Evaluation of Oxford Nanopore MinION™ Sequencing for 16S rRNA Microbiome Characterization

Xiao Ma[1], Elyse Stachler[1], Kyle Bibby[1, 2*]

[1]Department of Civil and Environmental Engineering, University of Pittsburgh, Pittsburgh, PA 15261,

[2]Department of Computational and Systems Biology, University of Pittsburgh Medical School, Pittsburgh, PA 15261, USA

[*]Corresponding author

[*]Corresponding Author: Kyle Bibby, 709 Benedum Hall, Pittsburgh, PA 15261 BibbyKJ@pitt.edu, 412-624-9207

Keywords: Nanopore, microbiome, 16S rRNA, sequencing

Abstract

In this manuscript we evaluate the potential for microbiome characterization by sequencing of near-full length 16S rRNA gene region fragments using the Oxford Nanopore MinION (hereafter 'Nanopore') sequencing platform. We analyzed pure-culture *E. coli* and *P. flourescens*, as well as a low-diversity mixed community sample from hydraulic fracturing produced water. Both closed and open reference operational taxonomic unit (OTU) picking failed, necessitating the direct use of sequences without OTU picking. The Ribosomal Database Project classifier against the Green Genes database was found to be the optimal annotation approach, with average pure-culture annotation accuracies of 93.8% and 82.0% at the phyla and genus levels, respectively. Comparative analysis of an environmental sample using Nanopore and Illumina MiSeq sequencing identified high taxonomic similarity when using a weighted metric (Bray-Curtis), and significantly reduced similarity when using an unweighted metric (Jaccard). These results highlight the great potential of Nanopore sequencing to analyze broad microbial community trends, and the challenge of applying Nanopore sequencing to discern rare taxa in mixed microbial communities. Finally, we observed that between-run carryover accounted for >10% of sequence reads, necessitating future development to either prevent carryover or filter sequences of interest (e.g. barcoding).

Introduction

Interest in studying the microbiome, microbiota associated with various environments, has exploded in recent years largely due to the rapid expansion in 'next-generation sequencing' capabilities and subsequent reduction in cost. It has been recognized that the human microbiome plays an important role in many different clinical outcomes, including obesity (Turnbaugh et al.

46 2006), immune state (Kumar et al.), and infection (Khoruts et al. 2010). The human microbiome
47 is comprised of a diverse array of commensal microorganisms in and on the human body, and
48 emerging research has suggested a clinical role for the microbiome in either therapeutic
49 development (e.g. probiotics) (Good et al. 2014) or diagnostics (Pflughoeft and Versalovic 2012).
50 Concurrently, significant interest has emerged in the microbiomes of various other environments,
51 such as buildings (Adams et al. 2016, Lax and Gilbert 2015, Prussin et al. 2016) and water
52 systems (Baron et al. 2014).

53 Currently, the most common microbiome analysis approach is high-throughput amplicon
54 sequencing of the 16S rRNA gene region. The most widely used sequencing technologies are
55 Illumina sequencing platforms (e.g. the MiSeq and HiSeq). These platforms are accurate and
56 generate a large amount of data, but are limited by capital costs, the necessity to pool samples to
57 reduce per-sample costs, sequence read length, and a turnaround time of days to weeks that may
58 be inadequate for many applications. Recently, Oxford Nanopore has released a small and
59 inexpensive sequencing platform called the MinION™, which has been previously reviewed
60 (Laver et al. 2015, Mikheyev and Tin 2014). Capital costs are reduced to per-run costs that are
61 comparable with the current Illumina platforms, and data analysis is possible in near-real time,
62 enabling investigators to generate sequence data as-needed. For example, samples were correctly
63 assigned to the *Salmonella* species in 20 minutes and serotype in 40 minutes (Quick et al. 2015).
64 Amplicon (non-metagenomic) sequencing has also been successfully employed to identify both
65 bacterial and viral origin (Kilianski et al. 2015, Quick et al. 2016). Additionally, this technology
66 routinely produces sequences >10kb in length, enabling sequencing of the full 16S rRNA gene
67 region and more reliable taxonomic placement. Despite the benefits, the primary drawback to
68 this technology is relatively high error rates (currently reported to be ~8%), hindering

69 metagenomic analysis of highly diverse microbiome samples and requiring additional
70 development and validation. Previous investigations have demonstrated successful 16S rRNA
71 microbial community classification via sequencing of a single mouse gut microbiome (Shin et al.
72 2016) and a single mock microbial community (Benítez-Páez et al. 2016), including species-
73 level assignment (Benítez-Páez et al. 2016); however, a formal assessment of 16S rRNA
74 sequencing on the Nanopore platform, including analysis of pure-culture samples for annotation
75 validation, is currently lacking. Development of Nanopore technology for microbiome analysis
76 would enable rapid (<12 hours from sample to results) and low cost microbiome characterization
77 that would be applicable to both the clinic and other applications.

78

79 In the current study we evaluate the potential for microbiome characterization via Nanopore
80 sequencing of near full-length 16S rRNA PCR amplicons. First, we evaluate the accuracy and
81 annotation strategies of sequences from pure culture *E. coli* and *P. flourescens* to determine the
82 most appropriate sequence analysis approach. We then compare performance of Nanopore
83 sequencing against the current state of the art Illumina sequencing using a sample from hydraulic
84 fracturing wastewater. Finally, we investigate an apparent between-run carryover phenomena,
85 and propose necessary future investigations to enable 16S rRNA microbiome characterization on
86 the Nanopore platform.

87

88 **Materials and Methods**

89 **Overview of DNA samples and sequencing libraries**

90 Three pure culture bacterial DNA samples and an environmental DNA sample were analyzed in
91 the current study. The three pure culture bacterial DNA samples were extracted from pure
92 cultures of: (1) *Escherichia coli* (ATCC 15597), (2) *Pseudomonas fluorescens* (ATCC 13525),
93 and (3) *Mycobacterium smegmatis* str. mc2 155, respectively. The environmental DNA samples,
94 which had previously undergone 16S rRNA gene sequencing using the Illumina MiSeq platform,
95 were a hydraulic fracturing produced water sample (Lipus et al. In Revision) and a river water
96 sample (unpublished).
97
98 All Nanopore sequencing runs were conducted using the MinION Mk IB platform following
99 recommended sequencing protocols (Oxford Nanopore Technologies). The *E. coli* 16S rRNA
100 amplicon 2D library and the hydraulic fracturing produced water 16S rRNA amplicon 2D library
101 were sequenced individually using a Nanopore MIN-105 flow cell and a Nanopore MIN-106
102 flow cell, respectively. The remaining libraries were sequenced on a MIN-106 flow cell
103 following a sequential order: (1) *P. fluorescens* 2D library; (2) *M. smegmatis* 2D library; (3) river
104 water sample 2D library; (4) *M. smegmatis* 1D library. For the sequential sequencing runs, flow
105 cell washing was conducted immediately following the completion of the previous sequencing
106 run using a Nanopore washing kit WSH002 (Oxford Nanopore). The Oxford Nanopore
107 recommended washing protocol was used between runs, namely 150 µL of WSH002 solution A
108 was loaded to the flow cell through priming port and incubated at room temperature for 10
109 minutes, then 150 µL of WSH002 solution B was loaded through the priming port before the
110 next sequencing run and incubated for another 10 minutes at room temperature.
111
112 **Nanopore sequencing library preparation**

113 Previously described universal primers targeting the 16S rRNA gene region (S-D-bact-0008-c-
114 S20 and S-D-bact-1391-a-A-17) (Klindworth et al. 2013) were used for PCR. Each PCR was
115 conducted in a total volume of 50 µL, containing 5 µL 10x buffer, 5 µL dNTP mix, 2.5 µL of
116 each forward and reverse primer, 0.25 µL DreamTaq, 1 µL template DNA, and 33.75 uL
117 nuclease free molecular grade water. The temperature condition for the PCR was 3 minutes at
118 95°C; 30 cycles composed of 20 seconds at 95 °C, 30 seconds at 47 °C for annealing, 1 minute at
119 72 °C; and a final elongation at 72 °C for 15 minutes. All PCR products were purified using
120 Ampure XP beads and normalized to 45 µL containing 1 µg of purified PCR products. Negative
121 controls were used for all PCR reactions and DNA extractions, and all controls were negative.
122
123 2D libraries were prepared using a Nanopore NSK007 sequencing kit and recommended protocol
124 (Oxford Nanopore Technologies). The end repair step of the purified PCR products was
125 conducted by adding 7 µL Ultra II End-Prep buffer, 3 µL Ultra II End-Prep enzyme mix (New
126 England Biolabs), and 5 µL control DNA provided with Nanopore NSK007 sequencing kit. The
127 end repair reaction mix was incubated at 20ºC for 5 minutes and 65ºC for 5 minutes. The end-
128 repaired PCR products were further purified using AMPure XP beads and ligated to the
129 sequencing adapters by adding 8 µL molecular grade water, 10 µL Nanopore NSK007 adapter
130 mix, 2 µL Nanopore NSK007 HPA solution, and 50 µL Blunt/TA Master Mix (New England
131 Biolabs), and then incubated at room temperature for 10 minutes. 1 µL HPT solution from the
132 NSK007 kit was added and incubated for an additional 10 minutes at room temperature. The
133 ligated and tethered 2D libraries were purified by using MyOne C1 beads (Thermo Scientific)
134 and eluted in 25 µL elution buffer (Oxford Nanopore Technologies). A description of the *M.*
135 *smegmatis* 1D library preparation is included in the Supplementary Information.

136

137 All sequencing flow cells were primed using 500 µL Running Buffer Fuel Mix diluted in 500 µL
138 molecular grade water following the recommended priming protocol (Oxford Nanopore
139 Technologies). After priming, 6 µL of each 2D sequencing library was mixed with 37.5 µL
140 Running Buffer Fuel Mix (Oxford Nanopore Technologies) and 31.5 µL molecular grade water,
141 then loaded to Nanopore flow cell for sequencing.

142

143 **Sequence data processing**

144 **Base-calling and initial format conversion**

145 The raw FAST5 files were base-called using Metrichor v2.42.2 with 2D Basecalling for FLO-
146 MIN106 250bps workflow and 1D Basecalling for FLO-MIN106 450bps workflow. Passed 2D
147 reads of 2D sequencing libraries and passed template reads of 1D sequencing libraries were
148 converted to FASTA files for downstream analysis using Poretools (Loman and Quinlan 2014).

149

150 **Operational Taxonomic Unit Evaluation**

151 Operational taxonomic unit (OTU) picking was conducted by using both closed-reference and *de*
152 *novo* picking strategies implemented in QIIME 1.9.2 (Caporaso et al. 2010a). Closed-reference
153 OTU picking was conducted by using Greengenes 13.8 as the reference database (DeSantis et al.
154 2006).

155

156 **Taxonomy Assignment**

157 For pure culture *E. coli* and *P. fluorescens* sequencing data, taxonomy was assigned to each
158 sequence read within QIIME 1.9.0 (Caporaso et al. 2010b) using the RDP classifier (Wang et al.

159 2007) against Greengenes 13.8 (DeSantis et al. 2006) and RDP 16S rRNA training set (Wang et
160 al. 2007) as the reference database, respectively; as well as using BLASTn (Altschul et al. 1990)
161 against Greengenes 13.8 (DeSantis et al. 2006) as a reference database. For subsequent analyses,
162 taxonomy was assigned to each individual sequence read using the RDP classifier (Wang et al.
163 2007) against Greengenes 13.8 (DeSantis et al. 2006) as this approach was found to achieve the
164 highest taxonomy assignment accuracy for the pure culture *E. coli* and *P. fluorescens* sequence
165 data.
166
167 Illumina 16S rRNA amplicon sequencing data of the produced water and river water samples
168 were re-processed using the same approach as Nanopore sequence data. The Illumina data were
169 clustered into OTUs using 100% similarity threshold, i.e. each identical Illumina sequencing
170 read was assigned taxonomy using RDP classifier (Wang et al. 2007) against Greengenes 13.8
171 (DeSantis et al. 2006).
172
173 Jaccard and Bray-Curtis distances, which are dissimilarity distances measuring level of
174 dissimilarity between two microbial communities, were calculated using QIIME 1.9.2 (Caporaso
175 et al. 2010a) to measure the degree of similarity of the produced water microbial community
176 similarity between Nanopore and Illumina platforms. Significance of correlation between the two
177 technical replicates of the produced water sample across sequencing platforms was conducted
178 using Pearson's correlation implemented in Minitab 16.
179
180 **Results and Discussion**
181 ***Sequencing***

182 Sequencing was performed on eight 16S rRNA libraries; however, due to apparent between-run
183 carryover (discussed below), only the initial run on each of three flow cells was used for
184 subsequent analyses of annotation approach and accuracy. These sequencing libraries were
185 derived from two pure-culture samples, *E. coli* and *P. fluorescens*, and a low-diversity sample
186 from hydraulic fracturing produced water. Detailed sequencing results are shown in Table 1.
187 Raw sequence data can be found on Figshare using DOI:
188 https://dx.doi.org/10.6084/m9.figshare.4515752.v3.
189

190 ***Pure Culture Analyses***

191 We first evaluated the suitability of Operational Taxonomic Unit (OTU) picking methods using
192 Nanopore sequence data. OTU picking via comparison with a reference library, i.e. closed-
193 reference OTU picking, failed, with no sequences being assigned to an OTU (i.e. all sequences
194 were excluded). *De novo* OTU clustering, i.e. OTU clustering by determining between-sequence
195 similarity, was then evaluated using similarity thresholds between 90-100%. At the typically
196 used similarity threshold of 97%, all sequences from both pure-culture samples were assigned to
197 unique OTUs (i.e. the ratio of OTUs to sequences was one). Results from *de novo* OTU
198 clustering evaluation are shown in Figure 1. These results highlighted the challenge of clustering
199 reads from long, error-prone sequences, and necessitated analyzing the taxonomy of sequences
200 individually without OTU picking.
201

202 We next evaluated the ability to accurately annotate the taxonomy of pure-culture Nanopore 16S
203 rRNA sequences using three different annotation approaches: the naïve Bayesian Ribosomal
204 Database Project (RDP) classifier with the Green Genes database; the RDP classifier against the

205 RDP database; and BLAST against the Green Genes database. Results from this evaluation are
206 shown in Figure 2. The RDP classifier against the Green Genes database was found to be the best
207 performing annotation strategy. Using this approach, the annotation accuracy for *E. coli* was
208 96.7% and 81.9% at the phyla and genus level, respectively, and was 90.9% and 82.0% for *P.*
209 *fluorescens* at the phyla and genus level, respectively.
210
211 ***Comparison with Illumina Sequencing***
212 We next evaluated Nanopore sequencing to characterize an environmental sample from
213 hydraulic fracturing produced water (Lipus et al. In Revision). This sample was selected as it
214 exhibited low alpha-diversity in previous analyses. The same DNA extract was used for both
215 analyses.
216
217 By assigning taxonomy to each individual sequence, five phyla were identified by the Nanopore
218 platform, and eleven phyla were identified by the Illumina platform. Among them, four phyla
219 that together accounted for greater than 99% of sequence relative abundance were identified by
220 both platforms (Table 2). Nine shared genera were detected by both platforms, accounting for
221 relative abundances of 98.3% on the Nanopore platform and 81.6% on the Illumina platform
222 (Table 2). Both the Nanopore and Illumina platforms revealed similar microbial community
223 structure of the produced water sample. The Firmicutes Phylum dominated the microbial
224 community with relative abundance higher than 90% with both platforms (Table S1). Phyla
225 unique to the Nanopore and Illumina platforms accounted for less than 0.5% of relative
226 abundance (Table S1). At the genus level, the produced water microbial community was
227 dominated by the genus *Halanaerobium* (Table S2), with 95.6% of Nanopore sequences and

228 76.8% of Illumina sequences being assigned to *Halanaerobium*. In addition, 14.6% of Illumina
229 16S rRNA sequence reads were assigned to Clostridiales, which is within the same class with
230 *Halanaerobium* (Clostridia) (Table S2).
231
232 We also calculated the Jaccard distance (solely based on presence/absence of each taxa) and
233 Bray-Curtis distance (based on both presence/absence and relative abundance of each taxa)
234 between the produced water microbial community revealed by Nanopore and Illumina platforms
235 at different taxonomic levels. Jaccard and Bray-Curtis distances measure the degree of microbial
236 community structure difference between two samples, with a value of one indicating no
237 community structure overlap and a value of zero indicating identical microbial communities. We
238 adopted these measures to evaluate the level of difference between the technical replicates of the
239 same produced water sample between the Nanopore and Illumina sequencing platforms. The
240 Jaccard distance increased from 0.62 at the phylum level to 0.89 at genus level; the Bray-Curtis
241 distance increased from 0.04 at phylum level to 0.22 at genus level (Figure 3). Jaccard distances
242 were higher than Bray-Curtis distances at all phylogenetic levels, because more taxa were
243 assigned using short sequence reads by Illumina sequencing (Table 2) and Jaccard distance only
244 accounts for the presence and absence of each assigned taxa whereas Bray-Curtis distances take
245 relative abundance into account.
246
247 Pearson correlation of the relative abundance of each taxon between Nanopore and Illumina data
248 was conducted to further evaluate the reproducibility of sequencing results between the two
249 platforms. Significant correlation was found at all phylogenetic levels from phylum to genus (R

250 values > 0.98, p values < 0.001), indicating reproducible taxonomic assignment results can be
251 obtained between Nanopore and Illumina platforms.
252
253 It should be noted in these comparisons that different primer sets were used for the two analysis
254 approaches, which has previously been shown to bias microbiome community structure as
255 analyzed by 16S rRNA sequencing (Pinto and Raskin 2012). Despite this additional source of
256 bias, the above analyses imply that the weighted community structure is comparable between the
257 two platforms, encouraging future development.
258
259 ***Between-Run Sample Carryover***
260 We noted an apparent carryover of sequences between pure culture runs of *P. fluorescens* and *M.*
261 *smegmatis*. The potential for sequence carryover has been anecdotally reported in the literature
262 (Greninger et al. 2015). We subsequently excluded all runs except the first run on each flow cell
263 from earlier analyses, and undertook a formal evaluation of sequence carryover.
264
265 Results from analysis of the *Mycobacterium smegmatis* run are shown in Figure 4. In this run,
266 76.5% of sequences were assigned to the correct Actinobacteria phyla, 11.0% of sequences were
267 incorrectly assigned to another domain or unable to be assigned, and 12.4% of sequences were
268 incorrectly assigned to the Proteobacteria phyla, presumptively resulting from sequence
269 carryover. 54.5% of sequences were assigned to the correct *Mycobacterium* genus whereas
270 10.8% of sequences were assigned to the *Pseudomonas* genus.
271

272 Following the *P. fluorescens* and *M. smegmatis* runs, we subsequently ran an additional
273 environmental sample derived from river water and re-ran a *M. smegmatis* sample using 1D
274 technology. In the second *M. smegmatis* run, we observed only 0.5% of sequences to be assigned
275 to the correct Mycobacteriaceae family, compared with 54.5% in the first *M. smegmatis* run.
276 These results imply that continued carry-over serves to significantly decrease output quality;
277 however, additional validation with controlled microbial community composition is necessary to
278 confirm this observation.
279
280 ***Areas of Future Development***
281 This investigation has identified multiple necessary areas of future development to enable 16S
282 rRNA microbiome characterization on the Nanopore platform. First, strategies to exclude
283 between-run carryover, either via improved washing between runs or a barcode approach, would
284 enable multiple runs on the same flow cell, significantly reducing per-run costs. In this
285 investigation we performed six runs on the same flow cell while observing minimal output loss.
286 Second, improved bioinformatics strategies are necessary to exclude poor quality sequences. In
287 the *E. coli* and *P. fluorescens* runs, 3.3 and 9.0% of sequences, respectively, were not assigned to
288 any phyla, suggesting poor sequence quality and confounding both alpha- and beta-diversity
289 analyses. Finally, it would be beneficial to develop a 16S rRNA annotation pipeline based upon
290 optimized analysis strategies that provides output in near real-time, facilitating field and clinic
291 applications and alleviating current bioinformatics challenges from interested investigators.
292 Ultimately, the development of a rapid and low-cost microbiome approach will facilitate the
293 application of clinical and environmental microbiome technologies.
294

295 **Acknowledgements**

296 Support for this project was provided by the University of Pittsburgh Central Research
297 Development fund. KB was a member of the Oxford Nanopore Early Access Program, which
298 initially provided access to the sequencing platform at reduced cost.

299

300 **References**

301 Adams, R.I., Bhangar, S., Dannemiller, K.C., Eisen, J.A., Fierer, N., Gilbert, J.A., Green, J.L.,
302 Marr, L.C., Miller, S.L. and Siegel, J.A. (2016) Ten questions concerning the microbiomes of
303 buildings. Building and Environment 109, 224-234.
304 Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment
305 search tool. Journal of Molecular Biology 215, 403-410.
306 Baron, J.L., Vikram, A., Duda, S., Stout, J.E. and Bibby, K. (2014) Shift in the Microbial
307 Ecology of a Hospital Hot Water System following the Introduction of an On-Site
308 Monochloramine Disinfection System. PLoS ONE 9(7), e102679.
309 Benítez-Páez, A., Portune, K.J. and Sanz, Y. (2016) Species-level resolution of 16S rRNA gene
310 amplicons sequenced through the MinION™ portable nanopore sequencer. Gigascience 5(1), 4.
311 Caporaso, J., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F., Costello, E., Fierer, N.,
312 Pena, A., Goodrich, J., Gordon, J., Huttley, G., Kelley, S., Knights, D., Koenig, J., Ley, R.,
313 Lozupone, C., McDonald, D., Muegge, B., Pirrung, M., Reeder, J., Sevinsky, J., Turnbaugh, P.,
314 Walters, W., Widmann, J., Yatsunenko, T., Zaneveld, J. and Knight, R. (2010a) QIIME allows
315 analysis of high-throughput community sequencing data. Nature Methods 7, 335 - 336.
316 Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K.,
317 Fierer, N., Pena, A.G., Goodrich, J.K. and Gordon, J.I. (2010b) QIIME allows analysis of high-
318 throughput community sequencing data. Nature methods 7(5), 335-336.
319 DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi,
320 D., Hu, P. and Andersen, G.L. (2006) Greengenes, a Chimera-Checked 16S rRNA Gene
321 Database and Workbench Compatible with ARB. Applied and Environmental Microbiology
322 72(7), 5069-5072.
323 Good, M., Sodhi, C.P., Ozolek, J.A., Buck, R.H., Goehring, K.C., Thomas, D.L., Vikram, A.,
324 Bibby, K., Morowitz, M.J., Firek, B., Lu, P. and Hackam, D.J. (2014) Lactobacillus rhamnosus
325 HN001 decreases the severity of necrotizing enterocolitis in neonatal mice and preterm piglets:
326 evidence in mice for a role of TLR9. American Journal of Physiology - Gastrointestinal and
327 Liver Physiology 306(11), G1021-G1032.
328 Greninger, A.L., Naccache, S.N., Federman, S., Yu, G., Mbala, P., Bres, V., Stryke, D., Bouquet,
329 J., Somasekar, S., Linnen, J.M., Dodd, R., Mulembakani, P., Schneider, B.S., Muyembe-Tamfum,
330 J.-J., Stramer, S.L. and Chiu, C.Y. (2015) Rapid metagenomic identification of viral pathogens in
331 clinical samples by real-time nanopore sequencing analysis. Genome medicine 7(1), 99.
332 Khoruts, A., Dicksved, J., Jansson, J.K. and Sadowsky, M.J. (2010) Changes in the composition
333 of the human fecal microbiome after bacteriotherapy for recurrent Clostridium difficile-
334 associated diarrhea. Journal of clinical gastroenterology 44(5), 354-360.

Kilianski, A., Haas, J.L., Corriveau, E.J., Liem, A.T., Willis, K.L., Kadavy, D.R., Rosenzweig, C.N. and Minot, S.S. (2015) Bacterial and viral identification and differentiation by amplicon sequencing on the MinION nanopore sequencer. Gigascience 4(12), 10.1186.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M. and Glöckner, F.O. (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Research 41(1), e1-e1.

Kumar, P., Monin, L., Castillo, P., Elsegeiny, W., Horne, W., Eddens, T., Vikram, A., Good, M., Schoenborn, Alexi A., Bibby, K., Montelaro, Ronald C., Metzger, Dennis W., Gulati, Ajay S. and Kolls, Jay K. Intestinal Interleukin-17 Receptor Signaling Mediates Reciprocal Control of the Gut Microbiota and Autoimmune Inflammation. Immunity 44(3), 659-671.

Laver, T., Harrison, J., O'Neill, P., Moore, K., Farbos, A., Paszkiewicz, K. and Studholme, D.J. (2015) Assessing the performance of the Oxford Nanopore Technologies MinION. Biomolecular detection and quantification 3, 1-8.

Lax, S. and Gilbert, J.A. (2015) Hospital-associated microbiota and implications for nosocomial infections. Trends in molecular medicine 21(7), 427-432.

Lipus, D., Vikram, A., Ross, D., Bain, D., Gulliver, D., Hammack, R. and Bibby, K. (In Revision) Predominance and Metabolic Potential of Halanaerobium in Produced Water from Hydraulically Fractured Marcellus Shale Wells. Applied and Environmental Microbiology.

Loman, N.J. and Quinlan, A.R. (2014) Poretools: a toolkit for analyzing nanopore sequence data. Bioinformatics 30(23), 3399-3401.

Mikheyev, A.S. and Tin, M.M. (2014) A first look at the Oxford Nanopore MinION sequencer. Molecular ecology resources 14(6), 1097-1102.

Pflughoeft, K.J. and Versalovic, J. (2012) Human microbiome in health and disease. Annual Review of Pathology: Mechanisms of Disease 7, 99-122.

Pinto, A.J. and Raskin, L. (2012) PCR Biases Distort Bacterial and Archaeal Community Structure in Pyrosequencing Datasets. PLoS ONE 7(8), e43093.

Prussin, A.J., II, Vikram, A., Bibby, K.J. and Marr, L.C. (2016) Seasonal Dynamics of the Airborne Bacterial Community and Selected Viruses in a Children?s Daycare Center. PLoS ONE 11(3), e0151004.

Quick, J., Ashton, P., Calus, S., Chatt, C., Gossain, S., Hawker, J., Nair, S., Neal, K., Nye, K. and Peters, T. (2015) Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of Salmonella. Genome Biol 16(114.2015), 10.1186.

Quick, J., Loman, N.J., Duraffour, S., Simpson, J.T., Severi, E., Cowley, L., Bore, J.A., Koundouno, R., Dudas, G., Mikhail, A., Ouédraogo, N., Afrough, B., Bah, A., Baum, J.H.J., Becker-Ziaja, B., Boettcher, J.P., Cabeza-Cabrerizo, M., Camino-Sánchez, Á., Carter, L.L., Doerrbecker, J., Enkirch, T., Dorival, I.G., Hetzelt, N., Hinzmann, J., Holm, T., Kafetzopoulou, L.E., Koropogui, M., Kosgey, A., Kuisma, E., Logue, C.H., Mazzarelli, A., Meisel, S., Mertens, M., Michel, J., Ngabo, D., Nitzsche, K., Pallasch, E., Patrono, L.V., Portmann, J., Repits, J.G., Rickett, N.Y., Sachse, A., Singethan, K., Vitoriano, I., Yemanaberhan, R.L., Zekeng, E.G., Racine, T., Bello, A., Sall, A.A., Faye, O., Faye, O., Magassouba, N.F., Williams, C.V., Amburgey, V., Winona, L., Davis, E., Gerlach, J., Washington, F., Monteil, V., Jourdain, M., Bererd, M., Camara, A., Somlare, H., Camara, A., Gerard, M., Bado, G., Baillet, B., Delaune, D., Nebie, K.Y., Diarra, A., Savane, Y., Pallawo, R.B., Gutierrez, G.J., Milhano, N., Roger, I., Williams, C.J., Yattara, F., Lewandowski, K., Taylor, J., Rachwal, P., J. Turner, D., Pollakis, G., Hiscox, J.A., Matthews, D.A., Shea, M.K.O., Johnston, A.M., Wilson, D., Hutley, E., Smit, E., Di Caro, A., Wölfel, R., Stoecker, K., Fleischmann, E., Gabriel, M., Weller, S.A., Koivogui, L.,

381 Diallo, B., Keïta, S., Rambaut, A., Formenty, P., Günther, S. and Carroll, M.W. (2016) Real-time,
382 portable genome sequencing for Ebola surveillance. Nature 530(7589), 228-232.
383 Shin, J., Lee, S., Go, M.-J., Lee, S.Y., Kim, S.C., Lee, C.-H. and Cho, B.-K. (2016) Analysis of
384 the mouse gut microbiome using full-length 16S rRNA amplicon sequencing. Scientific Reports
385 6, 29681.
386 Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R. and Gordon, J.I. (2006)
387 An obesity-associated gut microbiome with increased capacity for energy harvest. Nature
388 444(7122), 1027-1131.
389 Wang, Q., Garrity, G.M., Tiedje, J.M. and Cole, J.R. (2007) Naïve Bayesian Classifier for Rapid
390 Assignment of rRNA Sequences into the New Bacterial Taxonomy. Applied and Environmental
391 Microbiology 73(16), 5261-5267.

392

393

Table 1. Sequencing results and output from Oxford Nanopore MinION runs.

Sample	Sequencing Chemistry	Total Reads	Passed Reads (%)	Average Q-Score of Pass Reads (±SD)	Average Read Length (±SD)	Run Time (Min)
E. coli	2D	4093	1447 (35.4%)	11.95 ± 1.44	1304.23± 302.70	15
P. fluorescens	2D	7374	2671(36.2%)	13.73 ± 1.67	1277.99 ± 321.87	20
M. smegmatis[*]	2D	11627	4502(38.7%)	13.56 ± 1.7	1263.30 ± 300.29	20
Produced Water	2D	16525	5461 (33.0%)	13.79 ± 1.66	1320.96 ± 215.06	20
River Water	2D	9745	3143 (32.3%)	13.0 ± 1.76	1233.62 ± 385.52	20
M. smegmatis	1D	10383	7463 (71.9%)	7.82 ± 0.78	1414.44 ± 697.89	20

[*]*M. smegmatis* 2D, river water sample, and *M. smegmatis* 1D library were excluded from primary analyses due to apparent carryover from previous runs.

394

395

396

Table 2. Comparison of number of taxa at different phylogenetic levels for the produced water sample

	Phylum	Class	Order	Family	Genus

Number of taxa assigned using Nanopore data	5	9	13	17	20
Number of taxa assigned using Illumina data	11	21	36	63	84
Number of shared taxa	4	8	11	13	9
Total relative abundance of shared taxa (Nanopore)	99.9%	100.0%	99.9%	98.5%	98.3%
Total relative abundance of shared taxa (Illumina)	99.4%	98.0%	96.3%	81.8%	81.6%

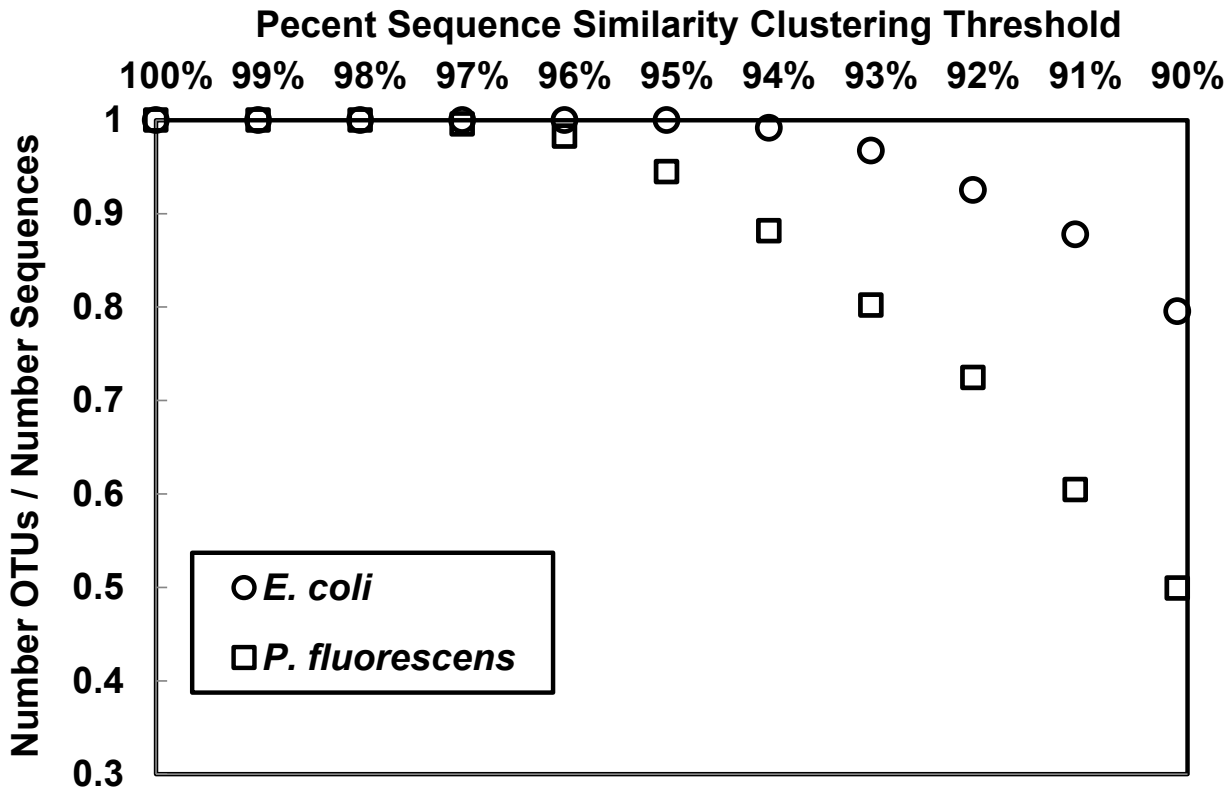
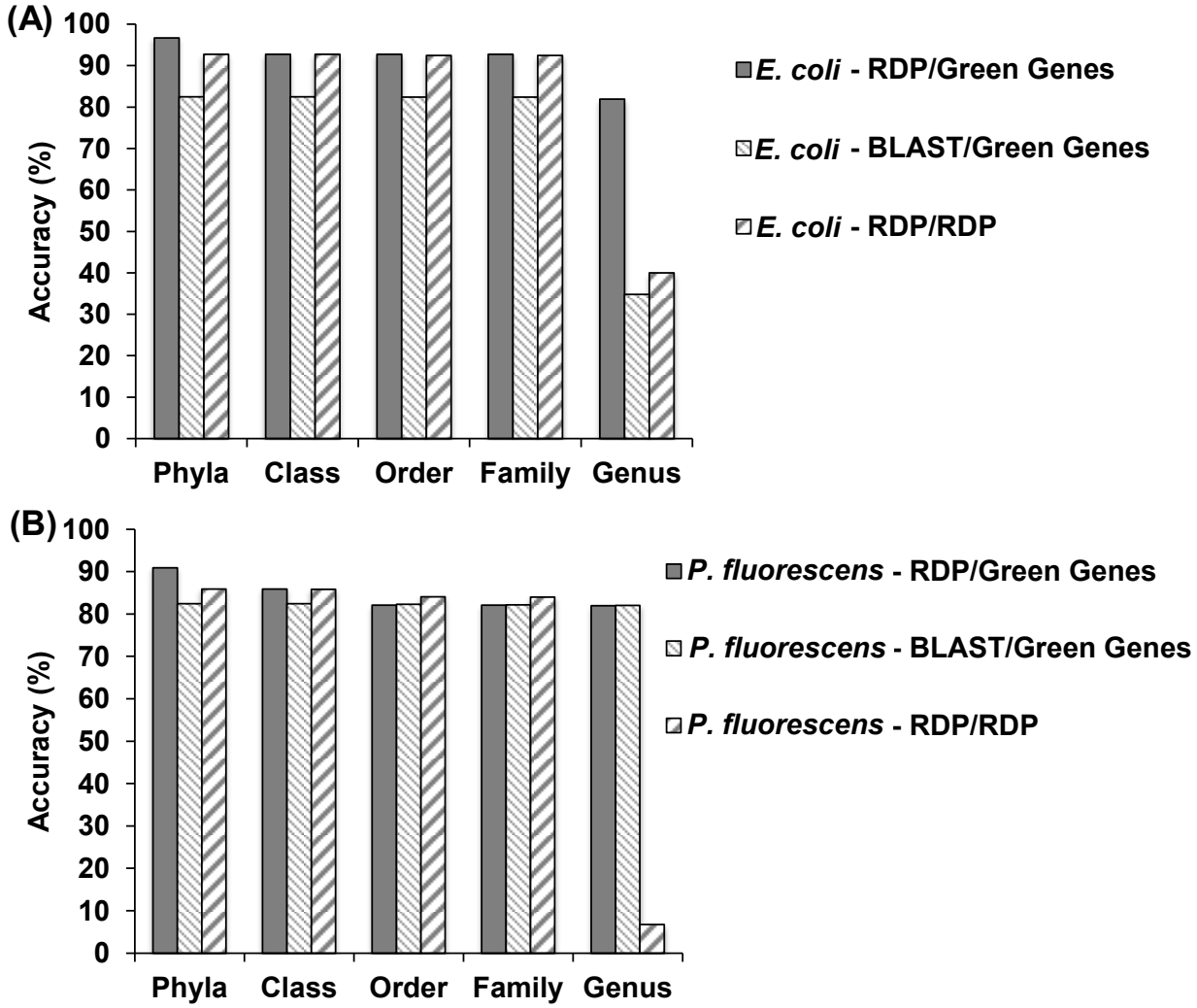


Figure 1. Number of observed *de novo* operational taxonomic units (OTUs) per number of sequences at different similarity thresholds for *E. coli* and *P. fluorescens* pure culture DNA sample sequenced by Nanopore.

405 in



406

407 Figure 2. Accuracy of taxonomy assignment at different phylogenetic levels for (A) *E. coli*; and
408 (B) *P. fluorescens*

409

410

411

412

413

414

Figure 3. Jaccard (unweighted) and Bray-Curtis (weighted) dissimilarity values for a hydraulic fracturing produced water sample analyzed by Nanopore and Illumina sequencing; distance value of one indicating no community structure overlap, and a value of zero indicating identical community structure

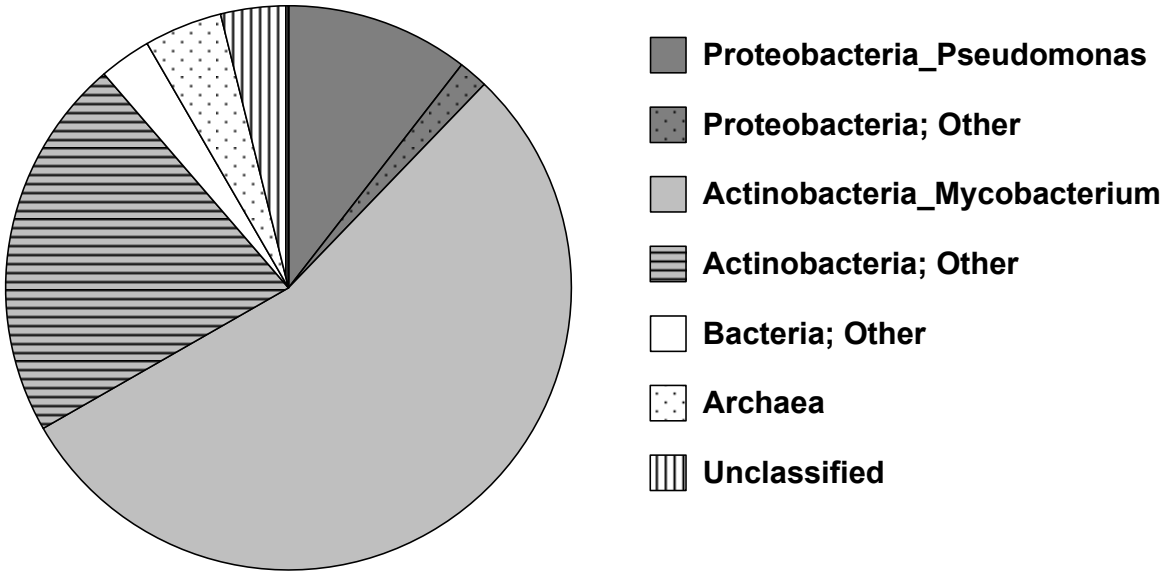


Figure 4. Taxonomy assignment of *Mycobacterium smegmatis* 16S rRNA pure culture sequencing following *Pseudomonas fluorescens* sequencing.