

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

3
4 Xiao Ma¹, Elyse Stachler¹, Kyle Bibby^{1, 2*}

5
6 ¹Department of Civil and Environmental Engineering, University of Pittsburgh, Pittsburgh, PA
7 15261,

8 ²Department of Computational and Systems Biology, University of Pittsburgh Medical School,
9 Pittsburgh, PA 15261, USA

10 *Corresponding author

11

12

13

14

15

16 *Corresponding Author: Kyle Bibby, 709 Benedum Hall, Pittsburgh, PA 15261

17 BibbyKJ@pitt.edu, 412-624-9207

18

19 **Keywords:** Nanopore, microbiome, 16S rRNA, sequencing

20

21

22

23 **Abstract**

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

40

41 **Introduction**

42 Interest in studying the microbiome, microbiota associated with various environments, has
43 exploded in recent years largely due to the rapid expansion in ‘next-generation sequencing’
44 capabilities and subsequent reduction in cost. It has been recognized that the human microbiome
45 plays an important role in many different clinical outcomes, including obesity (1), immune state

46 (2), and infection (3). The human microbiome is comprised of a diverse array of commensal
47 microorganisms in and on the human body, and emerging research has suggested a clinical role
48 for the microbiome in either therapeutic development (e.g. probiotics) (4) or diagnostics (5).
49 Concurrently, significant interest has emerged in the microbiomes of various other environments,
50 such as buildings (6-8) and water systems (9).

51 Currently, the most common microbiome analysis approach is high-throughput amplicon
52 sequencing of the 16S rRNA gene region. The most widely used sequencing technologies are
53 Illumina sequencing platforms (e.g. the MiSeq and HiSeq). These platforms are accurate and
54 generate a large amount of data, but are limited by capital costs, the necessity to pool samples to
55 reduce per-sample costs, sequence read length, and a turnaround time of days to weeks that may
56 be inadequate for many applications. Recently, Oxford Nanopore has released a small and
57 inexpensive sequencing platform called the MinIONTM, which has been previously reviewed (10,
58 11). Capital costs are reduced to per-run costs that are comparable with the current Illumina
59 platforms, and data analysis is possible in near-real time, enabling investigators to generate
60 sequence data as-needed. For example, samples were correctly assigned to the *Salmonella*
61 species in 20 minutes and serotype in 40 minutes (12). Amplicon (non-metagenomic) sequencing
62 has also been successfully employed to identify both bacterial and viral origin (13, 14).
63 Additionally, this technology routinely produces sequences >10kb in length, enabling
64 sequencing of the full 16S rRNA gene region and more reliable taxonomic placement. Despite
65 the benefits, the primary drawback to this technology is relatively high error rates (currently
66 reported to be ~8%), hindering metagenomic analysis of highly diverse microbiome samples and
67 requiring additional development and validation. Previous investigations have demonstrated
68 successful 16S rRNA microbial community classification via sequencing of a single mouse gut

69 microbiome (15) and a single mock microbial community (16), including species-level
70 assignment (16); however, a formal assessment of 16S rRNA sequencing on the Nanopore
71 platform, including analysis of pure-culture samples for annotation validation, is currently
72 lacking. Development of Nanopore technology for microbiome analysis would enable rapid (<12
73 hours from sample to results) and low cost microbiome characterization that would be applicable
74 to both the clinic and other applications.

75

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

84

85

86 **Results and Discussion**

87 *Sequencing*

88 Sequencing was performed on eight 16S rRNA libraries; however, due to apparent between-run
89 carryover (discussed below), only the initial run on each of three flow cells was used for
90 subsequent analyses of annotation approach and accuracy. These sequencing libraries were

91 derived from two pure-culture samples, *E. coli* and *P. fluorescens*, and a low-diversity sample
92 from hydraulic fracturing produced water. Detailed sequencing results are shown in Table 1.
93 Raw sequence data can be found on Figshare using DOI:
94 <https://dx.doi.org/10.6084/m9.figshare.4515752.v3>.

95

96 ***Pure Culture Analyses***

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

107

108 We next evaluated the ability to accurately annotate the taxonomy of pure-culture Nanopore 16S
109 rRNA sequences using three different annotation approaches: the naïve Bayesian Ribosomal
110 Database Project (RDP) classifier with the Green Genes database; the RDP classifier against the
111 RDP database; and BLAST against the Green Genes database. Results from this evaluation are
112 shown in Figure 2. The RDP classifier against the Green Genes database was found to be the best
113 performing annotation strategy. Using this approach, the annotation accuracy for *E. coli* was

114 96.7% and 81.9% at the phyla and genus level, respectively, and was 90.9% and 82.0% for *P.*
115 *fluorescens* at the phyla and genus level, respectively.

116

117 ***Comparison with Illumina Sequencing***

118 We next evaluated Nanopore sequencing to characterize an environmental sample from
119 hydraulic fracturing produced water (17). This sample was selected as it exhibited low alpha-
120 diversity in previous analyses. The same DNA extract was used for both analyses.

121

122 By assigning taxonomy to each individual sequence, five phyla were identified by the Nanopore
123 platform, and eleven phyla were identified by the Illumina platform. Among them, four phyla
124 that together accounted for greater than 99% of sequence relative abundance were identified by
125 both platforms (Table 2). Nine shared genera were detected by both platforms, accounting for
126 relative abundances of 98.3% on the Nanopore platform and 81.6% on the Illumina platform
127 (Table 2). Both the Nanopore and Illumina platforms revealed similar microbial community
128 structure of the produced water sample. The Firmicutes Phylum dominated the microbial
129 community with relative abundance higher than 90% with both platforms (Table S1). Phyla
130 unique to the Nanopore and Illumina platforms accounted for less than 0.5% of relative
131 abundance (Table S1). At the genus level, the produced water microbial community was
132 dominated by the genus *Halanaerobium* (Table S2), with 95.6% of Nanopore sequences and
133 76.8% of Illumina sequences being assigned to *Halanaerobium*. In addition, 14.6% of Illumina
134 16S rRNA sequence reads were assigned to Clostridiales, which is within the same class with
135 *Halanaerobium* (Clostridia) (Table S2).

136

137 We also calculated the Jaccard distance (solely based on presence/absence of each taxa) and
138 Bray-Curtis distance (based on both presence/absence and relative abundance of each taxa)
139 between the produced water microbial community revealed by Nanopore and Illumina platforms
140 at different taxonomic levels. Jaccard and Bray-Curtis distances measure the degree of microbial
141 community structure difference between two samples, with a value of one indicating no
142 community structure overlap and a value of zero indicating identical microbial communities. We
143 adopted these measures to evaluate the level of difference between the technical replicates of the
144 same produced water sample between the Nanopore and Illumina sequencing platforms. The
145 Jaccard distance increased from 0.62 at the phylum level to 0.89 at genus level; the Bray-Curtis
146 distance increased from 0.04 at phylum level to 0.22 at genus level (Figure 3). Jaccard distances
147 were higher than Bray-Curtis distances at all phylogenetic levels, because more taxa were
148 assigned using short sequence reads by Illumina sequencing (Table 2) and Jaccard distance only
149 accounts for the presence and absence of each assigned taxa whereas Bray-Curtis distances take
150 relative abundance into account.

151
152 Pearson correlation of the relative abundance of each taxon between Nanopore and Illumina data
153 was conducted to further evaluate the reproducibility of sequencing results between the two
154 platforms. Significant correlation was found at all phylogenetic levels from phylum to genus (R
155 values > 0.98 , p values < 0.001), indicating reproducible taxonomic assignment results can be
156 obtained between Nanopore and Illumina platforms.

157
158 It should be noted in these comparisons that different primer sets were used for the two analysis
159 approaches, which has previously been shown to bias microbiome community structure as

160 analyzed by 16S rRNA sequencing (18). Despite this additional source of bias, the above
161 analyses imply that the weighted community structure is comparable between the two platforms,
162 encouraging future development.

163

164 ***Between-Run Sample Carryover***

165 We noted an apparent carryover of sequences between pure culture runs of *P. fluorescens* and *M.*
166 *smegmatis*. The potential for sequence carryover has been anecdotally reported in the literature
167 (19). We subsequently excluded all runs except the first run on each flow cell from earlier
168 analyses, and undertook a formal evaluation of sequence carryover.

169

170 Results from analysis of the *Mycobacterium smegmatis* run are shown in Figure 4. In this run,
171 76.5% of sequences were assigned to the correct Actinobacteria phyla, 11.0% of sequences were
172 incorrectly assigned to another domain or unable to be assigned, and 12.4% of sequences were
173 incorrectly assigned to the Proteobacteria phyla, presumptively resulting from sequence
174 carryover. 54.5% of sequences were assigned to the correct *Mycobacterium* genus whereas
175 10.8% of sequences were assigned to the *Pseudomonas* genus.

176

177 Following the *P. fluorescens* and *M. smegmatis* runs, we subsequently ran an additional
178 environmental sample derived from river water and re-ran a *M. smegmatis* sample using 1D
179 technology. In the second *M. smegmatis* run, we observed only 0.5% of sequences to be assigned
180 to the correct Mycobacteriaceae family, compared with 54.5% in the first *M. smegmatis* run.
181 These results imply that continued carry-over serves to significantly decrease output quality;

182 however, additional validation with controlled microbial community composition is necessary to
183 confirm this observation.

184

185 ***Areas of Future Development***

186 This investigation has identified multiple necessary areas of future development to enable 16S
187 rRNA microbiome characterization on the Nanopore platform. First, strategies to exclude
188 between-run carryover, either via improved washing between runs or a barcode approach, would
189 enable multiple runs on the same flow cell, significantly reducing per-run costs. In this
190 investigation we performed six runs on the same flow cell while observing minimal output loss.
191 Second, improved bioinformatics strategies are necessary to exclude poor quality sequences. In
192 the *E. coli* and *P. fluorescens* runs, 3.3 and 9.0% of sequences, respectively, were not assigned to
193 any phyla, suggesting poor sequence quality and confounding both alpha- and beta-diversity
194 analyses. Finally, it would be beneficial to develop a 16S rRNA annotation pipeline based upon
195 optimized analysis strategies that provides output in near real-time, facilitating field and clinic
196 applications and alleviating current bioinformatics challenges from interested investigators.
197 Ultimately, the development of a rapid and low-cost microbiome approach will facilitate the
198 application of clinical and environmental microbiome technologies.

199 **Materials and Methods**

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

201 Three pure culture bacterial DNA samples and an environmental DNA sample were analyzed in
202 the current study. The three pure culture bacterial DNA samples were extracted from pure
203 cultures of: (1) *Escherichia coli* (ATCC 15597), (2) *Pseudomonas fluorescens* (ATCC 13525),
204 and (3) *Mycobacterium smegmatis* str. mc2 155, respectively. The environmental DNA samples,

205 which had previously undergone 16S rRNA gene sequencing using the Illumina MiSeq platform,
206 were a hydraulic fracturing produced water sample (17) and a river water sample (unpublished).

207
208 All Nanopore sequencing runs were conducted using the MinION Mk IB platform following
209 recommended sequencing protocols (Oxford Nanopore Technologies). The *E. coli* 16S rRNA
210 amplicon 2D library and the hydraulic fracturing produced water 16S rRNA amplicon 2D library
211 were sequenced individually using a Nanopore MIN-105 flow cell and a Nanopore MIN-106
212 flow cell, respectively. The remaining libraries were sequenced on a MIN-106 flow cell
213 following a sequential order: (1) *P. fluorescens* 2D library; (2) *M. smegmatis* 2D library; (3) river
214 water sample 2D library; (4) *M. smegmatis* 1D library. For the sequential sequencing runs, flow
215 cell washing was conducted immediately following the completion of the previous sequencing
216 run using a Nanopore washing kit WSH002 (Oxford Nanopore). The Oxford Nanopore
217 recommended washing protocol was used between runs, namely 150 μ L of WSH002 solution A
218 was loaded to the flow cell through priming port and incubated at room temperature for 10
219 minutes, then 150 μ L of WSH002 solution B was loaded through the priming port before the
220 next sequencing run and incubated for another 10 minutes at room temperature.

221

222 **Nanopore sequencing library preparation**

223 Previously described universal primers targeting the 16S rRNA gene region (S-D-bact-0008-c-
224 S20 and S-D-bact-1391-a-A-17) (20) were used for PCR. Each PCR was conducted in a total
225 volume of 50 μ L, containing 5 μ L 10x buffer, 5 μ L dNTP mix, 2.5 μ L of each forward and
226 reverse primer, 0.25 μ L DreamTaq, 1 μ L template DNA, and 33.75 μ L nuclease free molecular
227 grade water. The temperature condition for the PCR was 3 minutes at 95°C; 30 cycles composed

228 of 20 seconds at 95 °C, 30 seconds at 47 °C for annealing, 1 minute at 72 °C; and a final
229 elongation at 72 °C for 15 minutes. All PCR products were purified using Ampure XP beads and
230 normalized to 45 µL containing 1 µg of purified PCR products. Negative controls were used for
231 all PCR reactions and DNA extractions, and all controls were negative.

232
233 2D libraries were prepared using a Nanopore NSK007 sequencing kit and recommended protocol
234 (Oxford Nanopore Technologies). The end repair step of the purified PCR products was
235 conducted by adding 7 µL Ultra II End-Prep buffer, 3 µL Ultra II End-Prep enzyme mix (New
236 England Biolabs), and 5 µL control DNA provided with Nanopore NSK007 sequencing kit. The
237 end repair reaction mix was incubated at 20°C for 5 minutes and 65°C for 5 minutes. The end-
238 repaired PCR products were further purified using AMPure XP beads and ligated to the
239 sequencing adapters by adding 8 µL molecular grade water, 10 µL Nanopore NSK007 adapter
240 mix, 2 µL Nanopore NSK007 HPA solution, and 50 µL Blunt/TA Master Mix (New England
241 Biolabs), and then incubated at room temperature for 10 minutes. 1 µL HPT solution from the
242 NSK007 kit was added and incubated for an additional 10 minutes at room temperature. The
243 ligated and tethered 2D libraries were purified by using MyOne C1 beads (Thermo Scientific)
244 and eluted in 25 µL elution buffer (Oxford Nanopore Technologies). A description of the *M.*
245 *smegmatis* 1D library preparation is included in the Supplementary Information.

246
247 All sequencing flow cells were primed using 500 µL Running Buffer Fuel Mix diluted in 500 µL
248 molecular grade water following the recommended priming protocol (Oxford Nanopore
249 Technologies). After priming, 6 µL of each 2D sequencing library was mixed with 37.5 µL

250 Running Buffer Fuel Mix (Oxford Nanopore Technologies) and 31.5 μ L molecular grade water,
251 then loaded to Nanopore flow cell for sequencing.

252

253 **Sequence data processing**

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

255 The raw FAST5 files were base-called using Metrichor v2.42.2 with 2D Basecalling for FLO-
256 MIN106 250bps workflow and 1D Basecalling for FLO-MIN106 450bps workflow. Passed 2D
257 reads of 2D sequencing libraries and passed template reads of 1D sequencing libraries were
258 converted to FASTA files for downstream analysis using Poretools (21).

259

260 **Operational Taxonomic Unit Evaluation**

261 Operational taxonomic unit (OTU) picking was conducted by using both closed-reference and *de*
262 *novo* picking strategies implemented in QIIME 1.9.2 (22). Closed-reference OTU picking was
263 conducted by using Greengenes 13.8 as the reference database (23).

264

265 **Taxonomy Assignment**

266 For pure culture *E. coli* and *P. fluorescens* sequencing data, taxonomy was assigned to each
267 sequence read within QIIME 1.9.0 (24) using the RDP classifier (25) against Greengenes 13.8
268 (23) and RDP 16S rRNA training set (25) as the reference database, respectively; as well as
269 using BLASTn (26) against Greengenes 13.8 (23) as a reference database. For subsequent
270 analyses, taxonomy was assigned to each individual sequence read using the RDP classifier (25)
271 against Greengenes 13.8 (23) as this approach was found to achieve the highest taxonomy
272 assignment accuracy for the pure culture *E. coli* and *P. fluorescens* sequence data.

273
274 Illumina 16S rRNA amplicon sequencing data of the produced water and river water samples
275 were re-processed using the same approach as Nanopore sequence data. The Illumina data were
276 clustered into OTUs using 100% similarity threshold, i.e. each identical Illumina sequencing
277 read was assigned taxonomy using RDP classifier (25) against Greengenes 13.8 (23).

278
279 Jaccard and Bray-Curtis distances, which are dissimilarity distances measuring level of
280 dissimilarity between two microbial communities, were calculated using QIIME 1.9.2 (22) to
281 measure the degree of similarity of the produced water microbial community similarity between
282 Nanopore and Illumina platforms. Significance of correlation between the two technical
283 replicates of the produced water sample across sequencing platforms was conducted using
284 Pearson's correlation implemented in Minitab 16.

285
286 **Acknowledgements**
287 Support for this project was provided by the University of Pittsburgh Central Research
288 Development fund. KB was a member of the Oxford Nanopore Early Access Program, which
289 initially provided access to the sequencing platform at reduced cost.

290
291 **References**
292 1. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. 2006. An
293 obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*
294 444:1027-131.
295 2. Kumar P, Monin L, Castillo P, Elsegeiny W, Horne W, Eddens T, Vikram A, Good M,
296 Schoenborn Alexi A, Bibby K, Montelaro Ronald C, Metzger Dennis W, Gulati Ajay S,
297 Kolls Jay K. Intestinal Interleukin-17 Receptor Signaling Mediates Reciprocal Control of
298 the Gut Microbiota and Autoimmune Inflammation. *Immunity* 44:659-671.

- 299 3. Khoruts A, Dicksved J, Jansson JK, Sadowsky MJ. 2010. Changes in the composition of
300 the human fecal microbiome after bacteriotherapy for recurrent *Clostridium difficile*-
301 associated diarrhea. *Journal of clinical gastroenterology* 44:354-360.
- 302 4. Good M, Sodhi CP, Ozolek JA, Buck RH, Goehring KC, Thomas DL, Vikram A, Bibby
303 K, Morowitz MJ, Firek B, Lu P, Hackam DJ. 2014. *Lactobacillus rhamnosus* HN001
304 decreases the severity of necrotizing enterocolitis in neonatal mice and preterm piglets:
305 evidence in mice for a role of TLR9. *American Journal of Physiology - Gastrointestinal
306 and Liver Physiology* 306:G1021-G1032.
- 307 5. Pflughoeft KJ, Versalovic J. 2012. Human microbiome in health and disease. *Annual
308 Review of Pathology: Mechanisms of Disease* 7:99-122.
- 309 6. Prussin AJ, II, Vikram A, Bibby KJ, Marr LC. 2016. Seasonal Dynamics of the Airborne
310 Bacterial Community and Selected Viruses in a Children's Daycare Center. *PLoS ONE*
311 11:e0151004.
- 312 7. Adams RI, Bhangar S, Dannemiller KC, Eisen JA, Fierer N, Gilbert JA, Green JL, Marr
313 LC, Miller SL, Siegel JA. 2016. Ten questions concerning the microbiomes of buildings.
314 *Building and Environment* 109:224-234.
- 315 8. Lax S, Gilbert JA. 2015. Hospital-associated microbiota and implications for nosocomial
316 infections. *Trends in molecular medicine* 21:427-432.
- 317 9. Baron JL, Vikram A, Duda S, Stout JE, Bibby K. 2014. Shift in the Microbial Ecology of
318 a Hospital Hot Water System following the Introduction of an On-Site Monochloramine
319 Disinfection System. *PLoS ONE* 9:e102679.
- 320 10. Mikheyev AS, Tin MM. 2014. A first look at the Oxford Nanopore MinION sequencer.
321 *Molecular ecology resources* 14:1097-1102.
- 322 11. Laver T, Harrison J, O'Neill P, Moore K, Farbos A, Paszkiewicz K, Studholme DJ. 2015.
323 Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomolecular
324 detection and quantification* 3:1-8.
- 325 12. Quick J, Ashton P, Calus S, Chatt C, Gossain S, Hawker J, Nair S, Neal K, Nye K, Peters
326 T. 2015. Rapid draft sequencing and real-time nanopore sequencing in a hospital
327 outbreak of *Salmonella*. *Genome Biol* 16:10.1186.
- 328 13. Kilianski A, Haas JL, Corriveau EJ, Liem AT, Willis KL, Kadavy DR, Rosenzweig CN,
329 Minot SS. 2015. Bacterial and viral identification and differentiation by amplicon
330 sequencing on the MinION nanopore sequencer. *Gigascience* 4:10.1186.
- 331 14. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, Bore JA, Koundouno
332 R, Dudas G, Mikhail A, Ouédraogo N, Afrough B, Bah A, Baum JHJ, Becker-Ziaja B,
333 Boettcher JP, Cabeza-Cabrerizo M, Camino-Sánchez Á, Carter LL, Doerrbecker J,
334 Enkirch T, Dorival IG, Hetzelt N, Hinzmann J, Holm T, Kafetzopoulou LE, Koropogui
335 M, Kosgey A, Kuisma E, Logue CH, Mazzarelli A, Meisel S, Mertens M, Michel J,
336 Ngabo D, Nitzsche K, Pallasch E, Patrono LV, Portmann J, Repits JG, Rickett NY,
337 Sachse A, Singethan K, Vitoriano I, Yemanaberhan RL, Zekeng EG, Racine T, Bello A,
338 Sall AA, Faye O, et al. 2016. Real-time, portable genome sequencing for Ebola
339 surveillance. *Nature* 530:228-232.
- 340 15. Shin J, Lee S, Go M-J, Lee SY, Kim SC, Lee C-H, Cho B-K. 2016. Analysis of the
341 mouse gut microbiome using full-length 16S rRNA amplicon sequencing. *Scientific
342 Reports* 6:29681.

- 343 16. Benítez-Páez A, Portune KJ, Sanz Y. 2016. Species-level resolution of 16S rRNA gene
344 amplicons sequenced through the MinION™ portable nanopore sequencer. *GigaScience*
345 5:4.
- 346 17. Lipus D, Vikram A, Ross D, Bain D, Gulliver D, Hammack R, Bibby K. In Revision.
347 Predominance and Metabolic Potential of Halanaerobium in Produced Water from
348 Hydraulically Fractured Marcellus Shale Wells. *Applied and Environmental*
349 *Microbiology*.
- 350 18. Pinto AJ, Raskin L. 2012. PCR Biases Distort Bacterial and Archaeal Community
351 Structure in Pyrosequencing Datasets. *PLOS ONE* 7:e43093.
- 352 19. Greninger AL, Naccache SN, Federman S, Yu G, Mbala P, Bres V, Stryke D, Bouquet J,
353 Somasekar S, Linnen JM, Dodd R, Mulembakani P, Schneider BS, Muyembe-Tamfum J-
354 J, Stramer SL, Chiu CY. 2015. Rapid metagenomic identification of viral pathogens in
355 clinical samples by real-time nanopore sequencing analysis. *Genome Medicine* 7:99.
- 356 20. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. 2013.
357 Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-
358 generation sequencing-based diversity studies. *Nucleic Acids Research* 41:e1-e1.
- 359 21. Loman NJ, Quinlan AR. 2014. Poretools: a toolkit for analyzing nanopore sequence data.
360 *Bioinformatics* 30:3399-3401.
- 361 22. Caporaso J, Kuczynski J, Stombaugh J, Bittinger K, Bushman F, Costello E, Fierer N,
362 Pena A, Goodrich J, Gordon J, Huttley G, Kelley S, Knights D, Koenig J, Ley R,
363 Lozupone C, McDonald D, Muegge B, Pirrung M, Reeder J, Sevinsky J, Turnbaugh P,
364 Walters W, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows
365 analysis of high-throughput community sequencing data. *Nature Methods* 7:335 - 336.
- 366 23. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D,
367 Hu P, Andersen GL. 2006. Greengenes, a Chimera-Checked 16S rRNA Gene Database
368 and Workbench Compatible with ARB. *Applied and Environmental Microbiology*
369 72:5069-5072.
- 370 24. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer
371 N, Pena AG, Goodrich JK, Gordon JI. 2010. QIIME allows analysis of high-throughput
372 community sequencing data. *Nature methods* 7:335-336.
- 373 25. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naïve Bayesian Classifier for Rapid
374 Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied and*
375 *Environmental Microbiology* 73:5261-5267.
- 376 26. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment
377 search tool. *Journal of Molecular Biology* 215:403-410.

378

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

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

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

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

390

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

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

391

392

393

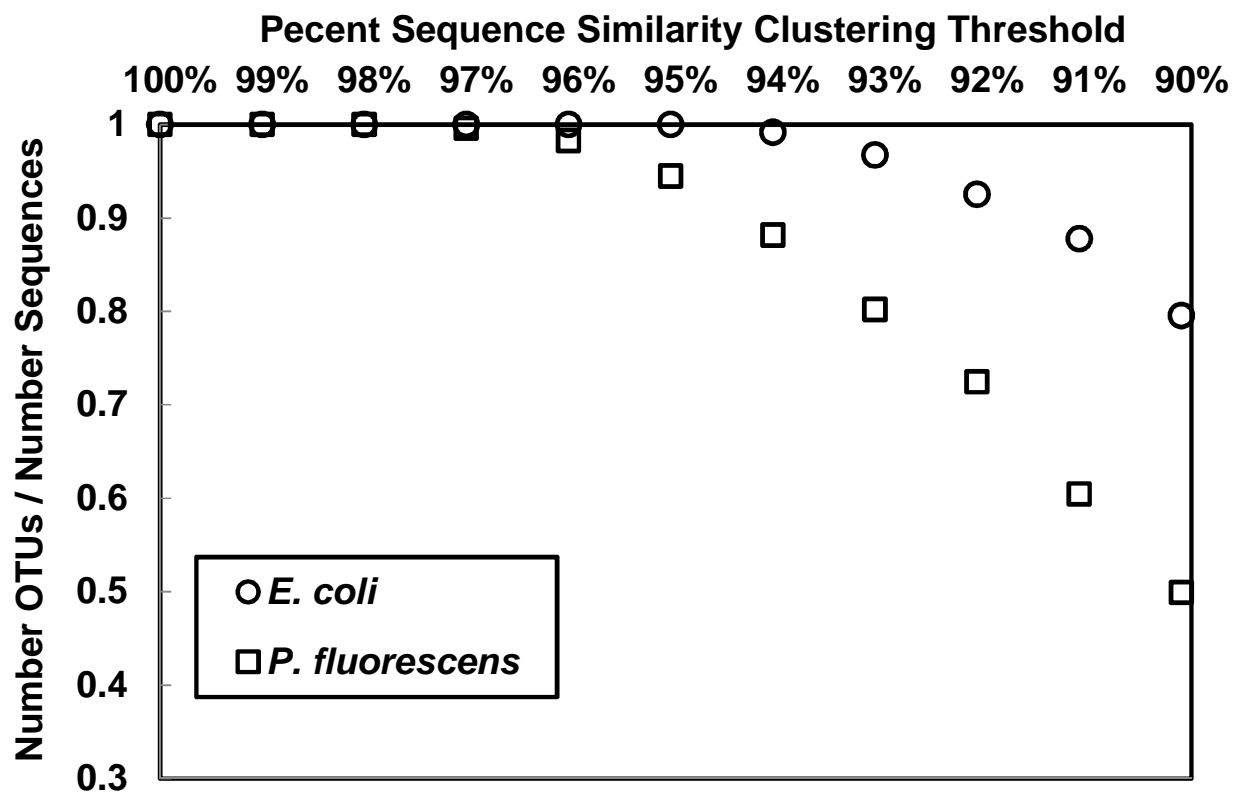
Table 2. Comparison of number of taxa at different phylogenetic levels for the produced water sample

	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%

394

395 Table S1. Relative abundance distribution at phylum level for the produced water sample.

396 Table S2. Relative abundance distribution of Firmicutes phylum for the produced water sample.

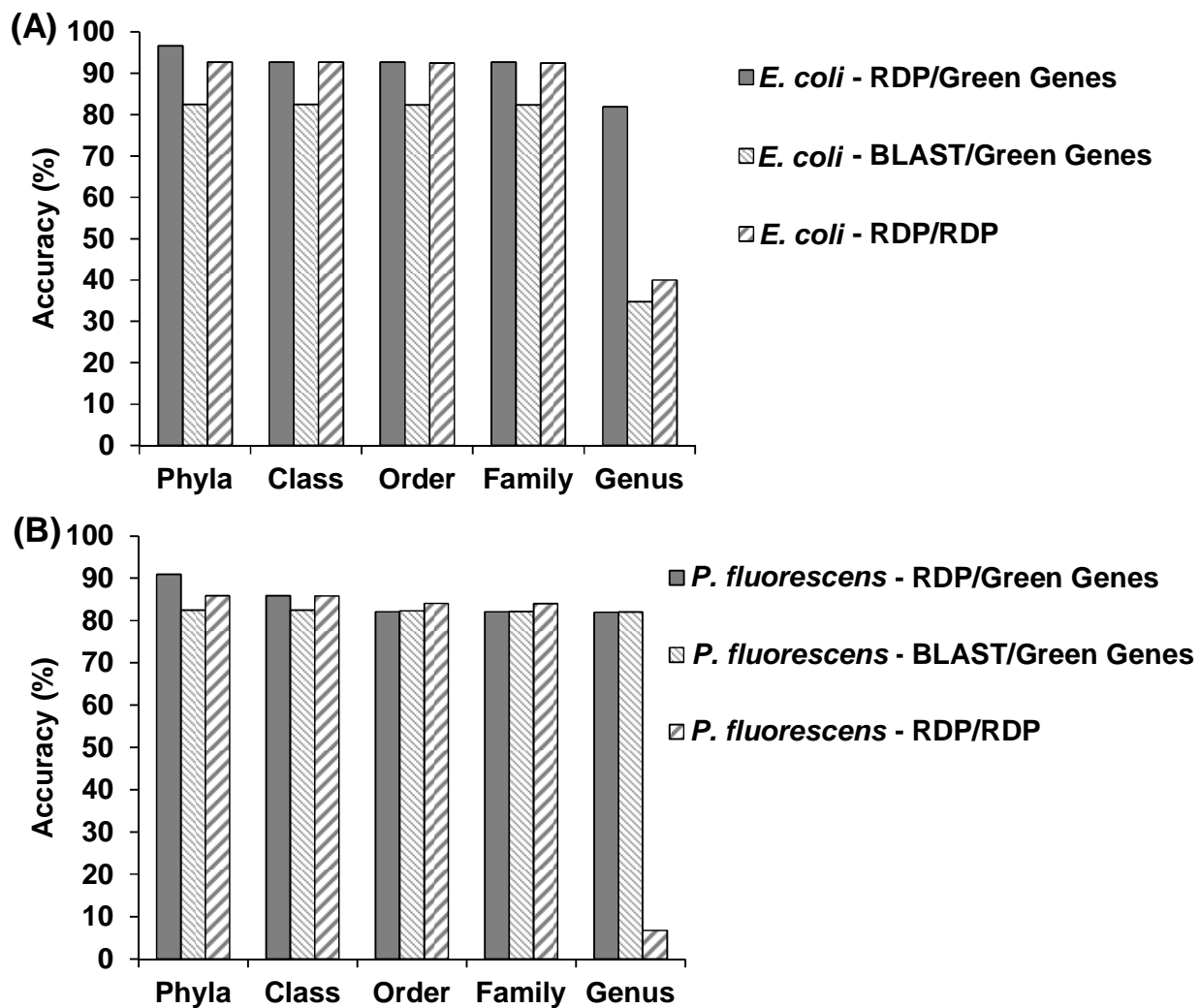


397

398 Figure 1.

399

400 in



401

402 Figure 2.

403

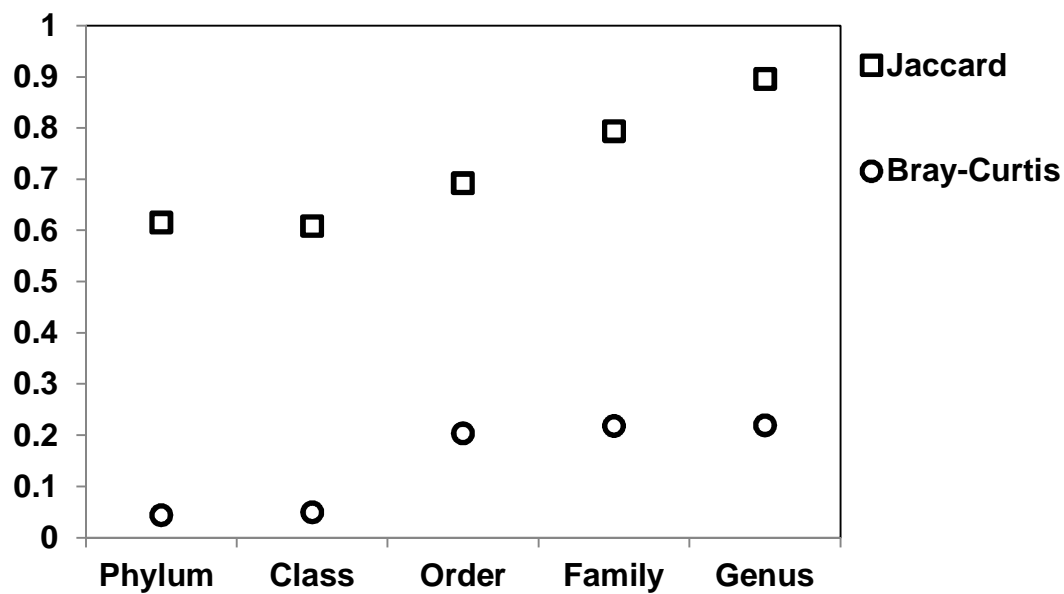
404

405

406

407

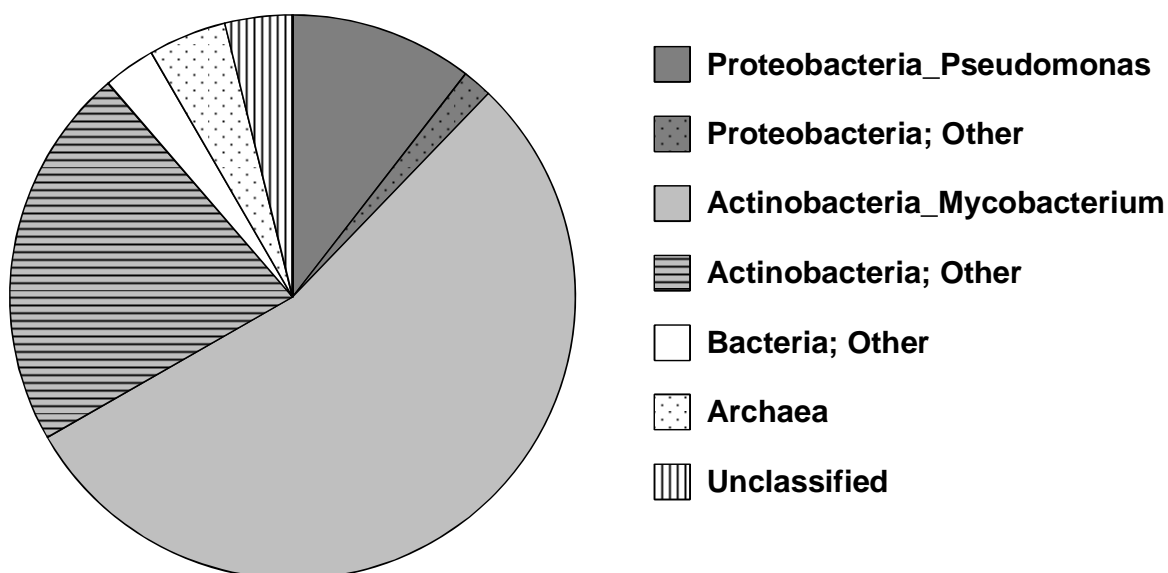
408



409

410 Figure 3.

411



412

413 Figure 4.

414