. Other sample types would require evaluation to determine how much

279 impact is observed. However, increasing the RA of any group makes it harder to identify minor

280 components in the community. EnzD could negatively impact the ability to identify low

281 abundance Gram-negative organisms as seen through the decrease in Gram-negative genera

282 shown in Table 2. This includes many important pathogens in CF (e.g. *Pseudomonas*) and should

283 be considered carefully.

284 **Conclusions**

285   In summary, we found that the use of EnzD on airway samples prior to DNA extraction

286 and sequencing alters microbiome community composition results, likely due to improved

287 detection of gram-positive bacterial taxa (e.g. *Streptococcus* and *Staphylococcus*). The use of

288 EnzD appears particularly important for analysis of OP swab samples. EnzD use with sputum

289 samples may be less critical and has the potential to decrease sensitivity for low abundance taxa.

290 Our findings highlight the need for a consistent approach to airway sample processing and

291 analysis, and suggest that EnzD should be applied routinely in studies using OP swabs and

292 studies requiring sensitive detection of *Staphylococcus* and *Streptococcus*.

293

294   **References**

295   Armstrong DS., Grimwood K., Carlin JB., Carzino R., Olinsky A., Phelan PD. 1996.

296   Bronchoalveolar Lavage or Oropharyngeal Cultures to Identify Lower Respiratory

297   Pathogens in Infants With Cystic Fibrosis. *Pediatric Pulmonology Pedlatr Pulmonol* 21.

298   Benjamini Y., Yekutieli D. 2001. THE CONTROL OF THE FALSE DISCOVERY RATE IN

299   MULTIPLE TESTING UNDER DEPENDENCY. *The Annals of Statistics* 29:1165–1188.

300   Bottone EJ., Thomas CA., Lindquist D., Janda JM. 1995. Difficulties encountered in

301   identification of a nutritionally deficient streptococcus on the basis of its failure to revert to

302   streptococcal morphology. *Journal of clinical microbiology* 33:1022–4.

303   Browder HP., Zygmunt WA., Young JR., Tavormina PA. 1965. Lysostaphin: Enzymatic mode

304   of action. *Biochemical and Biophysical Research Communications* 19:383–389. DOI:

305   10.1016/0006-291X(65)90473-0.

306   Burns JL., Emerson J., Stapp JR., Yim DL., Krzewinski J., Louden L., Ramsey BW., Clausen

307   CR. 1998. Microbiology of sputum from patients at cystic fibrosis centers in the United

308   States. *Clinical infectious diseases : an official publication of the Infectious Diseases*

309   *Society of America* 27:158–63.

310   Christensen JJ., Facklam RR. 2001. Granulicatella and Abiotrophia species from human clinical

311   specimens. *Journal of clinical microbiology* 39:3520–3. DOI: 10.1128/JCM.39.10.3520-

312   3523.2001.

313   Cummings LA., Kurosawa K., Hoogestraat DR., SenGupta DJ., Candra F., Doyle M., Thielges

314   S., Land TA., Rosenthal CA., Hoffman NG., Salipante SJ., Cookson BT. 2016. Clinical

315   Next Generation Sequencing Outperforms Standard Microbiological Culture for

316   Characterizing Polymicrobial Samples. *Clinical Chemistry* 62.

317   Edgar RC., Haas BJ., Clemente JC., Quince C., Knight R. 2011. UCHIME improves sensitivity

318   and speed of chimera detection. *Bioinformatics (Oxford, England)* 27:2194–200. DOI:

319   10.1093/bioinformatics/btr381.

320   Ewing B., Green P. 1998. Base-Calling of Automated Sequencer Traces Using *Phred.* II. Error

321   Probabilities. *Genome Research* 8:186–194. DOI: 10.1101/gr.8.3.186.

322   Gradisar H., Friedrich J., Krizaj I., Jerala R. 2005. Similarities and specificities of fungal

323   keratinolytic proteases: comparison of keratinases of Paecilomyces marquandii and

324   Doratomyces microsporus to some known proteases. *Applied and environmental*

325   *microbiology* 71:3420–6. DOI: 10.1128/AEM.71.7.3420-3426.2005.

326   Hara N., Alkanani AK., Ir D., Robertson CE., Wagner BD., Frank DN., Zipris D. 2012.

327   Prevention of Virus-Induced Type 1 Diabetes with Antibiotic Therapy. *The Journal of*

328   *Immunology* 189:3805–3814. DOI: 10.4049/jimmunol.1201257.

329   Heijerman H. 2005. Infection and inflammation in cystic fibrosis: a short review. *Journal of*

330   *cystic fibrosis : official journal of the European Cystic Fibrosis Society* 4 Suppl 2:3–5. DOI:

331   10.1016/j.jcf.2005.05.005.

332   Hoppe JE., Towler E., Wagner BD., Accurso FJ., Sagel SD., Zemanick ET. 2015. Sputum

333   induction improves detection of pathogens in children with cystic fibrosis. *Pediatric*

334   *pulmonology* 50:638–46. DOI: 10.1002/ppul.23150.

335   Johnson EJ., Zemanick ET., Accurso FJ., Wagner BD., Robertson CE., Harris JK. 2016.

336   Molecular identification of Staphylococcus aureus in airway samples from children with

337   cystic fibrosis. *PLoS ONE* 11:1–12. DOI: 10.1371/journal.pone.0147643.

338 LiPuma JJ. 2010. The Changing Microbial Epidemiology in Cystic Fibrosis. *Clinical*

339   *Microbiology Reviews* 23:299–323. DOI: 10.1128/CMR.00068-09.

340 Lozupone CA., Stombaugh J., Gonzalez A., Ackermann G., Wendel D., Vazquez-Baeza Y.,

341   Jansson JK., Gordon JI., Knight R. 2013. Meta-analyses of studies of the human microbiota.

342   *Genome Research* 23:1704–1714. DOI: 10.1101/gr.151803.112.

343 Markle JGM., Frank DN., Mortin-Toth S., Robertson CE., Feazel LM., Rolle-Kampczyk U., von

344   Bergen M., McCoy KD., Macpherson AJ., Danska JS. 2013. Sex differences in the gut

345   microbiome drive hormone-dependent regulation of autoimmunity. *Science (New York,*

346   *N.Y.)* 339:1084–8. DOI: 10.1126/science.1233521.

347 Pérez-Losada M., Crandall K., Freishtat RJ., Pé Rez-Losada M., Crandall KA. 2016. Comparison

348   of two commercial DNA extraction kits for the analysis of nasopharyngeal bacterial

349   communities APA Citation Comparison of two commercial DNA extraction kits for the

350   analysis of nasopharyngeal bacterial communities. *AIMS Microbiology* 2:108–119. DOI:

351   10.3934/microbiol.2016.2.108.

352 Pruesse E., Peplies J., Glöckner FO. 2012. SINA: accurate high-throughput multiple sequence

353   alignment of ribosomal RNA genes. *Bioinformatics (Oxford, England)* 28:1823–9. DOI:

354   10.1093/bioinformatics/bts252.

355 Quast C., Pruesse E., Yilmaz P., Gerken J., Schweer T., Yarza P., Peplies J., Glöckner FO. 2013.

356   The SILVA ribosomal RNA gene database project: improved data processing and web-

357   based tools. *Nucleic acids research* 41:D590-6. DOI: 10.1093/nar/gks1219.

358   Robertson CE., Harris JK., Wagner BD., Granger D., Browne K., Tatem B., Feazel LM., Park

359   K., Pace NR., Frank DN. 2013. Explicet: graphical user interface software for metadata-

360   driven management, analysis and visualization of microbiome data. *Bioinformatics (Oxford,*

361   *England)* 29:3100–1. DOI: 10.1093/bioinformatics/btt526.

362   Salazar O., Asenjo JA. 2007. Enzymatic lysis of microbial cells. DOI: 10.1007/s10529-007-

363   9345-2.

364   Schindler CA., Schuhardt VT. 1964. Lysostaphin: A New Bacteriolytic Agent for the

365   *Staphylococcus. Proceedings of the National Academy of Sciences of the United States of*

366   *America* 51:414–421. DOI: 10.1073/pnas.51.3.414.

367   Schloss PD., Westcott SL. 2011. Assessing and Improving Methods Used in Operational

368   Taxonomic Unit-Based Approaches for 16S rRNA Gene Sequence Analysis. *Applied and*

369   *Environmental Microbiology* 77:3219–3226. DOI: 10.1128/AEM.02810-10.

370   Trayer HR., Buckley CE. 1970. Molecular properties of lysostaphin, a bacteriolytic agent

371   specific for Staphylococcus aureus. *The Journal of biological chemistry* 245:4842–6.

372   Willner D., Daly J., Whiley D., Grimwood K., Wainwright CE., Hugenholtz P., Tringe S.,

373   Hugenholtz P. 2012. Comparison of DNA Extraction Methods for Microbial Community

374   Profiling with an Application to Pediatric Bronchoalveolar Lavage Samples. *PLoS ONE*

375   7:e34605. DOI: 10.1371/journal.pone.0034605.

376   Wolter DJ., Emerson JC., McNamara S., Buccat AM., Qin X., Cochrane E., Houston LS., Rogers

377   GB., Marsh P., Prehar K., Pope CE., Blackledge M., Deziel E., Bruce KD., Ramsey BW.,

378   Gibson RL., Burns JL., Hoffman LR. 2013. Staphylococcus aureus Small-Colony Variants

379   Are Independently Associated With Worse Lung Disease in Children With Cystic Fibrosis.

380   *Clinical Infectious Diseases* 57:384–391. DOI: 10.1093/cid/cit270.

381   Yuan S., Cohen DB., Ravel J., Abdo Z., Forney LJ. 2012b. Evaluation of Methods for the

382   Extraction and Purification of DNA from the Human Microbiome. *PLoS ONE* 7:e33865.

383   DOI: 10.1371/journal.pone.0033865.

384   Zemanick ET., Wagner BD., Harris JK., Wagener JS., Accurso FJ., Sagel SD. 2010. Pulmonary

385   exacerbations in cystic fibrosis with negative bacterial cultures. *Pediatric pulmonology*

386   45:569–77. DOI: 10.1002/ppul.21221.

387   Zemanick ET., Wagner BD., Robertson CE., Stevens MJ., Szefler SJ., Accurso FJ., Sagel SD.,

388   Harris JK. 2015. Assessment of Airway Microbiota and Inflammation in Cystic Fibrosis

389   Using Multiple Sampling Methods. *Annals of the American Thoracic Society* 12:221–229.

390   DOI: 10.1513/AnnalsATS.201407-310OC.

391   Zhao J., Carmody LA., Kalikin LM., Li J., Petrosino JF., Schloss PD., Young VB., LiPuma JJ.

392   2012. Impact of enhanced Staphylococcus DNA extraction on microbial community

393   measures in cystic fibrosis sputum. *PloS one* 7:e33127. DOI:

394   10.1371/journal.pone.0033127.

395

Table 1

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

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

Table 2

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

OP					
Genera	EnzD %	Non-EnzD %	Estimated change in RA %	P-Value	Corrected p-Value
Streptococcus	36.51	14.81	21.75	<0.01	<0.01
Prevotella	6.62	16.20	-10.90	<0.01	<0.01
Veillonella	9.27	19.22	-7.86	<0.01	<0.01
Neisseria	0.72	1.94	-3.62	<0.01	<0.01
Actinomyces	2.93	0.73	2.50	<0.01	<0.01
Granulicatella	1.80	2.44	-0.62	<0.01	<0.01
Gemella	1.89	0.64	1.23	<0.01	<0.01
Leptotrichia	1.16	3.79	-1.44	<0.01	<0.01
Fusobacterium	1.47	3.08	-0.56	<0.01	<0.01
Rothia	1.70	0.27	1.25	<0.01	<0.01
Haemophilus	0.45	1.69	-0.91	<0.01	<0.01
Johnsonella	0.97	0.40	1.19	<0.01	<0.01
Campylobacter	0.50	0.75	-0.25	<0.01	<0.01
Sputum					
Genera	EnzD %	Non-EnzD %	Estimated change in RA %	P-Value	Corrected p-Value
Streptococcus	27.85	22.48	1.33	0.50	0.50
Staphylococcus	1.39	0.62	5.47	<0.01	<0.01
Prevotella	0.88	3.54	-1.91	<0.01	<0.01
Veillonella	1.09	6.01	-3.55	<0.01	<0.01
Granulicatella	0.53	1.77	-1.30	<0.01	<0.01
Gemella	0.53	1.68	-0.62	<0.01	0.02
Rothia	0.84	0.57	0.21	0.24	0.27

Table 3

Displays the RA for CF pathogens detected in each sample. Samples are highlighted when the difference in relative abundance was at least 1%, red for increase with EnzD and blue with a decrease. The grey rows indicate the median (min, max) for the difference in relative abundance for the highlighted samples.

	Staphylococcus		Achromobacter		Haemophilus		Stenotrophomonas		Pseudomonas	
	Non EnzD	EnzD	Non EnzD	EnzD	Non EnzD	EnzD	Non EnzD	EnzD	Non EnzD	EnzD
					OP					
OP1	0	0.01%	0	0	1.63%	0.34%	0	0	0	0
OP2	0	0.08%	0	0	4.03%	1.56%	0	0	0	0
OP3	0.01%	0.07%	0	0	0.04%	0	0	0	0	0
OP4	0.01%	0.03%	0	0	1.96%	0.56%	0	0	0	0
OP5	0.01%	0.04%	0	0	2.65%	0.83%	0	0	0	0
OP6	0	0.01%	0	0	0.00%	0.01%	0	0.01%	0	0
OP7	0	0	0	0	5.71%	1.46%	0	0	0	0
OP8	0.15%	1.13%	0	0	2.05%	0.65%	0	0	0	0
OP9	0	0.02%	0	0	15.12%	7.11%	0	0	0	0
OP10	0	0.04%	0	0	0.25%	0.08%	0	0	0	0
OP11	0.01%	0.03%	0	0	5.86%	2.03%	0	0	0	0
OP12	0.02%	0.03%	0	0	0.21%	0.11%	0	0	0.01%	0.02%
OP13	0.08%	0.30%	0	0	1.19%	0.22%	0.01%	0	0	0
OP14	0.06%	0.37%	0	0	2.41%	1.43%	0	0	0.03%	0
OP15	0.07%	1.02%	0	0	0.03%	0.04%	0	0	0.01%	0.01%
OP16	0.04%	0.05%	0	0	0.09%	0.02%	0	0	0	0
OP17	0.01%	0.08%	0	0	0.57%	0.31%	0	0	0	0
OP18	0.01%	0.01%	0	0	3.49%	1.87%	0	0	0.01%	0.01%
OP19	0.01%	0.57%	0	0	1.91%	0.57%	0	0	0	0.01%
OP20	0	0	0	0	3.45%	1.79%	0	0	0	0
OP21	0.02%	0.10%	0	0	0.40%	0.18%	0	0	0.01%	0.02%
OP22	0	0	0	0	0.35%	0.05%	0	0	0.01%	0.05%
OP23	0.07%	0.18%	0	0	11.95%	7.84%	0.01%	0	0	0.01%
OP24	0	0	0	0	0.17%	0.03%	0	0	0	0
OP25	0.01%	0.01%	0	0	0.23%	0.07%	0.01%	0.01%	0.02%	0.01%
OP26	2.17%	3.92%	0	0	0.45%	0.04%	0	0	0.01%	0
OP27	0.11%	0.16%	0	0	0	0	0	0	0	0.01%
OP28	0.05%	0	0	0	2.44%	1.82%	0	0	0	0
OP29	0	0	0	0	1.75%	0.67%	0.05%	0.03%	0	0
OP30	0.01%	0.03%	0	0	40.38%	30.81%	0	0	0	0
OP31	0	0	0	0	0.61%	0.14%	0	0	0	0
OP32	0.01%	0.01%	0	0	3.40%	1.09%	0	0	0.07%	0.06%
OP33	0	0	0	0	0.30%	0.04%	0	0	0	0
OP34	3.34%	14.25%	0	0	2.01%	1.19%	0	0	0	0
OP35	0.05%	0.08%	0	0	4.44%	4.27%	0	0	0	0
OP36	0.02%	0.22%	0	0	0.03%	0.01%	0	0	0	0
OP37	0.11%	0.19%	0	0	2.06%	0.58%	0	0	0	0
OP38	0.01%	0.04%	0	0	3.68%	1.54%	0	0	0	0
OP39	0.06%	0.07%	0	0	0	0	0	0	0.01%	0
OP40	0.13%	0.11%	0	0	0.01%	0	0	0	0.01%	0.02%

	Staphylococcus		Achromobacter		Haemophilus		Stenotrophomonas		Pseudomonas	
	Non EnzD	EnzD	Non EnzD	EnzD	Non EnzD	EnzD	Non EnzD	EnzD	Non EnzD	EnzD
OP41	0.05%	0.03%	0	0	1.91%	1.20%	0	0	0	0
OP42	0.01%	0.01%	0	0	0.05%	0.01%	0	0	0.02%	0.01%
	6.3 (1.8, 10.9)		--		-1.8 (-1.1, -9.6)		--		--	
	Sputum									
ES1	2.24%	1.99%	0	0	0	0	0	0	0.66%	0.51%
ES2	0.01%	0.05%	0	0	2.89%	2.79%	0	0	39.12%	79.91%
ES3	0	0	0	0	8.77%	0.83%	0	0	0	0.01%
ES4	0.05%	0.12%	0	0	9.52%	6.47%	0	0	1.27%	1.02%
ES5	34.61%	72.13%	0	0	0	0	0	0	0	0.01%
ES6	43.91%	84.67%	0	0	0.01%	0	0	0	5.04%	1.39%
ES7	89.38%	97.47%	0	0	0	0	0.07%	0.02%	0.01%	0
ES8	16.72%	32.21%	0	0	0	0	0	0	52.01%	9.96%
ES9	0.49%	8.24%	0	0	1.53%	0.71%	0	0	26.46%	33.92%
ES10	0.20%	0.35%	0	0	0.54%	0.09%	0	0	0	0.01%
ES11	31.47%	96.31%	0	0	4.11%	0.23%	0.14%	0.02%	0.01%	0
ES12	0.04%	0.04%	0	0	0	0	0	0	10.08%	3.89%
ES13	26.05%	11.45%	0	0	1.08%	0.72%	0	0	31.81%	58.28%
ES14	0.06%	0.13%	0	0	0	0	0	0	0.02%	0.03%
ES15	16.08%	29.44%	0	0	11.39%	7.50%	0	0	13.04%	14.15%
ES16	1.48%	7.26%	0	0	2.26%	1.25%	0	0	52.82%	63.76%
ES17	67.38%	86.70%	0	0	0.17%	0.02%	0	0	1.69%	0.88%
ES18	0.03%	0.05%	0	0	0	0	0	0	0.01%	0.03%
ES19	0.62%	1.39%	0	0	6.75%	2.09%	11.95%	12.85%	0.01%	0
ES20	88.34%	98.55%	0	0	0	0	0	0	0.01%	0
ES21	0.01%	0.01%	0	0	0	0	56.44%	37.37%	0	0
ES22	1.08%	1.31%	0	0	6.74%	8.94%	0	0	0	0
ES23	0.07%	0.34%	0	0	0.01%	0	0	0	0.16%	0.23%
ES24	0.07%	0.15%	0	0	0.60%	0.28%	0.01%	0	87.24%	81.12%
ES25	0	0	0	0	1.18%	0.41%	2.41%	1.51%	0	0
ES26	41.52%	87.96%	0	0	5.60%	0.64%	0	0	0	0
ES27	64.41%	97.56%	0	0	0.43%	0.02%	2.93%	0.11%	0	0
ES28	0.02%	0.04%	0	0	0.01%	0	0.01%	0.04%	0.02%	0.01%
ES29	73.98%	98.00%	0.01%	0	0	0	0	0	0.05%	0.01%
ES30	72.00%	87.13%	0	0	0	0	0	0	0.18%	0.14%
ES31	90.32%	96.74%	0	0	0	0	0.01%	0.01%	0	0
ES32	0.02%	0.01%	0	0	0.02%	0	33.62%	57.14%	0.07%	0.04%
ES33	49.36%	78.78%	0	0	0	0	0.07%	0	0.03%	0.06%
ES34	0.05%	0.15%	0	0	0	0	0	0	0	0.03%
ES35	1.84%	7.29%	0	0	0.66%	0.40%	0.17%	0.10%	22.94%	21.79%
ES36	0.30%	0.16%	0	0	0.32%	0.02%	0.01%	0.02%	28.07%	96.42%
ES37	0	0.02%	0	0	1.90%	1.22%	0	0	0.18%	0.31%

ES38	0.16%	0.53%	0.08%	0.10%	0.05%	0.10%	0	0	0.01%	0.03%
ES39	0.01%	0	0	0	0	0	0	0	0	0.01%
	15.3 (-14.6, 64.8)		--		-3.9 (-7.9, 2.2)		-2.8 (-19.1, 23.5)		1.1 (-42.0, 68.4)	

Figure 1

Figure 1

Figure 1. Boxplots for individual taxa that change within paired samples. The distribution of differences in paired samples is shown for phyla (A) and genera (B). Taxa 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%. P-values correspond to Benjamini-Hochberg corrected Wilcoxon Signed rank test for paired samples.
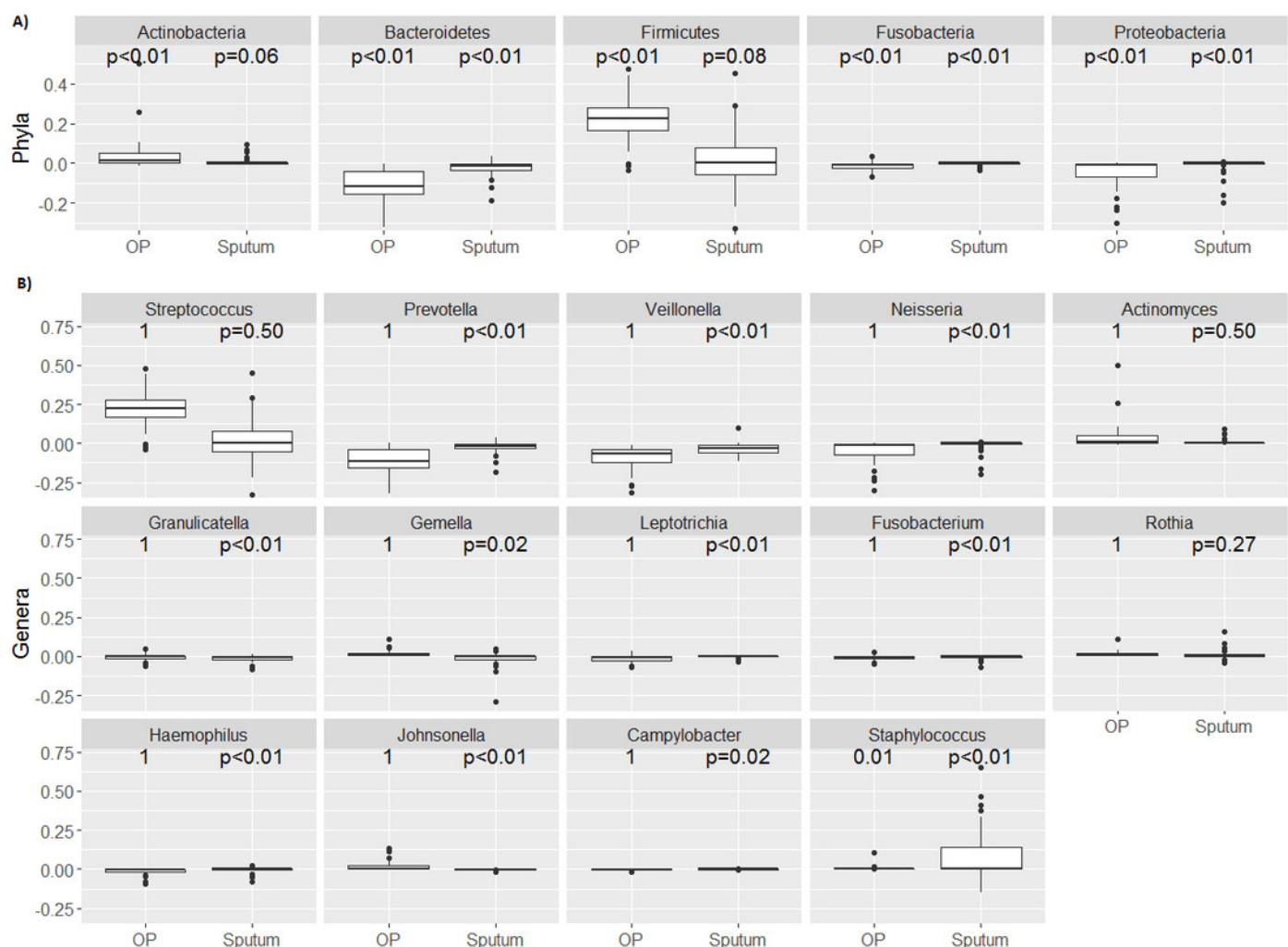
Figure 2

Figure 2

Boxplots showing distribution of Morisita-Horn (MH) similarity metric for paired samples in OP and sputum. Wilcoxon rank sum tests indicate the distributions between the two sample types were statistically different (Benjamini-Hochberg corrected $p<0.01$). The MH metric ranges from 0-1 with a value of 1 indicating identical communities, and 0 indicating no overlap between communities. The line indicates the median value and the box ranges from the 25th and 75th percentiles, the whiskers extend to 1.5 times the interquartile range.
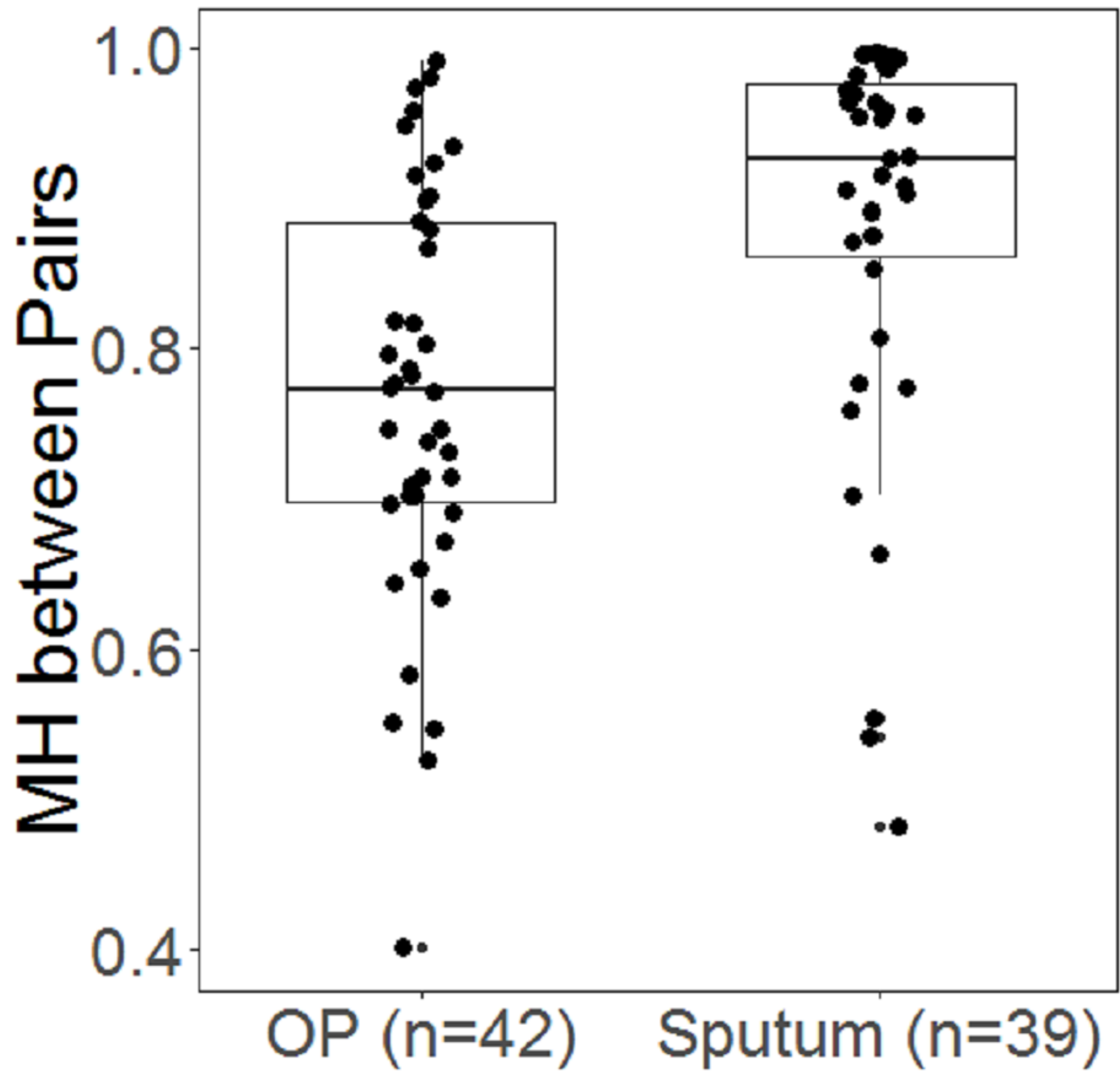
Figure 3

Figure 3

Figure 3. Ordination biplot, using 1-MH beta diversity values for OP (A) and sputum (B). Values closer together are more similar. Paired samples are connected with a line, vectors for the genera with at least 1% RA and $p<0.005$ (by permutation test) for R^2 of the loading.
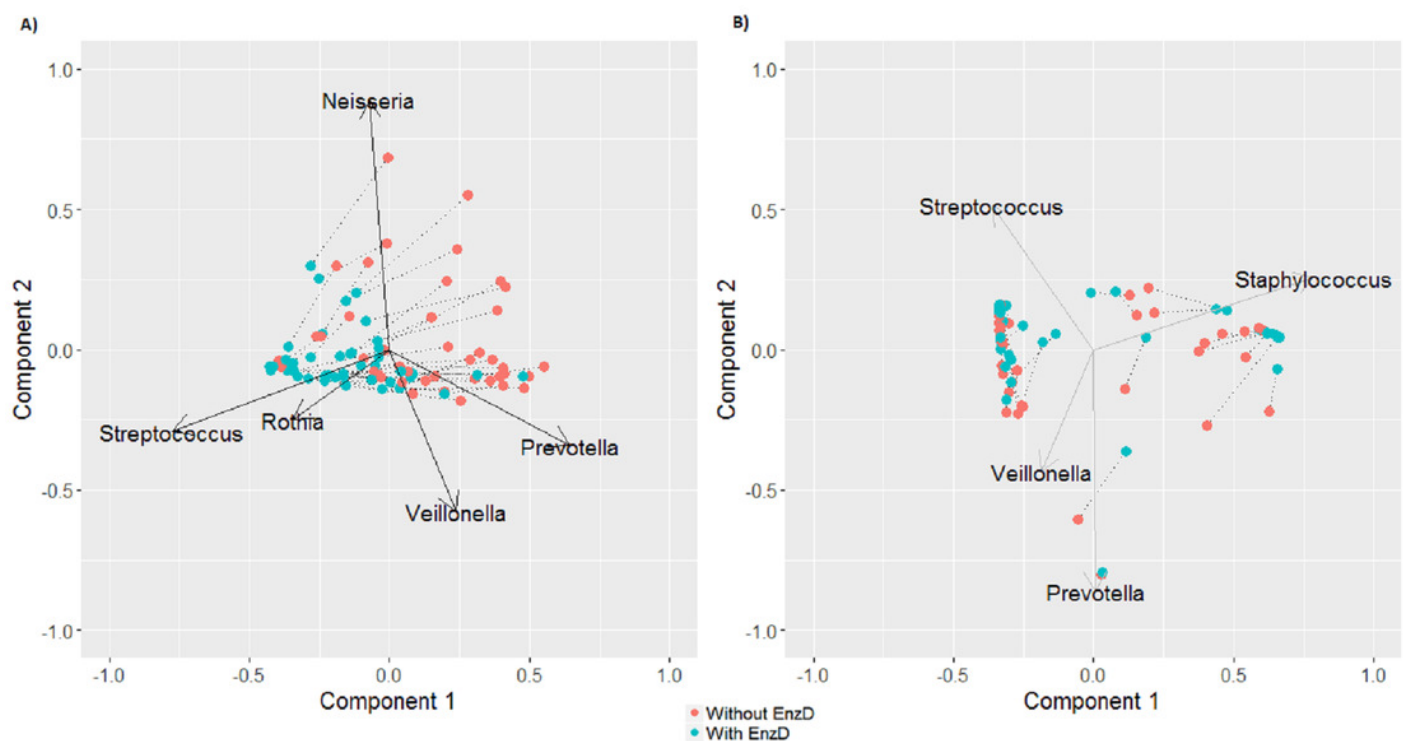
Figure 4

Figure 4

A) Scatterplot of relative abundance of Gram-positive taxa in non-EnzD samples versus Morisita-Horn (MH) similarity metric of paired samples. MH metric ranges from 0-1 with a value of 1 indicating identical communities, and 0 indicating no overlap between communities. OP samples are shown in red and sputum in blue. B) Scatterplot of the difference in relative abundance of Gram-positive taxa versus MH of paired samples OP samples are shown in red and sputum in blue.