1
2  **Profiling of bacterial and fungal microbial communities in cystic fibrosis sputum using RNA**
3
4  Running Title: RNA-based profiling of bacteria and fungi
5

6  Nora Grahl[1#], Emily L. Dolben[1#], Laura M. Filkins[1], Alex W. Crocker[1], Sven D. Willger[1], Hilary G.
7  Morrison[2], Mitchell L. Sogin[2], Alix Ashare[3], Alex H. Gifford[3], Nicholas J. Jacobs[1], Joseph D.
8  Schwartzman[4], and Deborah A. Hogan[1]*
9

10  [1]Microbiology and Immunology, Geisel School of Medicine at Dartmouth, Hanover, NH, USA.

11
12  [2]Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological
13  Laboratory, Woods Hole, MA, USA.

14
15  [3]Pulmonary and Critical Care Medicine, Dartmouth-Hitchcock Medical Center, Lebanon, NH,
16  United States of America

17

18  [4]Department of Pathology and Laboratory Medicine, Dartmouth-Hitchcock Medical Center,
19  Lebanon, NH, USA.
20
21
22  *To whom correspondence should be addressed
23  Department of Microbiology and Immunology, Geisel School of Medicine at Dartmouth
24  Rm 208 Vail Building, Hanover, NH 03755
25  E-mail: dhogan@dartmouth.edu
26  Tel: (603) 650-1252

**Abstract**

Here, we report an approach to detect diverse bacterial and fungal taxa in complex samples by direct analysis of community RNA in one step using NanoString probe sets. We designed rRNA-targeting probe sets to detect forty two bacterial and fungal genera or species common in cystic fibrosis (CF) sputum, and demonstrated taxon-specificity of these probes as well as a linear response over more than three logs of input RNA. Culture-based analyses correlated qualitatively with relative abundance data on bacterial and fungal taxa obtained by NanoString and the analysis of serial samples demonstrated the use of this method to simultaneously detect bacteria and fungi and to detect microbes at low abundance without an amplification step. The relative abundances of bacterial taxa detected by analysis of RNA correlated with the relative abundances of the same taxa as measured by sequencing of the V4V5 region of the 16S rRNA gene amplified from community DNA from the same sample. We propose that this method may complement other methods designed to understand dynamic microbial communities, may provide information on bacteria and fungi in the same sample with a single assay, and, with further development, may provide quick and easily-interpreted diagnostic information on diverse bacteria and fungi at the genus or species level.

**Importance**

44

45    Here we demonstrate the use of an RNA-based analysis of specific taxa of interest, including

46    bacteria and fungi, within microbial communities. This multiplex method may be useful as a

47    means to identify samples with specific combinations of taxa and to gain information on how

48    specific populations vary over time and space or in response to perturbation. A rapid means to

49    measure bacterial and fungal populations may aid in the study of host response to changes in

50    microbial communities.

**Introduction**

51

52        Cystic Fibrosis (CF) is a life-limiting genetic disease associated with chronic lung infection

53    that often leads to progressive lung function decline interspersed with pulmonary

54    exacerbations (1-4). CF lung infections are often polymicrobial, heterogeneous between

55    subjects, variable within individuals over time and can contain bacteria, fungi, and viruses (5-

56    14). The characterization of microbes within CF sputum samples containing diverse bacteria and

57    fungi has been performed using various methods including culturing on different media,

58    sequencing of PCR amplicons for ribosomal RNA genes (rDNA) or ITS (internal transcribed

59    spacer) sequences from bacteria and fungi, respectively, and metagenomic sequencing of

60    community DNA. In this work, we assess the use of the NanoString nCounter technology (15) to

61    detect specific bacterial and fungal taxa in sputum samples from individuals with CF-related

62    respiratory infections.

63        The NanoString methodology detects RNAs using color-coded probe pairs that are

64    optically detected. Together, a fluorescent reporter probe and a capture probe recognize an

65    approximately one hundred base pair sequence within the target RNA. The target RNA-probe

66    complexes are then captured and counted based on their unique fluorescent probe signatures

67    (**Fig. 1**). Barczak *et al.* (16) previously demonstrated the use of this technology to detect

68    species-specific mRNAs of bacteria, parasites, and fungal species and viruses in samples of

69    blood and urine containing one pathogen, but not in clinical samples with complex

70    communities such as CF sputum. In addition, NanoString-based methods were used to monitor

71    seventy-five *Pseudomonas aeruginosa* virulence-associated mRNAs in RNA from CF sputum (17)

4

72    and to detect pathogen mRNAs in samples from single-species mammalian infection models

73    (18-21).

74        We designed and tested taxon-specific NanoString probe sets for the simultaneous

75    detection of forty-two CF-associated bacterial and fungal genera or species by monitoring levels

76    of rRNAs. Probe fidelity was determined using control RNA mixtures as well as by comparison to

77    results obtained by culture and previously established DNA-based community profiling methods.

78    This technology can be used to monitor diverse taxa of interest, including but not limited to,

79    bacterial and fungal targets, in complex microbe-containing samples. We demonstrate its use in

80    monitoring changes in populations over time. Future applications of this approach include the

81    development of improved diagnostics and strategies to monitor changes in microbial

82    communities in response to perturbation.

83

84    **Results**

85    **Probe selection and design**

86        The bacteria and fungi commonly found in CF respiratory sputum have been well-

87    described by culture and deep sequencing methods (5, 6, 14, 22-26). With this knowledge of

88    species likely  to be present, we designed a set of NanoString probes to detect taxon-specific

89    rRNA sequences within total RNA isolated from sputum. In addition to the forty-two bacterial

90    and fungal genera or species that are frequently detected in CF sputum, we included a few

91    other taxa that commonly occur in non-CF respiratory samples (**Table 1**). In some cases, we

92    designed genus-level probes (e.g. *Streptococcus* and *Staphylococcus*) in addition to probes that

93    could detect specific species or complexes within these genera. For other taxa, such as

94    *Burkholderia*, only a genus-level probe was included to streamline analysis and limit costs for

95    this proof of concept study. For each taxon that we chose to monitor, a NanoString probe set

96    was designed to target rRNA sequences from bacterial (16S or 23S rRNA) or fungal taxa (18S or

97    28S rRNA). Two probes that recognized adjacent sequences within a consecutive 100-

98    nucleotide region were designed and one was modified to serve as the capture probe and the

99    other modified to serve as the reporter probe (**Fig. 1**). The sequences for the two probes were

100   designed to have >95% identity to the comparable region in all publically available genomes for

101   the target taxon (inclusivity criteria) and less than 90% sequence identity to any non-ribosomal

102   RNA sequence in the genomes of bacteria, fungi or humans available in Genbank at the time of

103   probe design (exclusivity criteria) as described in the methods. The inclusivity and exclusivity

104   assessments are presented in **Table S1**.

105

106   **Analysis of specificity of probes using mixes of RNA from single-species cultures.**

107        To accompany the computational analysis, which suggested that taxon-targeting rRNA

108   probes would be specific based on exclusivity and inclusivity criteria, we performed control

109   experiments using mixes of total RNA from single species cultures of twelve bacteria and two

110   different fungi to assess potential cross-reactivity. We prepared mixtures of RNA extracted from

111   single species cultures. Each mixture, containing equivalent amounts of RNA from three to five

112   species, was analyzed using the panel of NanoString probe sets. The species RNA included in

113   each mix are indicated in **Fig. 2** (see **Table S2** for strain information).

114        Taxon-specific probe sets gave strong positive signals, measured as counts, when

115   cognate rRNAs were present in the mix (**Fig. 2**). The average positive signal was 1000-fold

6

116 higher than the signals from other probe sets. There were no cases of false positives due to

117 cross-reaction with other probes (counts for all probes shown in **Table S3A**). As shown for

118 Mixes 2 and 5, the *Streptococcus salivarius* RNA reacted with the species specific *Streptococcus*

119 *salivarius* probe (St.s) and the *Streptococcus* spp. probe (St. sp.), while *Streptococcus anginosus*

120 only reacted with the genus level *Streptococcus* spp. probe as no species-specific probe was

121 present. Both *Staphylococcus aureus* and *Staphylococcus epidermidis* reacted with the

122 *Staphylococcus* spp. (S.sp) probe. Two different *Burkholderia* species, *B. multivorans* and *B.*

123 *cenocepacia*, were detected by the *Burkholderia* species probe (B. sp.). Each mixture that

124 included *Pseudomonas aeruginosa* included RNA from a distinct clinical isolate (strains detailed

125 in **Table S1**), and each reacted equally well with the species-specific *P. aeruginosa* probe.

126 *Candida parapsilosis* and *Aspergillus fumigatus* RNAs reacted specifically with their cognate

127 probes, and did not cross-react with probes from other *Candida* or *Aspergillus* species,

128 respectively.

129 To assess the range of counts for which probe set signals were linear, we analyzed RNA

130 Mix 1, which contained RNA from *Pseudomonas aeruginosa*, *Haemophilus influenzae*, and

131 *Staphylococcus aureus*, with total RNA amounts from 0.0005 ng to 5 ng. We found that the

132 probe sets for *P. aeruginosa*, *H. influenzae* and *S. aureus* all yielded the expected linear increase

133 in counts that correlated with sample input with the linear range between 100 and >70,000

134 counts ($R^2$>0.999 for all three taxa) (**Table S3A**).

135

136 **RNA-derived NanoString profiles were consistent across replicates.**

137     As a first step towards evaluating the NanoString panel above for the analysis of

138     microbial communities within sputum, we obtained de-identified sputa that were expected to

139     contain many of the taxa targeted by the probes that we had designed. We extracted total RNA

140     from five sputum samples (sputum samples A-E), and analyzed the RNA in two independent

141     runs. In every case, we found that rRNA counts for each bacterial and fungal taxon strongly

142     correlated between replicates indicating the high reproducibility of the method (**Fig. S1A,**

143     $R^2$>0.999 each sample). The counts from bacteria- and fungus-targeting probes are plotted

144     separately (**Fig. S1A**) as the counts for bacterial RNAs were 100- to 1000-fold higher than those

145     for the fungi suggesting that the bacteria were more abundant. In addition, we separately

146     recovered RNA from homogenized sputum that had been divided into three aliquots (samples F,

147     G and H). Again, we found that the relative abundances of different taxa present in the three

148     replicates from each sample were highly similar (**Fig. S1B,** $R^2$>0.9 in each pairwise comparison

149     per patient**)**.

150     In the NanoString method, probes bind directly to their RNA targets without a cDNA

151     synthesis step or any other enzymatic manipulation. Double-stranded DNA is not available for

152     hybridization (**Fig. 1**)**.** To confirm that the NanoString analysis of rRNA was detecting RNA and

153     not DNA, three sputum-derived RNA samples were analyzed before and after RNase treatment.

154     Total number of counts for all probe sets above background was reduced by >99% for all

155     samples after RNase treatment and the community patterns determined by NanoString were

156     markedly altered after RNase treatment indicating that RNA, and not DNA, was giving rise to

157     the detected counts (**Fig. S2 and Table S3B**).

158

159 **Simultaneous analysis of bacteria and fungi using NanoString in sputum sample series from**

160 **six patients hospitalized for disease exacerbations**

161 One of the impetuses for designing this RNA-based method for complex culture analysis

162 of CF sputum is to overcome several challenges associated with current culture based protocols

163 including 1) the need to analyze CF samples using at least six different culture media (Blood

164 agar, Chocolate Agar, MacConkey Agar, Saboraud Dextrose Agar, Mannitol Salt Agar, and OFPBL

165 (Oxidation/Fermentation-Polymyxin-Bacitracin-Lactose) medium to detect potential pathogens

166 of interest, 2) the inability to readily identify some pathogens, such as *Achromobacter*

167 *xylosoxidans*, by growth characteristics alone, 3) the wide variation in growth kinetics and

168 colony morphology types within a species, and 4) the use of a small, fixed volume of sputum

169 which may limit the detection of less abundant or heterogenously-distributed species. To both

170 assess the accuracy of NanoString 16S rRNA analysis as a technique for detecting species of

171 interest within a mixed community and for assessing changes in sputum microbiota across

172 samples from the same subject, we obtained and analyzed sample series from six hospitalized

173 individuals undergoing treatment for disease exacerbation (designated E1 to E6). In each series,

174 a sample was obtained on the first day of treatment then at least one more and as many as six

175 more samples were obtained on subsequent days. A total of twenty-five sputa were analyzed.

176 In addition to the clinical culture analysis of the Day 0 sample (**Table 2**), all specimens were

177 plated on Sheep Blood Agar (SBA), Pseudomonas Isolation Agar (PIA) and Sabouraud Dextrose

178 Agar (SDA) in our research lab (culture plates shown in **Fig. S3)**. In addition, total RNA was

179 extracted from 100 µl of each sputum sample and 18 ng of total extracted sputum RNA was

180     analyzed by NanoString. All probe sets were included in the analysis of each sample. **Fig. 3A**

181     presents data for all positive signals.

182        The analysis of sputum RNA from the E1 series showed high counts for the fungus

183     *Exophilia dermatitidis* (*E.d.*) with no counts above background for any other fungi; the counts

184     for *E. dermatitidis* were ~100-fold higher than for any bacterial species suggesting that *E.*

185     *dermatitidis* was a dominant microbe in these samples (**Fig. 3A**). While *E. dermatitidis* was not

186     evident on culture plates after 48 hours of incubation at 37°C, the characteristic brown colonies

187     of *E. dermatitidis* were clearly abundant on both SBA and SDA after more than a week (**Fig. 3B**).

188     *E. dermatitidis* was in the final clinical culture report of the Day 0 sample (**Table 2**).

189        In the two samples from the E2 series, NanoString counts for *Achromobacter*

190     *xylosoxidans* were ten-fold higher than any other bacterial or fungal species at both time points

191     (**Fig. 3**), and *Achromobacter xylosoxidans* was the only species detected in the clinical

192     microbiology analysis of the Day 0 sample. Small, non-descript white colonies consistent with

193     those formed by *Achromobacter xylosoxidans* were also detected on blood agar, PIA and SBA

194     plates in our culture based studies (**Fig. S3-E2**). Low counts for *Candida albicans* were also

195     detected and yeast colonies were evident on the SDA medium plates (**Fig. S3**).

196        In series E3, the most abundant NanoString signal from CF-associated pathogens was *P.*

197     *aeruginosa* that are routinely reported clinically (**Fig. 3**-green bar); this species was the only

198     species reported in the clinical microbiology analysis of the Day 0 sample and was clearly

199     evident on the Day 0 sample on PIA (**Table 2** and **Fig. S3-E3**). *P. aeruginosa* was not detected by

200     either NanoString or by culture at the later time points. The genus-level *Staphylococcus* probe

201     (black bar) set gave very high counts and the *Staphylococcus aureus* (black hatched bar) probe

202   set yielded ten-fold fewer counts in the Day 0 sample. The presence of non-aureus

203   *Staphylococcus* species was confirmed by the presence of large numbers of non-hemolytic

204   *Staphylococci* on the blood agar plate at the Day 0 time point, and the number of

205   staphylococcal colonies on the blood agar plate decreased concomitant with the reduction in

206   *Staphylococcus* spp. NanoString signal (**Fig. S3-E3**). Interestingly, the *S. aureus* NanoString

207   counts remained very low but constant over the samples examined, suggesting that this signal

208   was not due to cross reaction with other RNA from other *Staphylococcus* spp. that grew on the

209   blood agar plates.

210        In samples from subject E4, counts for *Burkholderia* (pink bar) were high in the RNA

211   from the sputum samples (**Fig. 3**); *Burkholderia* colonies that were slow-growing, characteristic

212   of small colony variants, began to emerge on SBA after 48 h (**Fig. S3-E4**). The weak growth

213   observed on plates from E4 was confirmed to be *Burkholderia* by 16S rRNA gene sequencing of

214   amplified DNA from single isolates and the clinical microbiology reported *Burkholderia*

215   *cenocepacia* as the dominant pathogen with "few" *S. aureus*. *Staphylococcus aureus* counts

216   were elevated in the Day 0 sample, but not above the conservative threshold set in this study

217   (see methods) (**Table S3C**). *Candida* colonies on SDA were more abundant at later time points

218   and the signals from *Candida albicans* and *Candida dubliniensis* increased concomitantly in

219   samples from later days. We confirmed that both *C. albicans* and *C. dubliniensis* were present

220   by sequencing of the ITS1 region amplified from genomic DNA from individual colonies.

221        In samples from series E5, *Pseudomonas aeruginosa* (green bar), *Staphylococcus* spp.

222   (black bar) and *Staphylococcus aureus* (black hatched bar) represented the strongest signals in

223   the early time points (Day 0 and Day 1) (**Fig. 3-E5**). Both *P. aeruginosa* (mucoid) and *S. aureus*

11

224  (MRSA) were detected in the clinical microbiology analysis of the Day 0 sample (**Table 2**). *P.*

225  *aeurignosa* was detected on PIA medium and non-aureus staphylococci were evident on the

226  SBA plate in our laboratory analyses (**Fig. S3-E5**). E5 sample from Day 6 showed *P. aeruginosa*

227  and *S. aureus* signals to be 100-fold lower than on Days 0 and 1, and the counts for

228  *Achromobacter* (*A.x.*-brown bar) were the highest among all taxa at this time point (>7000

229  counts) (**Fig. 3** and **Table S3C**). Small grey colonies that are typical for *Achromobacter* appeared

230  on the PIA and SBA plates at these late time points (**Fig. S3**). *C. albicans* and *C. dubliniensis* were

231  detected by NanoString with increasing levels at later days and this was concordant with the

232  culture analysis (**Fig. S3-E5**). We confirmed that both of these species were present by

233  sequencing of the ITS1 region amplified from DNA from individual colonies. *Candida glabrata*

234  and *Candida parapsilosis* were detected, though at low levels, in E5 by both NanoString and

235  their presence among the colonies on the Day 6 SDA plate was confirmed by sequencing of the

236  ITS1 sequence amplified from the DNA of single isolates. The presence of multiple *Candida* spp.

237  within single CF sputum samples is consistent with previous reports (14).

238      In series E6, the strongest signals were for *P. aeruginosa* (green bar) and

239  *Stenotrophomonas maltophilia* (red bar) with much lower, but positive signals from

240  *Achromobacter xylosoxidans* and *Burkholderia* species (**Fig. 3**). Of these taxa, only *P. aeruginosa*

241  was reported in the clinical microbiology report for Day 0 (**Table 2**). The clinical lab reported

242  isolation of *Staphylococcus aureus* from the E6 Day 0 culture (**Table 2**) and while the NanoString

243  counts were above the highest background count, they were much lower than for the other

244  bacterial pathogens (**Table S3C**). The levels of *S. maltophilia* rRNA increased on days 7 and 10.

245  Bacterial growth from the E6 samples was slow and weak on Blood agar plates, and species

12

246 determination of the colonies that finally emerged was not performed (**Fig. 3-E6**). *Candida*

247 colonies were only observed on Days 0 and 3 and a positive NanoString signal for *Candida*

248 *albicans* was elevated (**Fig. S3**) but below the threshold we had set.

249 Together, these data indicate that all of the pathogens reported upon clinical

250 microbiology analysis of the Day 0 sputum samples were present and abundant in the

251 NanoString data, with the exception of reports of *Staphylococcus aureus* in the clinical cultures

252 from E1 and E6 (**Table 2**). *Staphylococcus aureus* was not evident in our laboratory culture of a

253 separate aliquot of the same sample. The NanoString counts for the predominant bacterial

254 pathogen were high (>3,800 counts) regardless of whether the strains grew robustly on

255 medium, such as *P. aeruginosa* in E3, or slowly, such as *Burkholderia cenocepacia* in E4,

256 *Pseudomonas aeruginosa* in E6 and *Exophiala* in E1. Together, these data indicate that the

257 NanoString methodology can specifically detect bacteria (E2-E6) and fungi (E1, E2, E4, and E5)

258 within samples using a single assay.

259 In addition, the NanoString approach can provide insights into dynamics of bacterial and

260 fungal species abundance over time or in response to treatment, especially when one taxon is

261 of lower abundance. As one example, in subjects E4 and E5, the number of *Candida* colonies

262 and *Candida* NanoString counts increased over the course of treatment especially when sputum

263 was analyzed on SDA plus gentamicin, a medium that suppresses the growth of bacteria (**Fig.**

264 **S3**) consistent with previous reports of bacterial suppression of fungal growth in clinical

265 cultures (27). The only fungi detected among these samples were *Exophiala* and *Candida* spp.,

266 and no samples had counts above background from probe sets for other fungi encountered in

267 CF samples (*Aspergillus fumigatus*, *Scedosporium* and *Trichosporon*). Except for samples from

13

268     the E1 series, the total probe counts from all fungal rRNA-specific probe sets were more than

269     1,000x lower than the total counts from probe sets designed to target bacterial rRNAs.

270

271     **Relative abundance of CF pathogens determined by NanoString correlates with Illumina**

272     **sequencing of community DNA and culture-based analysis of sputum**.

273             Eighteen sputum samples collected during outpatient visits were divided for separate

274     RNA and DNA extractions. We compared data obtained by NanoString-analysis of RNA to data

275     obtained by well-validated deep sequencing analysis of 16S rRNA sequences amplified from

276     DNA. Comparison to clinical microbiological culture data was also performed. We only focused

277     on the bacteria in these analyses as the fungi were predicted to be of low abundance in most

278     samples based on the clinical culture analysis. Total RNA was analyzed using the NanoString

279     probes described above and the total DNA was subjected to amplification and sequencing of

280     V4-V5 16S rRNA-encoding genes by Illumina HiSeq (**Fig. 4A**). Clinical culture analysis of the

281     eighteen samples found that most of the bacterial pathogens frequently reported from CF

282     sputum cultures were detected in at least one sample including *P. aeruginosa*, *Achromobacter*

283     *xylosoxidans*, *Burkholderia* spp., *Ralstonia* spp., *Stenotrophomonas maltophilia,* and

284     *Staphylococcus aureus*; two other commonly-reported CF pathogens were not detected

285     *Haemophilus influenzae* and *Mycobacterium* spp. in the clinical microbiology analyses.

286             All of the genera at greater than 5% abundance in the Illumina-based community

287     profiles of CF sputum were the taxa monitored by NanoString probes (**Table S4**). This result was

288     not surprising since the taxa included in the NanoString codeset were selected based on their

289     representation in prior Illumina-based community profiling of CF sputum. On average 92% ± 6%

14

290     of the deep sequencing reads corresponded to taxa that were represented among the

291     NanoString probes (**Table S4A**). For both data sets, the relative abundances were determined

292     by division of the counts, for NanoString, or reads, for HiSeq, for the taxon of interest by the

293     total number of counts or reads. For the HiSeq relative abundances, total reads, including those

294     for taxa that were not assessed by NanoString, were used in the percent abundance calculation.

295     Overall, there was high concordance between the data obtained by the two methods with an

296     average Pearson correlation coefficient of 0.803 across all eighteen samples with all taxa,

297     including both CF-related pathogens and species that are also frequent members of the oral

298     microbiome. Together, these data indicate that there is minimal cross reaction between

299     different NanoString probes. Furthermore, these data show that for CF sputum collected during

300     outpatient visits, NanoString analysis of RNA and Illumina analysis of sequences amplified from

301     DNA yielded similar results. Future studies will determine if data obtained from community

302     RNA versus DNA differs in other contexts, such as over the course of intensive antibiotic

303     treatment that might increase the fraction of non-viable or dead cells.

304         Clinical culture results from analysis of these eighteen samples described twenty-six

305     instances in which one of the above bacterial pathogens was detected by culture and several

306     samples contained two or three pathogens. For twenty-four of the twenty six bacterial

307     pathogen detection events, there was a clear positive signal in the NanoString data that

308     indicated the presence of the cultured pathogen (highlighted in green font with blue

309     background in **Fig. 4B**); there was also a signal for the pathogen in the HiSeq data in each case.

310     In two instances, culture analysis detected only "rare" colonies of *P. aeruginosa* (CF3) or

311     *Stenotrophomonas maltophilia* (CF5), which can mean as few as one colony, and the NanoString

312    signal was not clearly positive (red text with blue background in **Fig. 4B**). HiSeq analysis found

313    *Pseudomonas* or *Stenotrophomonas* to be at very low relative abundance (0.01 and 0.2%) in

314    samples CF3 and CF5, respectively. In one case (CF18), the pathogen in the clinical culture

315    report was *Staphylococcus aureus* and this species was robustly detected by NanoString but at

316    very low levels in the High Seq data (**Fig. 4A**). The reason for this discrepancy is not known, but

317    it is worth noting that the sequencing depth was not aberrantly low for this sample. These

318    data suggest that nucleic acid based-methods, which involve the analysis of larger sample

319    aliquots, may be capable of providing quantitative- or semi-quantitative information on the

320    levels or activity of a taxon and may reinforce clinical microbiology data when few colonies of a

321    species are observed in culture. One benefit of the NanoString method is that no amplification

322    is performed prior to analysis aiding in quantitation. The high concordance between the

323    NanoString, Hi-Seq, and culture data further support the validity of the probes and their use in

324    describing microbial communities.

325

326    **Discussion**

327        Here, we report a RNA-based method for detecting target species within complex

328    microbial communities. We describe this approach for the analysis of sample types that have

329    been well characterized by non-targeted approaches such as the sequencing of the

330    metagenome or of amplified rRNA and ITS regions from community DNA in order to identify the

331    taxa present. While the NanoString technology has been applied to the detection of mRNAs

332    from microbial species within clinical samples or in tissues from animal experiments, we show

333    the application of this technology to the analysis of rRNAs which are highly abundant molecules

334    with sequences that are highly conserved across the genus and species level. In contrast to

335    amplification based methods in which conserved primer-binding sequences are required, the

336    regions detected within the rRNAs can be different for different taxa thus allowing for different

337    taxonomic levels to be detected in the same multiplex reaction. Because of the use of a large

338    (~100 nt) region for hybridization to the two probes, computational analysis of the probes

339    suggests that this method can accommodate some sequence divergence within species or

340    genus, but this was not explicitly tested in these experiments.

341        Because fungi can grow slowly in laboratory cultures, as shown for the *Exophiala* in

342    samples from subject E1, and may be unsuspected in some chronic contexts, a method that

343    detects both bacteria and fungi has significant clinical value. We found that taxon-specific

344    NanoString probe sets successfully detected fungi including *C. albicans*, *C. dubliniensis*, *C.*

345    *glabrata*, *A. fumigatus*, and *Exophiala*. In our initial analyses using the NanoString method, we

346    included probes that were complementary to ITS1 RNA sequences in addition to probes that

17

347    bound fungal 28S rRNA (Table S1 for sequences). The ITS1 RNA counts were over 100-fold less

348    abundant suggesting that ITS1 RNA, which is processed from rRNA containing RNAs, is less

349    stable and thus less abundant than rRNAs.

350        We demonstrate significant concordance between this method and Illumina-based

351    sequencing of amplified 16S rRNA sequences from community DNA for many of the organisms

352    examined (**Fig. 4A**) including *Pseudomonas*, *Achromobacter*, *Burkholderia*, *Ralstonia*,

353    *Stenotrophomonas, Staphylococcus*, *Actinomyces*, *Fusobacterium*, *Gemella*, *Neisseria*, *Rothia*,

354    *Porphyromonas* and *Veillonella*. Consistent with a recent study that showed a relationship

355    between deep sequencing and detection by culture when extended culture techniques were

356    used (12), we found correlations between NanoString data and clinical microbiology culture

357    assessment as well.

358        This method is most useful in the analysis of sample sets that have already been well-

359    characterized by non-targeted approaches such as the sequencing of the metagenome or of

360    amplified rRNA and ITS regions from community DNA. Once microbial community members of

361    interest have been identified within a particular habitat, the number of tools available for

362    gaining additional data on specific taxa within the community increases to include quantitative

363    PCR, fluorescent in situ hybridization methods, and plate-based studies, among other methods.

364    We propose that the NanoString method be added to the list of techniques for the analysis of

365    specific taxa within a community and that it may be particularly useful for analyzing the

366    relationships between microbes from different domains, which cannot be profiled using single

367    primer sets, and between species within a genus which may lack resolution by other methods.

368   Multiple probes for each taxon that target both small and large subunit rRNAs may enhance

369   confidence in the detection of low abundance species.

370        In addition to the rRNA analyses described here, we have published a NanoString

371   analysis of various *P. aeruginosa* mRNAs in clinical samples, and the inclusion of species- or

372   strain-specific mRNAs could be useful in analyzing phylogenetically close taxa (17). The

373   combination of probes for rRNAs and mRNAs, however, could be challenged by the difficulties

374   in detecting lower abundance mRNA signals when rRNAs, which are orders of magnitude more

375   abundant. Interfering complementary RNAs may be useful in reducing signals from highly

376   abundant RNAs in order to enable the simultaneous analysis of probes for RNAs that are

377   present in vastly different amounts, but we have not explored this strategy as part of this work.

378   With further development, it is also possible to monitor host and microbial transcripts at the

379   same time to better understand the interplay between host and microbe, particularly during

380   changes in health.

381        The agreement between NanoString and 16S rDNA deep sequencing data for most taxa

382   was striking considering the differences in target molecule (rRNA versus rRNA-encoding DNA)

383   and the technological differences between the methodologies. A number of environmental

384   community profiling studies have compared results from the analysis of extracted rRNA and

385   rDNA sequences often with the goal of determining if RNA can be used to identify the more

386   metabolically active members of the community (28-31). Ribosomal RNA, while more stable

387   than mRNA, is degraded by stable, active RNases and thus it is tempting to speculate that the

388   presence of RNA may serve as a more sensitive indicator of an active cell than the presence of

389   DNA (32). If this is the case, the high concordance between NanoString and HiSeq data would

19

390     suggest that the majority of microbes in the CF sputum samples from these stable patients are

391     alive. While our data do not address whether RNA and DNA equivalently reflect the presence of

392     live and dead cells, with taxon-specific probes in hand, studies that test this idea can now be

393     performed.

394         This detection method, like any nucleic acid analysis method, has the potential for

395     methodological bias. Complicating factors for the NanoString technology could include variable

396     lysis, differences in transcription frequency of rRNA genes in different taxa, and differences in

397     probe hybridization efficiency, thus validation strategies must be employed. While the bead

398     beating based methods used here have been shown to be efficient at lysing fungi and other

399     difficult-to-lyse taxa, the nucleic acid extraction methods used need to be assessed for other

400     taxa that were not present in our sample sets such as *Mycobacteria*.

401         The studies described here represent the first step in employing the NanoString

402     technology to the characterization of complex microbial communities using rRNA. In addition to

403     analyzing and monitoring bacteria and fungi, host markers could also be added to the code sets

404     to gain knowledge about the host state and response (33-36). Furthermore, the NanoString

405     technology has already been introduced into clinical practice for the purpose of cancer cell

406     profiling. While many additional validation experiments would need to be performed before a

407     clinical diagnostic tool could be developed, the quick sample turnaround time and semi-

408     quantitative outputs may allow for more rapid assessment of samples (as compared to

409     traditional culture techniques, which may take days to weeks) as well as protocols for the

410     assessment of the efficacy of treatments. This method provides direct quantitation of rRNAs

411     present in a sample and thus can give insight into the number of rRNA molecules by sample

412    weight, volume, or total RNA. Future studies are necessary to determine if rRNA levels correlate

413    with meaningful metrics such as cell number or population activity.

414        One application of this methodology is the identification of community RNAs that

415    contain significant levels of specific mixed-species microbial populations for use in more

416    extensive analyses such as total RNA sequencing. In addition, the development of a sensitive

417    methodology for the detection of fungi in human-derived samples can be readily adapted to

418    address other research questions such as those regarding the roles of fungi in ulcerative colitis

419    or asthma and reactive airway disease (37-41) as well as in environmental settings. Lastly, the

420    development of taxon-specific probes that detect microbial rRNAs sets the stage for future

421    studies aimed at understanding whether an analysis of microbial rRNA can provide an

422    indication of the metabolic activity or pathogenic potential of specific microbial populations

423    within complex communities.

424

425    **Materials and Methods**

426    **Probe design**. We detected microbial rRNAs using custom-designed species- and genus-specific

427    probe sets. The probe sequences were designed by the NanoString probe design team to

428    recognize all available sequences in Genbank for the taxon of interest and to not cross hybridize

429    to off-target sequences in other bacteria, other fungi, or human genomes. All probe set

430    sequences are listed in **Table S1**. Measurements were made using the NanoString nCounter

431    system (NanoString Technologies, Seattle, WA) as previously described (15). As described in

432    **Table S1**, the probes were purchased in batches, referred to as code sets and the code set

433    composition was slightly modified over time to include additional taxa or to achieve different

21

434  levels of specificity (for example replacing a genus level probe with a species level probe). Each

435  probe contained either capture and detection elements. Each code set also contained six

436  synthetic positive control probes that detect spiked transcripts added at a range of different

437  concentrations and eight negative control probes that are not predicted to hybridize to any

438  transcripts in the bacterial, fungal or human transcriptomes analyzed (NanoString Technologies).

439

440  **Analysis of RNA from in vitro-grown cultures.**

441  To assess the specificity of the RNA probes for specific species, equal amounts of purified total

442  RNA from each microorganism was mixed, and 5 ng of the RNA mixture was used). The strains

443  used in this study are listed in **Table S2**. RNA from laboratory cultures was prepared from cells

444  harvested from 1-3 ml of liquid culture for all strains. *Pseudomonas aeruginosa* was grown in

445  lysogeny broth (LB) with aeration on a roller drum. *Achromobacter xylosoxidans,*

446  *Staphylococcus aureus, Burkholderia cenocepacia,* and *Mycobacterium fortuitum* were grown in

447  tryptic soy broth (TSB) with aeration on a roller drum. *Prevotella melaninogenica* was grown

448  under anoxic conditions using a GasPak EZ anaerobic container system and *Streptococcus*

449  *salivarius,* and *Streptococcus anginosus* were grown statically in an atmosphere with 5% $CO_2$ in

450  Todd Hewitt broth amended with 0.5% yeast extract. *Haemophilus influenzae* was grown in

451  Brain Heart Infusion (BHI) broth supplemented with 10 μg/ml NADH and 10 μg/ml hemin with

452  static incubation in a 5% $CO_2$ atmosphere. All cultures were inoculated from a single colony into

453  5 ml of the specified medium. Liquid cultures were grown at 37°C for 16-18 h except for *M.*

454  *fortuitum* which was grown for 48 h. Cells were frozen and lyophilized prior to RNA extraction

455  using the methods described in the following section.

22

456

457   **RNA and DNA isolation from sputum.** Frozen aliquots (100 µl) were lyophilized for 5-16 h, as

458   necessary. Lyophilized cells were disrupted using a mixture of 0.1 mm, 0.5 mm, and 1 mm

459   zirconia/silicon beads (70 µl each) that was added to the dry pellet prior to agitation. Samples

460   were lysed using 5 cycles of 30 sec, with 20 sec pauses in between, on an Omni Bead Ruptor

461   Homogenizer (Omni International Inc.) set at 5.65 m/s. The disrupted sample was resuspended

462   in Lysis Buffer (200 µl of 0.25 µg/µl lysostaphin and 3 µg/µl lysozyme in 10 mM Tris and 1 mM

463   EDTA (TE) buffer, pH 8.0), and incubated at room temperature for 5-10 min. RNA isolation was

464   then performed using Direct-zol kit (Zymo Research), with TRIzol, according to the

465   manufacturer's suggested protocol. RNA was eluted in two 20 µl volumes of elution buffer

466   before freezing at -80°C until use. No DNAase treatment was performed. For RNA isolation

467   reproducibility experiments, sputa were split into triplicate samples after homogenization, and

468   then processed in parallel following the RNA isolation protocol described above.

469       For DNA isolation from sputum, lyophilized and disrupted sputum aliquots were

470   resuspended in Lysis Buffer (described above) and incubated at room temperature for 5-10 min.

471   Total DNA was isolated using Qiagen genomic DNA purification reagents, following the

472   manufacturer's protocol for Gram-negative bacterial DNA isolation. DNA pellets were

473   resuspended in 40 µl TE buffer and frozen at -80°C until use.

474

475   **rRNA analysis using NanoString probe sets.** We empirically determined that 15-20 ng total RNA

476   was best for balancing the ability to detect less abundant species without overloading the

477   optical detection method. For control mixes of RNA isolated from pure culture, 5 ng total RNA

478   was analyzed, unless otherwise specified. Sample RNA was hybridized to the probe sets for 12-

479   16 h at 65°C using the NanoString nCounter Prep Station instrument according to

480   manufacturer's instructions. Targets were counted on the nCounter using 255 fields of view per

481   sample and output results were extracted using nSolver Analysis Software (NanoString

482   Technologies).

483

484   **Background determination and detection of positive samples.** A series of internal negative

485   control probes that do not target sequences in any known organisms are included in every

486   sample, and the signals associated with these probes inform the relative quality and

487   background hybridization in each individual analysis. Following the manufacturer's

488   recommendation of using limits of detection based on the maximum signal for internal negative

489   control probes, we chose a background subtraction method in which three times the maximum

490   detection by internal negative controls (3xMax) was subtracted from all probe counts. Prior to

491   analysis, a background subtraction of the 3xMax value was performed for all control mixes and

492   patient samples. For comparisons of levels across samples, we normalized samples based on

493   the mean of the positive controls.

494

495   **Patient cohort for CF sputum collection.** The samples were collected in accordance with

496   protocols approved by Center for the Protection of Human Subjects at Dartmouth.

497   Expectorated sputum samples used in the comparison of NanoString and Hi-Seq methods were

498   collected from adult subjects with CF during a routine office visit. Enrolled subjects provided

499   two separate sputum samples during a single visit, one of which was analyzed by standard

500    culture methods for this specimen type at the Dartmouth Hitchcock Medical Center clinical

501    microbiology laboratory. The second sputum sample was divided into aliquots (one for DNA and

502    one for RNA extraction) then frozen on-site using a portable -80°C freezer shuttle (Sterling

503    Ultracold) prior to storage in an upright -80°C freezer. Sputum samples were stored at -80°C for

504    at least 1 h and often for multiple days or weeks before analysis.

505        In serial samples collected from the same patient, samples were collected upon

506    admission for treatment of a disease exacerbation as well as over the course of treatment.

507    Sputum samples were frozen upon collection and stored at -80°C until samples were processed

508    for RNA isolation as described below.

509

510    **DNA deep sequence analysis of bacterial 16S rDNA**. Total genomic DNA (gDNA) isolated from

511    18 sputum samples was used as template for the amplification and analysis of the V4V5 regions

512    of 16S rRNA encoding genes by deep sequencing at the Marine Biological Laboratory (Woods

513    Hole, MA). Sequencing was performed on the HiSeq System (Illumina Inc., San Diego, CA) as

514    previously described (42). Sequencing reads were processed and taxonomic assignments were

515    called using a custom bioinformatic pipeline at the Marine Biological Laboratory. Briefly, raw

516    reads were quality filtered as previously described (43, 44). Chimeric reads were identified and

517    removed using the UChime (45) algorithm and *de novo* reference database modes. Taxon

518    assignments were performed with GAST (43) against a curated SILVA database (46). Operational

519    taxonomic unit (OTU) clustering was determined using UCLUST with a 97% similarity threshold

520    (47). For the comparative analyses with the NanoString methods, a threshold of 0.01% was set

521    for positive signals.

25

522

523    **Analysis of RNase treated RNA.** To determine specificity of NanoString detection for RNA

524    versus DNA, RNAse treatment was performed on 3 samples after RNA isolation (F, G, and H).

525    For these three samples, 1 μl RNaseA was added to 5 μl of isolated RNA and incubated at 37°C

526    for 1 h. The samples were then diluted 1:10 in 45 μl of 10 mM Tris-HCl (pH 7.5), 300 mM NaCl, 5

527    mM EDTA (pH 7.5) and 2 μl of RNase A/T1 mix (Thermo Fisher Cat # EN0551). This mixture was

528    incubated for 1 h at 37°C. Non-treated RNA samples were then diluted 1:10 and equal volumes

529    of RNAse treated and non-treated matched samples were used for the NanoString code set.

530

531    **Culture analysis of serial sputum samples.** Approximately 10 μl of sputum was removed with

532    metal Scienceware Microspoon and plated on Tryptic Soy Agar (TSA) with 5% Sheep Blood

533    (Northeast Laboratory Service, P1100), Pseudomonas Isolation Agar, or Sabouraud Dextrose

534    Agar + 100 μg/ml gentamicin made according to the manufacturer's directions. Plates were

535    incubated for 24 to 48 h at 37°C then imaged.

536        In order to identify isolates that grew on plates, clinical isolates were struck to single

537    colonies, grown as single-colony cultures in either liquid LB medium (bacteria), liquid YPD

538    medium (*Candida* species), or liquid glucose minimal medium (*Aspergillus fumigatus*). Bacterial

539    16S    rDNA    (forward    primer:    GTGSTGCAYGGYTGTCGTCA;    reverse    primer:

540    ACGTCRTCCMCACCTTCCTC) and fungal ITS1 region (forward primer: GTA AAA GTC GTA ACA

541    AGG TTT C; reverse primer: GTT CAA AGA YTC GAT GAT TCA C) were PCR amplified, and the

542    identity of each clinical isolate was determined after sequencing each amplicon using the

543    Applied Biosystems 3730 DNA analyzer (ThermoFisher Scientific).

544

545 **Statistics**. Significance of correlation to determine reproducibility and linearity of microbial

546 detection by NanoString was calculated using linear regression in Graphpad Prism 5. Correlation

547 between NanoString and HiSeq sequencing were using R to determine the Pearson *r* correlation

548 coefficient and its p-value for each subject across selected taxa and for the selected taxa across

549 all subjects using the "cor.test" function in the "stats" package (v3.3.0) with default settings.

550 Significance of correlation between *Candida* spp. colony numbers and NanoString counts and

551 the significance of differences were calculated using linear regression or the Mann-Whitney test

552 in GraphPad Prism 5.

553

554 **Acknowledgements**

565    Health (NIH). The content is solely the responsibility of the authors and does not necessarily

566    represent the official views of the NIH.

567

568    **Literature Cited**

569    1.    Farrell PM. 2000. Improving the health of patients with cystic fibrosis through newborn

570          screening. Wisconsin Cystic Fibrosis Neonatal Screening Study Group. Adv Pediatr 47:79-

571          115.

572    2.    Hoegger MJ, Fischer AJ, McMenimen JD, Ostedgaard LS, Tucker AJ, Awadalla MA,

573          Moninger TO, Michalski AS, Hoffman EA, Zabner J, Stoltz DA, Welsh MJ. 2014. Impaired

574          mucus detachment disrupts mucociliary transport in a piglet model of cystic fibrosis.

575          Science 345:818-22.

576    3.    Stoltz DA, Meyerholz DK, Pezzulo AA, Ramachandran S, Rogan MP, Davis GJ, Hanfland

577          RA, Wohlford-Lenane C, Dohrn CL, Bartlett JA, Nelson GAt, Chang EH, Taft PJ, Ludwig PS,

578          Estin M, Hornick EE, Launspach JL, Samuel M, Rokhlina T, Karp PH, Ostedgaard LS, Uc A,

579          Starner TD, Horswill AR, Brogden KA, Prather RS, Richter SS, Shilyansky J, McCray PB, Jr.,

580          Zabner J, Welsh MJ. 2010. Cystic fibrosis pigs develop lung disease and exhibit defective

581          bacterial eradication at birth. Sci Transl Med 2:29ra31.

582    4.    Abou Alaiwa MH, Reznikov LR, Gansemer ND, Sheets KA, Horswill AR, Stoltz DA, Zabner J,

583          Welsh MJ. 2014. pH modulates the activity and synergism of the airway surface liquid

584          antimicrobials beta-defensin-3 and LL-37. Proc Natl Acad Sci U S A 111:18703-8.

585    5.    Zhao J, Schloss PD, Kalikin LM, Carmody LA, Foster BK, Petrosino JF, Cavalcoli JD,

586          VanDevanter DR, Murray S, Li JZ, Young VB, LiPuma JJ. 2012. Decade-long bacterial

587          community dynamics in cystic fibrosis airways. Proc Natl Acad Sci U S A 109:5809-14.

588    6.    Price KE, Hampton TH, Gifford AH, Dolben EL, Hogan DA, Morrison HG, Sogin ML,

589          O'Toole GA. 2013. Unique microbial communities persist in individual cystic fibrosis

590          patients throughout a clinical exacerbation. Microbiome 1:27.

591    7.    Carmody LA, Zhao J, Schloss PD, Petrosino JF, Murray S, Young VB, Li JZ, LiPuma JJ. 2013.

592          Changes in cystic fibrosis airway microbiota at pulmonary exacerbation. Ann Am Thorac

593          Soc 10:179-87.

594    8.    Salipante SJ, Sengupta DJ, Rosenthal C, Costa G, Spangler J, Sims EH, Jacobs MA, Miller SI,

595          Hoogestraat DR, Cookson BT, McCoy C, Matsen FA, Shendure J, Lee CC, Harkins TT,

596          Hoffman NG. 2013. Rapid 16S rRNA next-generation sequencing of polymicrobial clinical

597          samples for diagnosis of complex bacterial infections. PLoS One 8:e65226.

598    9.    Kim SH, Clark ST, Surendra A, Copeland JK, Wang PW, Ammar R, Collins C, Tullis DE,

599          Nislow C, Hwang DM, Guttman DS, Cowen LE. 2015. Global Analysis of the Fungal

600          Microbiome in Cystic Fibrosis Patients Reveals Loss of Function of the Transcriptional

601          Repressor Nrg1 as a Mechanism of Pathogen Adaptation. PLoS Pathog 11:e1005308.

602   10.    Carmody LA, Zhao J, Kalikin LM, LeBar W, Simon RH, Venkataraman A, Schmidt TM,

603          Abdo Z, Schloss PD, LiPuma JJ. 2015. The daily dynamics of cystic fibrosis airway

604          microbiota during clinical stability and at exacerbation. Microbiome 3:12.

605   11.    Boutin S, Graeber SY, Weitnauer M, Panitz J, Stahl M, Clausznitzer D, Kaderali L,

606          Einarsson G, Tunney MM, Elborn JS, Mall MA, Dalpke AH. 2015. Comparison of

607           microbiomes from different niches of upper and lower airways in children and

608           adolescents with cystic fibrosis. PLoS One 10:e0116029.

609   12.   Cummings LA, Kurosawa K, Hoogestraat DR, SenGupta DJ, Candra F, Doyle M, Thielges S,

610           Land TA, Rosenthal CA, Hoffman NG, Salipante SJ, Cookson BT. 2016. Clinical next

611           generation sequencing outperforms standard microbiological culture for characterizing

612           polymicrobial samples. Clin Chem 62:1465-1473.

613   13.   Hahn A, Sanyal A, Perez GF, Colberg-Poley AM, Campos J, Rose MC, Perez-Losada M.

614           2016. Different next generation sequencing platforms produce different microbial

615           profiles and diversity in cystic fibrosis sputum. J Microbiol Methods 130:95-99.

616   14.   Willger SD, Grim SL, Dolben EL, Shipunova A, Hampton TH, Morrison HG, Filkins LM,

617           O'Toole GA, Moulton LA, Ashare A, Sogin ML, Hogan DA. 2014. Characterization and

618           quantification of the fungal microbiome in serial samples from individuals with cystic

619           fibrosis. Microbiome 2:40.

620   15.   Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, Fell HP, Ferree S,

621           George RD, Grogan T, James JJ, Maysuria M, Mitton JD, Oliveri P, Osborn JL, Peng T,

622           Ratcliffe AL, Webster PJ, Davidson EH, Hood L, Dimitrov K. 2008. Direct multiplexed

623           measurement of gene expression with color-coded probe pairs. Nat Biotechnol 26:317-

624           25.

625   16.   Barczak AK, Gomez JE, Kaufmann BB, Hinson ER, Cosimi L, Borowsky ML, Onderdonk AB,

626           Stanley SA, Kaur D, Bryant KF, Knipe DM, Sloutsky A, Hung DT. 2012. RNA signatures

627           allow rapid identification of pathogens and antibiotic susceptibilities. Proc Natl Acad Sci

628           U S A 109:6217-22.

629    17.    Gifford AH, Willger SD, Dolben EL, Moulton LA, Dorman DB, Bean H, Hill JE, Hampton TH,

630            Ashare A, Hogan DA. 2016. Use of a multiplex transcript method for analysis of

631            *Pseudomonas aeruginosa* gene expression profiles in the cystic fibrosis lung. Infect

632            Immun 84:2995-3006.

633    18.    Fanning S, Xu W, Solis N, Woolford CA, Filler SG, Mitchell AP. 2012. Divergent targets of

634            *Candida albicans* biofilm regulator Bcr1 in vitro and in vivo. Eukaryot Cell 11:896-904.

635    19.    Xu W, Solis NV, Filler SG, Mitchell AP. 2016. Pathogen gene expression profiling during

636            infection using a NanoString nCounter platform. Methods Mol Biol 1361:57-65.

637    20.    Chung D, Barker BM, Carey CC, Merriman B, Werner ER, Lechner BE, Dhingra S, Cheng C,

638            Xu W, Blosser SJ, Morohashi K, Mazurie A, Mitchell TK, Haas H, Mitchell AP, Cramer RA.

639            2014. ChIP-seq and in vivo transcriptome analyses of the *Aspergillus fumigatus* SREBP

640            SrbA reveals a new regulator of the fungal hypoxia response and virulence. PLoS Pathog

641            10:e1004487.

642    21.    Cheng S, Clancy CJ, Xu W, Schneider F, Hao B, Mitchell AP, Nguyen MH. 2013. Profiling of

643            *Candida albicans* gene expression during intra-abdominal candidiasis identifies biologic

644            processes involved in pathogenesis. J Infect Dis 208:1529-37.

645    22.    Fodor AA, Klem ER, Gilpin DF, Elborn JS, Boucher RC, Tunney MM, Wolfgang MC. 2012.

646            The adult cystic fibrosis airway microbiota is stable over time and infection type, and

647            highly resilient to antibiotic treatment of exacerbations. PLoS One 7:e45001.

648    23.    Foundation CF. 2011. Cystic Fibrosis Foundation patient registry 2010 annual data report.

649            Cystic Fibrosis Foundation.

650   24.   Gilligan PH. 2014. Infections in patients with cystic fibrosis: diagnostic microbiology

651         update. Clin Lab Med 34:197-217.

652   25.   Filkins LM, Hampton TH, Gifford AH, Gross MJ, Hogan DA, Sogin ML, Morrison HG,

653         Paster BJ, O'Toole GA. 2012. Prevalence of streptococci and increased polymicrobial

654         diversity associated with cystic fibrosis patient stability. J Bacteriol 194:4709-17.

655   26.   Madan JC, Koestler DC, Stanton BA, Davidson L, Moulton LA, Housman ML, Moore JH,

656         Guill MF, Morrison HG, Sogin ML, Hampton TH, Karagas MR, Palumbo PE, Foster JA,

657         Hibberd PL, O'Toole GA. 2012. Serial analysis of the gut and respiratory microbiome in

658         cystic fibrosis in infancy: interaction between intestinal and respiratory tracts and

659         impact of nutritional exposures. MBio 3.

660   27.   Hockey LJ, Fujita NK, Gibson TR, Rotrosen D, Montgomerie JZ, Edwards JE, Jr. 1982.

661         Detection of fungemia obscured by concomitant bacteremia: in vitro and in vivo studies.

662         J Clin Microbiol 16:1080-5.

663   28.   Kamke J, Taylor MW, Schmitt S. 2010. Activity profiles for marine sponge-associated

664         bacteria obtained by 16S rRNA vs 16S rRNA gene comparisons. ISME J 4:498-508.

665   29.   West NJ, Obernosterer I, Zemb O, Lebaron P. 2008. Major differences of bacterial

666         diversity and activity inside and outside of a natural iron-fertilized phytoplankton bloom

667         in the Southern Ocean. Environ Microbiol 10:738-56.

668   30.   Brinkmann N, Martens R, Tebbe CC. 2008. Origin and diversity of metabolically active

669         gut bacteria from laboratory-bred larvae of *Manduca sexta* (*Sphingidae*, *Lepidoptera*,

670         *Insecta*). Appl Environ Microbiol 74:7189-96.

671    31.    Inkinen J, Jayaprakash B, Santo Domingo JW, Keinanen-Toivola MM, Ryu H, Pitkanen T.

672           2016. Diversity of ribosomal 16S DNA- and RNA-based bacterial community in an office

673           building drinking water system. J Appl Microbiol doi:10.1111/jam.13144.

674    32.    McKillip JL, Jaykus LA, Drake M. 1998. rRNA stability in heat-killed and UV-irradiated

675           enterotoxigenic *Staphylococcus aureus* and *Escherichia coli* O157:H7. Appl Environ

676           Microbiol 64:4264-8.

677    33.    Sweeney TE, Wong HR, Khatri P. 2016. Robust classification of bacterial and viral

678           infections via integrated host gene expression diagnostics. Sci Transl Med 8:346ra91.

679    34.    Wang RY, Weng KF, Huang YC, Chen CJ. 2016. Elevated expression of circulating miR876-

680           5p is a specific response to severe EV71 infections. Sci Rep 6:24149.

681    35.    Pallin DJ, Bry L, Dwyer RC, Lipworth AD, Leung DY, Camargo CA, Jr., Kupper TS, Filbin MR,

682           Murphy GF. 2016. Toward an oobjective diagnostic test for bacterial cellulitis. PLoS One

683           11:e0162947.

684    36.    Jangi S, Gandhi R, Cox LM, Li N, von Glehn F, Yan R, Patel B, Mazzola MA, Liu S, Glanz BL,

685           Cook S, Tankou S, Stuart F, Melo K, Nejad P, Smith K, Topcuolu BD, Holden J, Kivisakk P,

686           Chitnis T, De Jager PL, Quintana FJ, Gerber GK, Bry L, Weiner HL. 2016. Alterations of the

687           human gut microbiome in multiple sclerosis. Nat Commun 7:12015.

688    37.    Wheeler ML, Limon JJ, Bar AS, Leal CA, Gargus M, Tang J, Brown J, Funari VA, Wang HL,

689           Crother TR, Arditi M, Underhill DM, Iliev ID. 2016. Immunological Consequences of

690           Intestinal Fungal Dysbiosis. Cell Host Microbe 19:865-73.

691    38.    Underhill DM, Pearlman E. 2015. Immune interactions with pathogenic and commensal

692           fungi: a two-way street. Immunity 43:845-58.

693   39.   Iliev ID, Funari VA, Taylor KD, Nguyen Q, Reyes CN, Strom SP, Brown J, Becker CA,

694         Fleshner PR, Dubinsky M, Rotter JI, Wang HL, McGovern DP, Brown GD, Underhill DM.

695         2012. Interactions between commensal fungi and the C-type lectin receptor Dectin-1

696         influence colitis. Science 336:1314-7.

697   40.   Goodridge HS, Simmons RM, Underhill DM. 2007. Dectin-1 stimulation by *Candida*

698         *albicans* yeast or zymosan triggers NFAT activation in macrophages and dendritic cells. J

699         Immunol 178:3107-15.

700   41.   Noverr MC, Falkowski NR, McDonald RA, McKenzie AN, Huffnagle GB. 2005.

701         Development of allergic airway disease in mice following antibiotic therapy and fungal

702         microbiota increase: role of host genetics, antigen, and interleukin-13. Infect Immun

703         73:30-8.

704   42.   Eren AM, Vineis JH, Morrison HG, Sogin ML. 2013. A filtering method to generate high

705         quality short reads using illumina paired-end technology. PLoS One 8:e66643.

706   43.   Huse SM, Dethlefsen L, Huber JA, Mark Welch D, Relman DA, Sogin ML. 2008. Exploring

707         microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. PLoS

708         Genet 4:e1000255.

709   44.   Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM. 2007. Accuracy and quality of

710         massively parallel DNA pyrosequencing. Genome Biol 8:R143.

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

712         and speed of chimera detection. Bioinformatics 27:2194-200.

713    46.    Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. 2013.

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

715           based tools. Nucleic Acids Res 41:D590-6.

716    47.    Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST.

717           Bioinformatics 26:2460-1.

718

719    **Table 1.** Taxa detected with the NanoString probe sets

| Bacteria | Fungi |
|---|---|
| *Achromobacter xylosoxidans* | *Aspergillus fumigatus* |
| *Actinomyces odontolyticus* | *Candida albicans* |
| *Anaerococcus prevotii* | *Candida dubliniensis* |
| *Burkholderia* spp. | *Candida glabrata* |
| *Corynebacterium pseudodiptheriticum* | *Candida lusitaniae* |
| *Fusobacterium nucleatum* | *Candida tropicalis* |
| *Gemella haemolysans* | *Candida parapsilosis* |
| *Gemella morbillorum* | *Exophiala dermatitidis* |
| *Haemophilus influenza* | *Malassezia globosa* |
| *Haemophilus parainfluenzae* | *Malassezia restricta* |
| *Mycobacterium* spp. | *Rhodotorula glutinis* |
| *Mycobacterium abscessus* | *Rhodotorula mucilaginosa* |
| *Neisseria* spp. | *Scedosporium apiospermum* |
| *Porphyromonas catoniae* | *Trichosporon mycotoxinivorans* |
| *Prevotella melaninogenica* | |
| *Prevotella oris* | |
| *Pseudomonas aeruginosa* | |
| *Ralstonia pickettii* | |
| *Rothia dentocariosa/mucilaginosa* | |
| *Staphylococcus* spp. | |
| *Staphylococcus aureus* | |
| *Stenotrophomonas maltophilia* | |
| *Streptococcus* spp. | |
| *Streptococcus oralis/mitis/pneumonia* | |
| *Streptococcus anginosus* | |
| *Streptococcus intermedius/constellatus* | |
| *Streptococcus salivarius* | |
| *Veillonella atypica and Veillonella dispar* | |

1  **Table 2. Clinical culture report for Day 0 samples from E-series.** MSSA represents methicillin-
2  sensitive *Staphylococcus aureus* and MRSA represents methicillin-resistant *Staphylococcus*
3  *aureus*.

| Day 0 sample | Clinical culture report |
|---|---|
| E1 | Moderate *Exophiala dermatitidis* and few MSSA |
| E2 | *Achromobacter xylosoxidans* |
| E3 | *Pseudomonas aeruginosa* |
| E4 | *Burkholderia cenocepacia,* MSSA |
| E5 | MRSA, mucoid *Pseudomonas aeruginosa* |
| E6 | MRSA, MSSA, mucoid and non-mucoid *Pseudomonas aeruginosa* |

**Figures**



| Nanostring barcode | Target RNA | Transcript count |
|---|---|---|
| | *P. aeruginosa* 16S | 30,000 |
| | *S. salivarius* 16S | 10,000 |
| | *A. fumigatus* 28S | 10,000 |
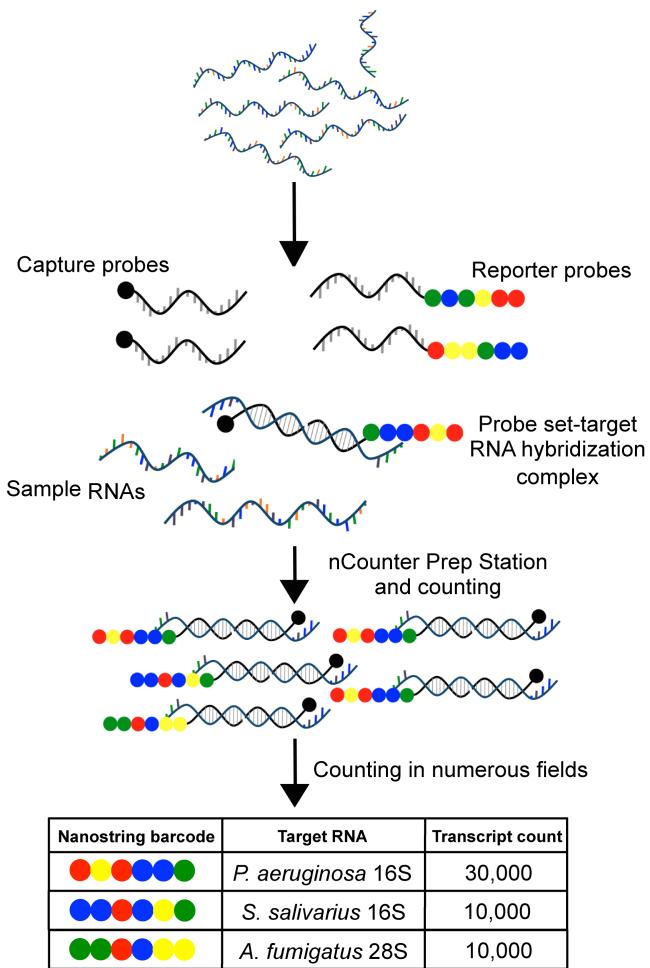
**Fig 1. NanoString detection method schematic.** Frozen samples are lyophilized and total RNA is extracted. Total RNA is mixed with the probe sets, with each containing a reporter and capture probe for each target RNA. After hybridization, the excess probes are removed and the probe-target complexes are aligned and immobilized in the nCounter cartridge. Fluorescent color sequences are counted and tabulated yielding counts that correspond to the number of target molecules.
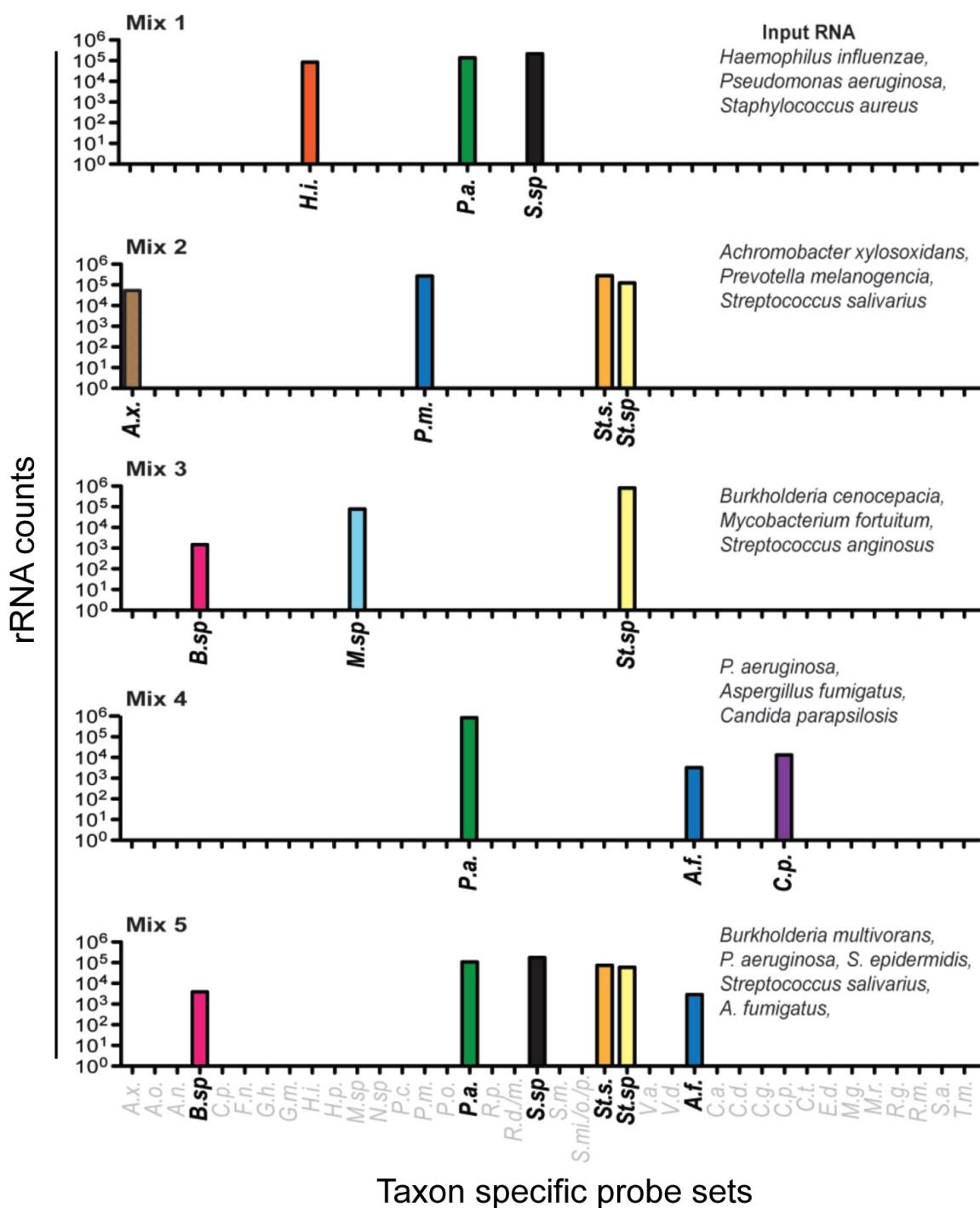
**Fig. 2. Specificity of select bacterial rRNA targeting probe sets.** Control mixes were prepared

with total RNA from three to five species per mix and analyzed using the full panel of probe

sets. Strain details are in Table S2. The probe sets for taxa not present in any of the RNA mixes (in grey font) yielded no counts above background, indicating no cross-reaction with RNAs in the mixes. All data are presented as counts, and all taxon abbreviations are defined in Table S1. The species detected by each probe set is defined by first letters of the genus and species name, except for *Streptococcus*-targeting probe sets which are abbreviated "St." to differentiate from *Staphylococcus*-targeting probe sets which are abbreviated "S". Probe sets that detect multiple species within a genus are indicated by "sp." as for the probe sets that detect *Staphyloccocus* species (*S.sp.*), *Streptococcus* species (*St.sp.*) and *Burkholderia* spp. (*B.sp.*).
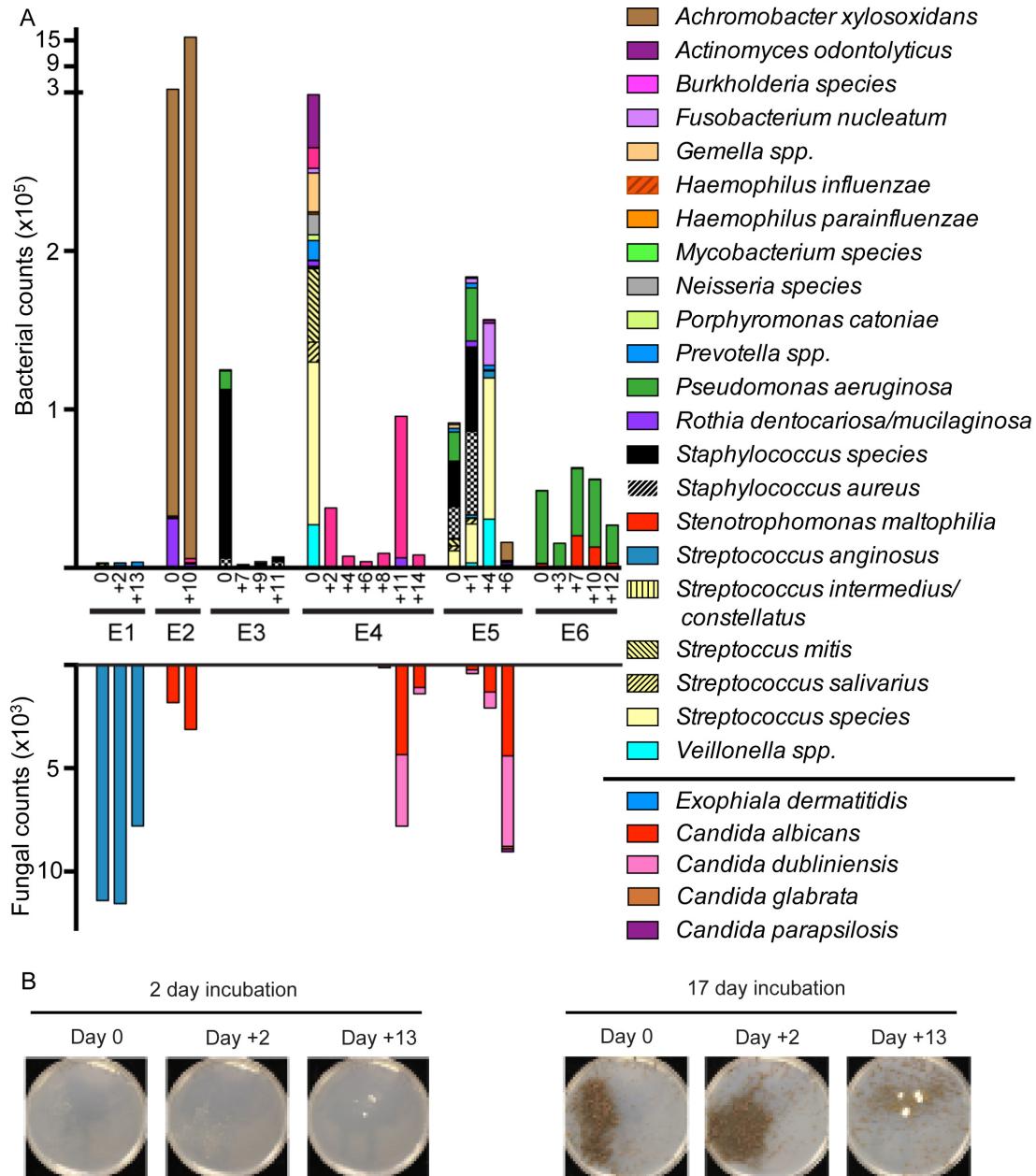
**Fig. 3 Bacterial and fungal community composition in exacerbation series from six subjects.**

**(A)** Serial sputum samples from six individuals with CF were obtained during treatment for pulmonary exacerbation and analyzed by NanoString (data in **Table S3C**). Microbes present in

multiple subjects or abundantly in at least one subject are graphed as total counts normalized to positive controls. Bacteria and fungi are plotted on the upper and lower y-axes, respectively, though the data for bacterial and fungal probe sets were obtained in the same reaction. **(B)** Exacerbation series E1 showed high levels of *Exophiala dermatitidis* by plate culture only after extended incubation.
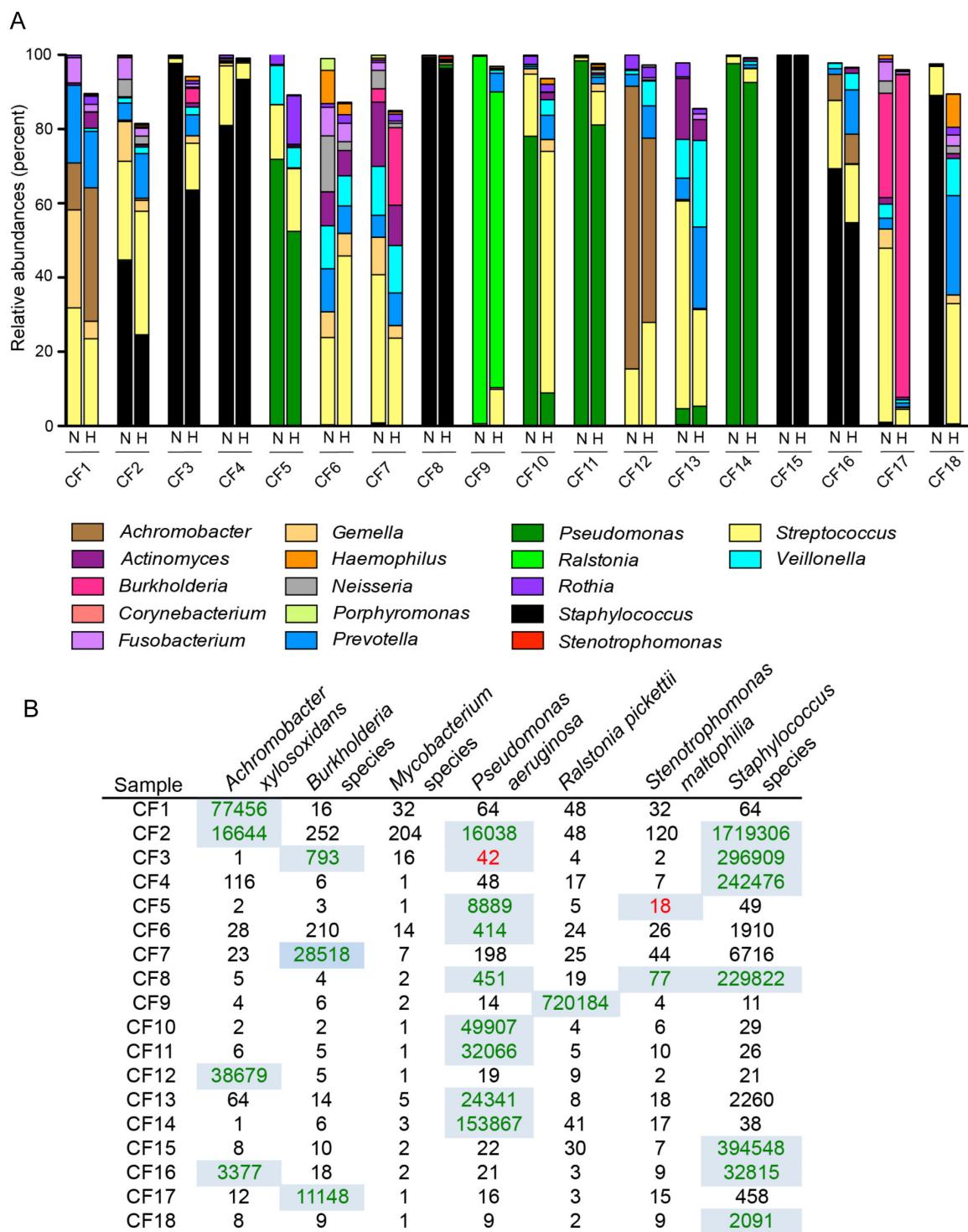
**Fig. 4. Bacterial community profiles detected by NanoString and 16S rDNA deep sequencing (HiSeq) methods correlate. (A)** Eighteen CF sputum samples, labeled with the subject number,

were analyzed by both NanoString (N) and 16S rDNA deep sequencing (HiSeq, H). Percent relative abundance of taxa detected by both methods is shown. Only the taxa measured in our NanoString code set are included in the deep sequencing profiles, but relative abundance was determined as a percentage of total reads, not only reads represented in the NanoString code set (data in Table S4). In cases where NanoString probes detected different species within the same genus, counts from individual species-specific probes were summed to correspond with the genus-level assignments for the HiSeq data. **(B)** The counts for probe sets correlating with CF pathogens are shown. Highlighted numbers indicate that the corresponding pathogen was also detected by clinical culture analysis. The numbers in red in CF3 and CF5 were reported as being "rare" in the clinical microbiology analysis and were at very low levels (<0.5%) in the HiSeq-based community profiles for those samples.

**A**

Legend:

- Achromobacter xylosoxidans
- Actinomyces odontolyticus
- Burkholderia species
- Corynebacterium pseudod.
- Fusobacterium nucleatum
- Gemella haemolysans
- Gemella morbillorum
- Haemophilus influenzae
- Haemophilus parainfluenzae
- Neisseria species
- Porphyromonas catoniae
- Prevotella melaninogenica
- Prevotella oris
- Pseudomonas aeruginosa
- Rothia dentocariosa/mucilagino
- Staphylococcus species
- Stenotrophomonas maltophilia
- Streptococcus species
- Streptococcus mitis group
- Streptococcus salivarius
- Veillonella atypica
- Veillonella dispar

-----------------------------------

- Candida parapsilosis
- Candida albicans
- Candida dubliniensis
- Candida glabrata
- Aspergillus fumigatus

**B**

**Fig. S1 Reproducibility of NanoString analysis of sputum. (A)** Microbial profiles in CF sputum RNA (sputum samples A-E) analyzed in duplicate by NanoString analysis are shown. Bacteria and fungi are presented separately. Correlation of count numbers by NanoString between duplicates for samples (A-E). **(B)** Microbial community profiles presented as percent abundance in three replicate RNAs extracted in parallel from three separate sputum samples (F, G, and H). The legend is the same as in A.
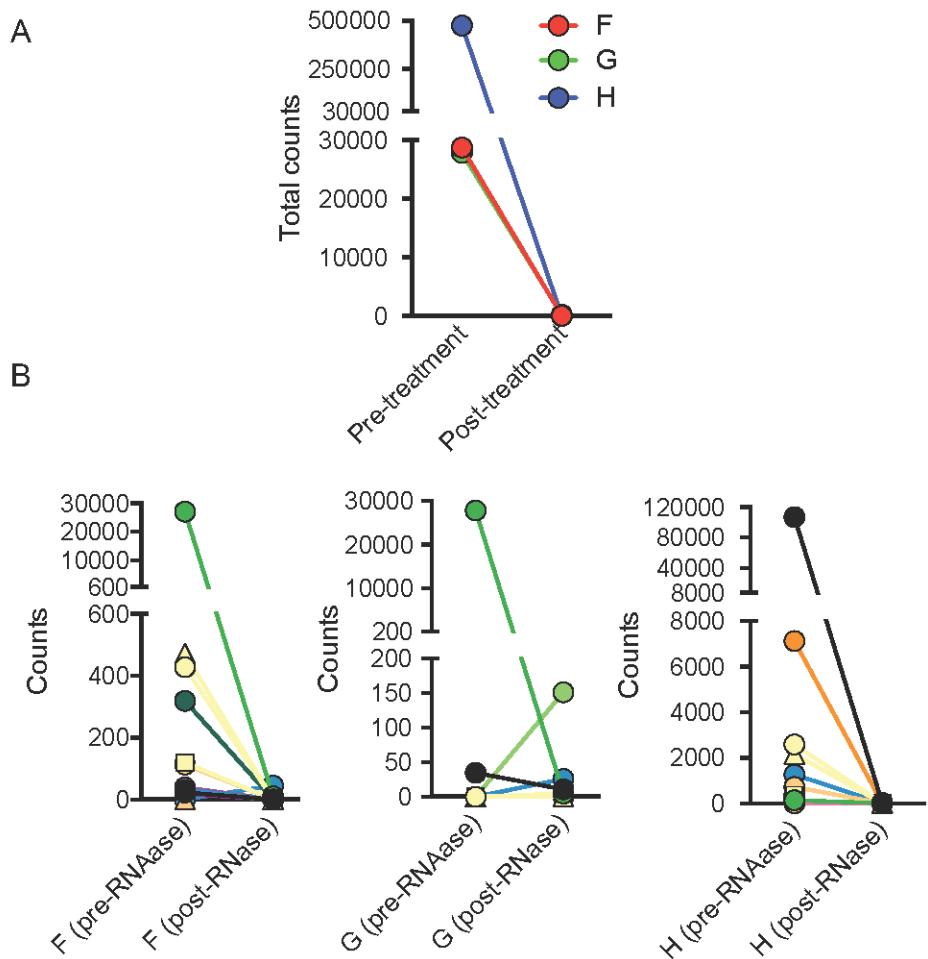
**Fig. S2 RNAse treatment of extracted RNA depletes detection of patient sample microbial communities by NanoString.** Microbial communities were detected from total RNA of three sputum samples (F, G, and H) pre- and post- RNAse treatment with NanoString probes. **(A)** Total detection (sum of counts from all probe sets) from pre- and post- RNAse-treated samples. **(B)**

47

Probe-specific transcript detection pre- and post- RNAse-treatment. Different species/genera

detected are indicated by the different colors and symbols. The data values are presented in

**Table S3B**.
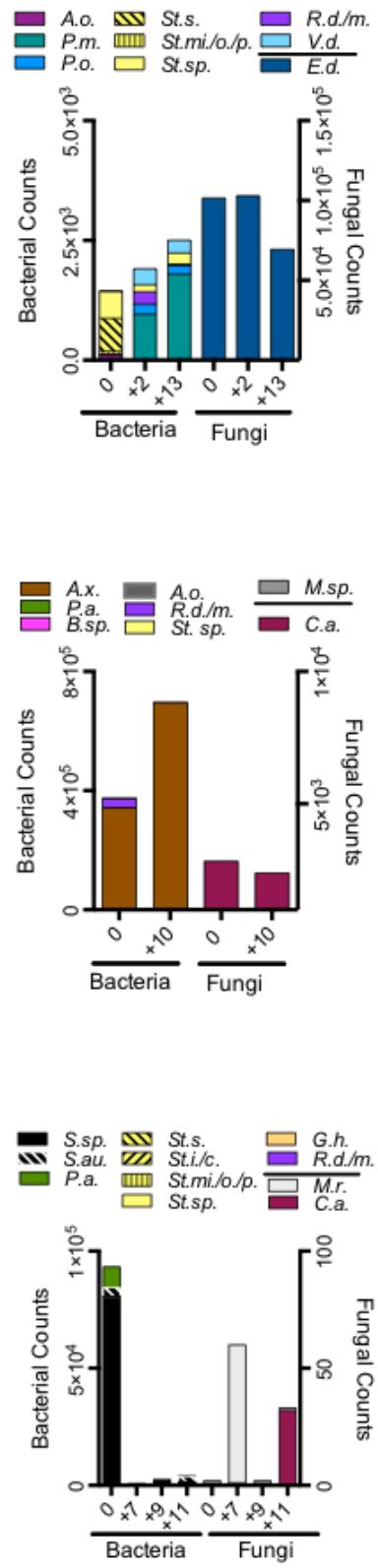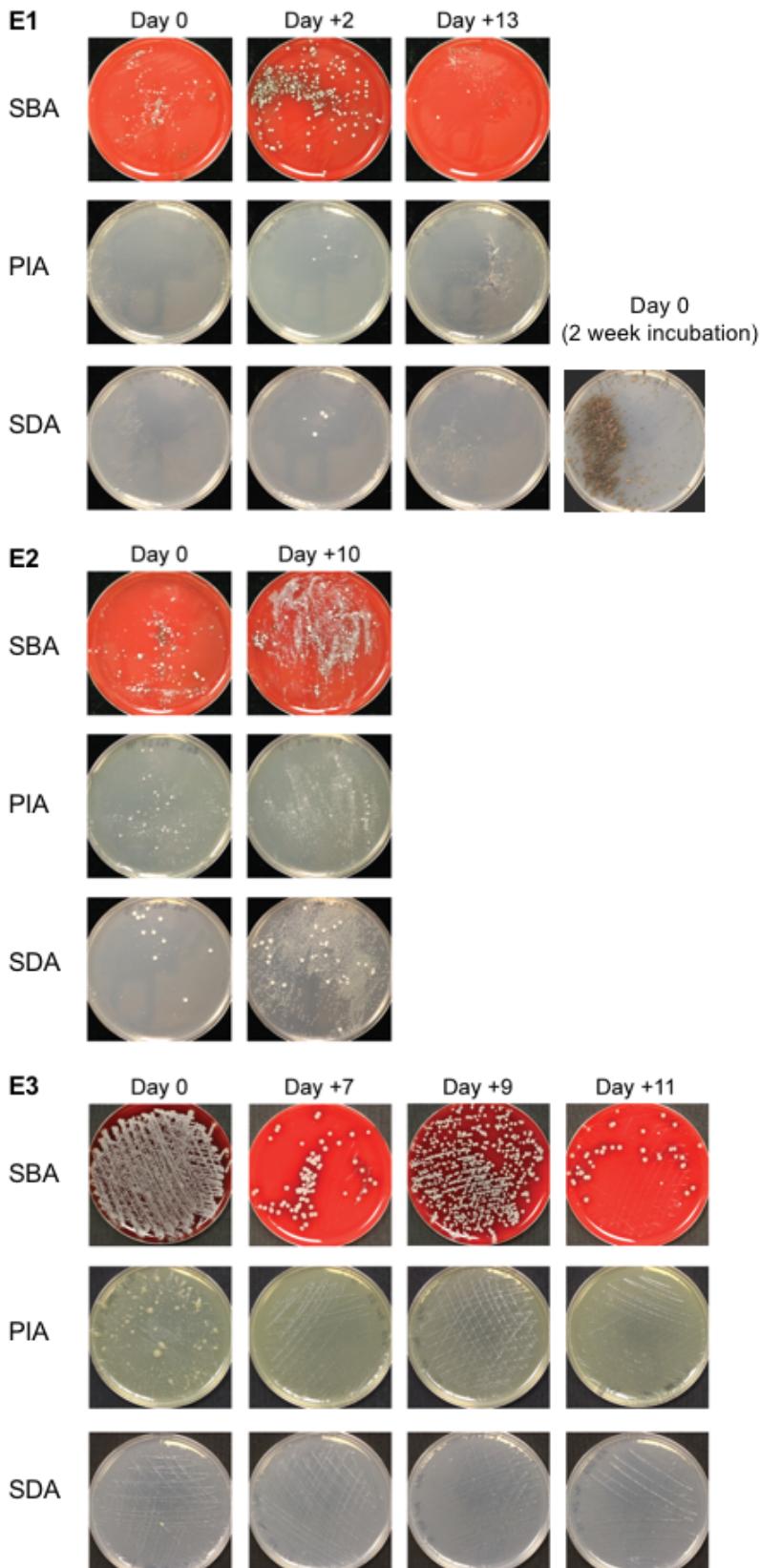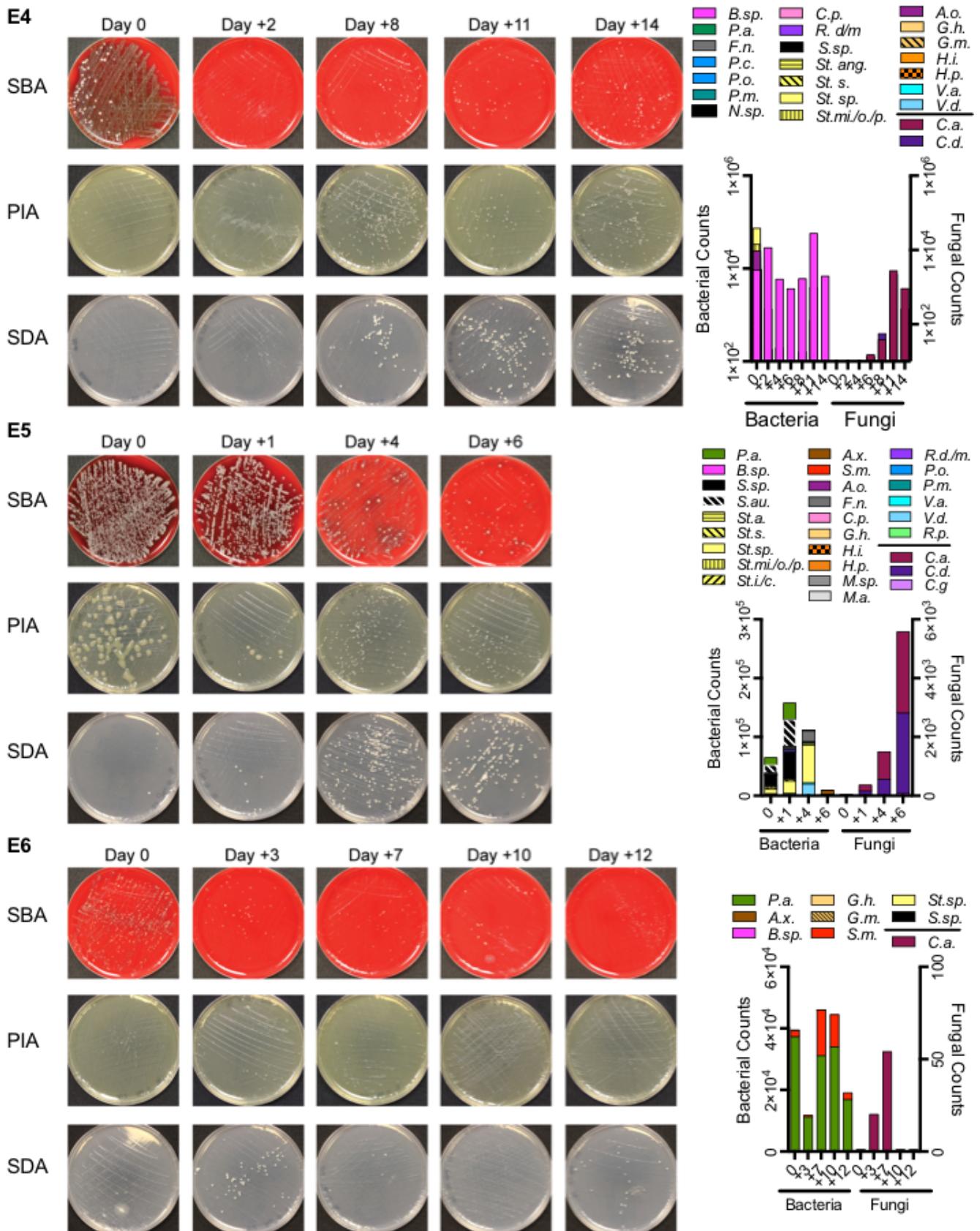
**Fig. S3. Culture analysis and NanoString analysis of sputum series from six hospitalized patients.** Culture of E1-E6 sputum series on sheep blood agar, PIA, and Sabouraud dextrose agar plates. Plates were incubated for at least 48 h before imaging. Samples were collected on Day 0 (date of admission for treatment of exacerbation) and the days post treatment is shown for subsequent samples. For each series, RNA was analyzed by NanoString and data above background are shown for the bacterial and fungus targeting probes included in the code set. Exacerbation series E1 showed high levels of *Exophiala dermatitidis* in plate cultures only after extended incubation for all days and the plate from Day 0 is shown as an example. The clinical culture data for the day 0 samples are shown in **Table 2**. The NanoString data in the graphs for each sample are the same data as shown in Fig. 3, and are presented here for ease of comparison.

**Table S1**. NanoString probe sets designed for this study.

**Table S2**. Strains used in this study.

**Table S3.** All NanoString counts for samples.

**Table S4**. Illumina Reads and Illumina vs NanoString.