

Metagenomic survey of agricultural water using long read sequencing: considerations for a successful analysis

Meghan Maguire

Center for Food Safety and Applied Nutrition, FDA

Julie A. Kase

Center for Food Safety and Applied Nutrition, FDA

Eric W. Brown

Center for Food Safety and Applied Nutrition, FDA

Marc Allard

Center for Food Safety and Applied Nutrition, FDA

Steve Musser

Center for Food Safety and Applied Nutrition, FDA

Narjol González-Escalona (✉ narjol.gonzalez-escalona@fda.hhs.gov)

Center for Food Safety and Applied Nutrition, FDA

Research Article

Keywords: Foodborne pathogens, nanopore sequencing, agricultural water, metagenomics, Shiga toxin-producing *Escherichia coli*, STEC

Posted Date: September 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-825305/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

Leafy greens are responsible for nearly half of the produce-related Shiga toxin-producing *Escherichia coli* (STEC) outbreaks in the United States and recent investigations have implicated agricultural water as a potential source. Current FDA detection protocols require extensive analysis time. Development of methods for rapid detection of STECs in the field are imperative to maintain food safety. We aimed to use Oxford Nanopore rapid sequencing kits for an in-field determination of agricultural water microbiome and possible detection and characterization of STECs strain(s) in these samples.

Results

We tested the performance of the nanopore rapid sequencing kit (RAD004) for fast microbiome determination using the well characterized ZymoBIOMICS mock microbial community and the number of reads for each identified species was present in the expected proportion. Rapid sequencing kit (LRK001 and RAD004) library preparation of DNA extracted from agricultural water resulted in poor nanopore sequencing reactions, with low output (0.3–1.7 M reads), a high proportion of failed reads (50–60%), and highly sheared DNA before and after a magnetic bead clean up. To improve performance, we prepared a DNA library with the ligation kit (LSK109), which includes multiple cleaning steps, reducing inherent inhibitors and producing a better outcome (2.2 M reads, 15% failed reads). Metagenomic composition of four sample sites determined using the ligation kit, showed a highly diverse bacterial community for each site, with 11 total genera identified, including *Synechococcus* and *Cyanobium*. No definitive presence of STEC could be confirmed in any of the sites. Approximately 100 reads from each site (0.02% of total reads) were identified as *Escherichia coli*, but the specific strain or their virulence genes could not be detected. Sites 9, 10, and 12 were found to be positive for STEC presence by microbiological techniques after enrichment.

Conclusions

The rapid sequencing kits can be appropriate for genus or species level microbial identification, but we recommend the use of the ligation kit for increased sequencing depth and removal of contaminants in agricultural water. However, we were not able to identify any STEC strains in these nanopore microbiome samples, due to low initial concentrations. The results from this pilot study provide preliminary evidence that MinION sequencing of agricultural water using the ligation kit has the potential to be used for rapid microbiome determination in the field with optimal results for water quality surveillance.

Background

Shiga toxin-producing *Escherichia coli* (STEC) is a foodborne pathogen responsible for approximately 265,000 illnesses per year in the United States (1). STEC infection can cause severe disease, including bloody diarrhea and hemolytic uremic syndrome (HUS) (2–5). STECs are defined by the presence of Shiga toxin genes (*stx*) and are identified by serotype based on their O and H antigens. While the most common STEC associated with outbreaks and illness is *E. coli* O157:H7 (1, 6, 7), there are over 400 STEC serotypes with varying degrees of pathogenicity, which can be determined *in silico* by the presence of virulence genes (2, 8). Attachment and colonization genes can be found in the locus of enterocyte effacement (LEE), including intimin (*eae*) and type 3 secretion system (TTSS) effector proteins (*esp*, *esc*, *tir*), Additional non-LEE effectors (*nleA*, *nleB*, *nleC*) and other putative virulence genes (*ehxA*, *etpD*, *subA*,

toxB, *saa*) can also impact virulence (2, 9–11). It is, therefore, imperative that the serotype and virulence factors are identified to assess potential pathogenicity.

STEC infections have been linked to multiple commodities (e.g. beef, milk, yogurt), including a growing incidence in produce (12–14), especially leafy greens, with agricultural water implicated as a potential source (15–21). Agricultural water can be contaminated by adjacent land use, wild animal activity, or incomplete water sanitization (16). There are currently no approved antimicrobial treatments for agricultural water to prevent against foodborne pathogens; however the FDA has collaborated with the EPA to establish a new protocol for development and registration of treatment of agricultural water (22). This protocol was developed under the 2020 Leafy Greens STEC Action Plan in which the FDA is focused on improving the safety of leafy greens through a set of guidelines, including extensive meta-analysis of past outbreak data, longitudinal studies, promotion of tech-enabled traceability, monitoring of nearby agricultural land use, compost sampling, and improved whole genome sequencing (WGS) tracing (23).

WGS has increased the precision and responsiveness of food safety by the ability to produce closed genomes, determining serotype, virulence, antimicrobial resistance, and phylogenetic relationships, particularly during an outbreak (9, 24–27). While single colony isolation and WGS are the current standard procedure for FDA STEC detection and classification (FDA Bacteriological Analytical Manual Chap. 4A (28), these reporting methods require approximately two weeks of analysis time. U.S. Federal regulatory action will continue to require a single isolate, but metagenomic, culture-independent methods are being tested for expedited detection and characterization (27, 29, 30). Retrospective analysis of clinical fecal samples from *E. coli* O104:H4 and *Salmonella enterica* outbreaks have shown promising results in detecting and characterizing the outbreak strain (29, 30) Targeted microbial detection by metagenomic analysis using 16S rRNA profiling or shotgun metagenomic sequencing is increasingly being used (31–36). Mock microbial communities have been used in long read metagenomic studies and have shown the capability of obtaining closed metagenome assembled genomes (MAGs) for high concentration microbial species (37–39). Library preparation with the Oxford Nanopore ligation kit (SQK-LSK109) produces approximately 10–20 Gb of sequencing data. In a study by Nicholls, et al (2019) sequencing of the ZymoBIOMICS Mock Microbial Community resulted in more than 150X coverage for each of the 8 bacterial species. Long read sequencing is particularly useful in assembling complex, highly repetitive regions that can extend for hundreds of kilobases (40).

Oxford Nanopore has developed portable rapid sequencing kits that have less stringent storage conditions and require minimal time and equipment (LRK001 and RAD004). Combined with a culture-independent, metagenomics approach, these kits may be useful tools for the characterization of agricultural water microbiome. We aimed to test the efficacy of these rapid kits by using a mock microbial community and design a pilot study for a fast, field-based method for microbial analysis, including the detection and characterization of STECs in agricultural water.

Results

Characterization of a bacterial community standard using the RAD004 rapid sequencing kit. The first step in our investigation for the use of the rapid sequencing kit (RAD004) for fast taxonomic classification of the microbial composition of a sample was to test the performance of our proposed workflow (nanopore sequencing using the RAD004 library preparation followed by WIMP classification tool) with a known microbial standard (ZymoBIOMICS Microbial Community DNA Standard, Zymo Research). The successful characterization of the same microbial standard, using the same instrument, but with a different DNA library preparation, the ligation sequencing kit LSK109, and analysis pipeline has been demonstrated earlier (38). There are some fundamental differences between these two DNA library preparation methods: 1) DNA gets more fragmented in the rapid kit, 2) there is potential loss of DNA

in the ligation kit because of several DNA cleaning steps, and 3) the ligation kit is composed of several more steps than the rapid sequencing kit. All of these could affect the microbial profile of a sample, the speed of the sequencing, and in-field usage of the nanopore sequencing device. As shown in Figure 1 and Table 1, the preparation of the DNA library using the RAD004 kit followed by WIMP analysis resulted in a correct classification of the composition of the mock community across different sequencing intervals (<https://epi2me.nanoporetech.com/shared-report-226486?tokenv2=a812d53e-7d47-4294-975c-3550fd037336>). This experiment showed that the RAD004 kit could be used as efficiently as the ligation kit (LSK109) for determining the microbial composition of a sample, albeit with lower output.

Testing metagenomic characterization of agricultural water using nanopore rapid sequencing kits. Our original goal for this project was to test 1) the on-site, fast characterization of the bacterial composition and 2) detection of STECs in culture-independent, concentrated agricultural water samples (Fig. 2) using the Oxford Nanopore hand-held MinION and two versions of DNA library preparation kits (RAD004 rapid sequencing kit and LRK001 field sequencing kit). DNA extraction of each sample (10 ml) resulted in 2.5 ug total DNA per sample. A DNA library for sample 26 was prepared using the RAD004 kit and resulted in a non-productive sequencing reaction with a low output (388,130 reads and 0.4 Gb yield) (Additional File 1). Pore availability at the onset was modest (70%) with only ~40% of pores actively sequencing and steadily declined over the first 24 hours. Almost 60% of the total reads did not pass the quality filter (Additional File 1). Of the reads that passed the quality filter, a large majority were less than 1,000 bp (103,235 reads) and taxonomy was classified by WIMP for about 45% of them (<https://epi2me.nanoporetech.com/shared-report-243787?tokenv2=934deea4-2201-4e0d-bf82-d42c3a03078b>). The rapid sequencing kit does not contain a DNA cleaning step and requires the highest DNA quality to maintain optimal performance. The poor performance of sample 26 suggested that the sample contained an inhibitor or other interference with proper sequencing.

In order to reduce or eliminate this inhibition, samples 17 and 26 were further cleaned with an Agencourt magnetic bead cleaning step (as described in Methods) and prepared for sequencing using the LRK001 or RAD004 kit. The cleaning resulted in a loss of 40% of the total DNA and the sequencing continued to show inhibition, although less pronounced and with better results (Fig. 3). The sequencing run with the LRK001 kit for sample 26 showed a rapid decay of the sequencing pores in less than 24 hours, resulted in low read output (322,000 reads and 0.4Gb yield) (Fig. 3A), and only 50% of reads passing the quality filter (Fig. 3B). A similar result was obtained with the RAD004 kit for the same sample 26, but with a higher read output (1,700,000 reads and 1.8 Gb yield) (Fig. 3C) and a slight increase in the number of reads passing the quality filter (Fig. 3D). Nevertheless, the majority of the read sizes were below 1,000 base pairs in length (~560,000 reads) which resulted in more than 86% of the reads being unclassified (<https://epi2me.nanoporetech.com/shared-report-214408?tokenv2=9ed5fce3-da1c-434c-a388-5d98953a7e1c>). The DNA was highly sheared due to the Agencourt cleaning and the use of the rapid sequencing kit. Sequencing of sample 17 using the RAD004 kit showed similar results. The total read output was 1,680,000 reads with 67% of reads passing the quality filter and 23% of those reads were classified by WIMP. Additionally, more than 430,000 reads were under 1,000 base pairs in length (<https://epi2me.nanoporetech.com/shared-report-214241?tokenv2=e55ce2ae-6d63-4dd4-a1a5-dd3c4f0c567a>).

Use of the LSK109 ligation kit for eliminating agricultural water inhibitors. Because we found that the field sequencing kit and the rapid sequencing kit did not produce satisfactory results for our field application, we decided to test the ligation kit (LSK109). The ligation kit has several advantages over the rapid sequencing kits such as: higher output,

no enzymatic shearing, and with several cleaning steps the inhibitors will diminish to levels that would not interfere with the sequencing reaction. The drawback was that it takes almost 90 minutes for sample preparation compared to the 15 minutes required for the rapid sequencing or field sequencing kits and it contains more steps where sample can be lost. Testing produced promising results with 2,210,000 reads and 8.45 Gb yield with more than 60% of pores sequencing over the 24-hour sequencing run. Furthermore, over 85% of the reads passed the quality filter (Fig. 4). We decided to process three additional samples and compare their taxonomic composition to conduct a baseline metagenomic survey of samples collected across 3.7 contiguous miles in the Southwestern US (Fig. 2).

Agricultural water metagenomic taxonomic characterization. Each run produced an average of 2,200,000 reads with an average total yield of 8.5 Gb (Additional File 2). The base-called reads were passed through a quality filter and reads above 5,000 base pairs in length were analyzed by the EPI2ME WIMP workflow. The agricultural water samples had a diverse composition. The reads were predominately bacterial (89-92%) with the remaining of eukaryotic (6-9%), viral (1-2%), and archaeal (<1%) origin (Additional File 3) that could be organized into approximately 50 phyla, 90 classes, and 1,500 genera (Table 2).

Bacterial composition in agricultural water. We identified the bacterial genera with an abundance greater than 1% in at least one sample (Fig. 5). The 11 bacterial genera identified include *Synechococcus*, *Cyanobium*, *Pseudomonas*, *Streptomyces*, *Flavobacterium*, *Candidatus Fonsibacter*, *Limnohabitans*, *Hydrogenophaga*, *Acidovorax*, *Variovorax*, and *Rubrivivax*. A large portion of reads (~40%) were classified as various taxa, but the combined genus abundance was less than 1%. Sites 9, 10, and 11 displayed very similar composition with 30-40% *Synechococcus*, 4% *Cyanobium*, and 1-2% each *Pseudomonas*, *Streptomyces*, *Flavobacterium*, *Candidatus Fonsibacter*, and *Limnohabitans*. Site 12 is located approximately 6.9 miles from site 11 in a saltwater drainage canal and has a similar abundance of *Streptomyces* (1.3%), and *Limnohabitans* (1.3%). However, site 12 had almost no *Synechococcus* (0.3%), *Cyanobium* (0.1%), and *Candidatus Fonsibacter* (0.1%), While, *Pseudomonas* (3.1%), *Flavobacterium* (5.9%), *Hydrogenophaga* (3.1%), *Acidovorax* (1.9%), *Variovorax* (1.3%), and *Rubrivivax* (1.2%) were each present in approximately 2 - 6 times the abundance as sites 9-11.

Eukaryotic composition in agricultural water. Strikingly, approximately 6 - 9% of the total reads were identified as genus *Homo*. Eukaryotic DNA was represented by 19,645 to 56,173 reads (Additional File 3). Within the eukaryotic reads, approximately 98% of reads were identified as *Homo sapiens*, with the other 2% being largely fungal in origin. Thus, the fungal composition of the agricultural water was minimal.

Detection of STECs. Each sample site was analyzed for the presence of Shiga toxin-producing *E. coli* by the FDA BAM Chapter 4A methods. Sites 9, 10, and 12 were confirmed to be STEC positive after enrichment. Contrary to these results, the WIMP analysis of the nanopore sequencing output of the unenriched agricultural water revealed that between 46 – 152 reads were identified as *E. coli*. The strain-level identification further classified one read from site 11 and 2 reads from site 12 as the O157 serotype. An NCBI BLAST search of those individual reads revealed that only the read from site 11 matched the O157:H7 genome. Due to the limited coverage, strain level identification could not

be obtained. Therefore, the concentration of *E. coli* in unenriched agricultural water samples was not sufficient for the detection of STECs or *E. coli* O157:H7 by direct nanopore sequencing.

Discussion

Agricultural water has been implicated in the contamination of produce-related foodborne illness and outbreaks (15–20). Current FDA protocols for the detection and isolation of STECs require multiple rounds of selective plating and WGS of a single isolate. On-site field testing is increasingly becoming a priority to decrease the time to detection of pathogenic microbes and prevention of prospective corrective measures. We have designed a pilot study to test nanopore sequencing methods for the fast determination of concentrated agricultural water microbiome and detection and classification of STECs. We determined that the rapid library preparation kit produced expected results for a mock microbial community but performed poorly within the agricultural water matrix. The DNA library prepared with the ligation kit improved the sequencing output and the bacterial composition, however, we were unable to accurately detect STECs in concentrated agricultural water.

Mock microbial communities are standardized metagenomic samples and are typically used for benchmarking sequencing studies (38, 40, 41). Mock microbial communities sequenced by long-read nanopore technology using the ligation library preparation kit (LSK109) produced sufficient coverage expected to close all the microbial genomes (38). We have tested the same community as a benchmark for the RAD004 rapid sequencing library preparation kit. The sequencing run produced a total of 4,360,159 reads with an output of 6.5 Gb and with 93,448 reads longer than 5,000 base pairs. The EPI2ME cloud-based service WIMP classifies the reads by taxon and identified each of the expected microbial species. Additionally, we showed that the read abundance, calculated as a percentage of total reads classified, was correlated with the expected microbial proportions.

Oxford Nanopore sequencing has developed rapid and field sequencing DNA library preparation kits (RAD004 and LRK001, respectively) for fast, portable sequencing efforts. These advances have allowed and encouraged researchers to develop in-field testing kits for remote regions to identify the microbial composition in metagenomic samples. Increased access to nanopore technologies can provide rapid information to these remote areas that have previously been subject to outsourcing sequencing, which can take months. Nanopore sequencing has been used for epidemiological surveillance and early detection of Zika (42) and Ebola (43) viruses. In polar environments the relationship between the changing climate and the microbial community has been of particular interest (44–46). Nanopore sequencing has also aided the protection and maintenance of the cassava crop in Africa (39). These successful in-field metagenomic analyses suggest that the technology can be applied to agricultural water for microbiome analysis and possibly foodborne pathogen detection and assembly.

Confident that the rapid kit has the potential to produce a sequencing output appropriate for metagenomic analysis and closing bacterial genomes, we then tested the performance of the rapid sequencing kit on concentrated agricultural water samples. Surprisingly however, nearly half of the output reads failed initial quality control standards for base-calling. Low output despite an additional Agencourt DNA cleanup step, indicated the presence of a carryover inhibitor, such as humic acid which has similar solubility to DNA and is not easily separated (47, 48). The protocol for the LSK109 ligation kit, however, employs additional DNA cleanup steps, which improved the sequencing output and quality of the base-called reads (Fig. 4).

While the ligation kit adds time and resources to the rapid and field DNA library preparation kits, we were able to use the total reads sequencing output to identify 11 genera in the microbiome of the agricultural water. The three sites along the canal displayed remarkable similarity and we were able to distinguish these communities from an

unrelated, distant site. If the microbiome remains relatively constant over a particular distance, these preliminary results suggest that we may be able to reduce the proximity of the sites and that a distance limit can be established for future baseline survey studies using this type of nanopore metagenomic analysis. Overall, this could aid in reducing the sample number and costs associated with sampling and microbiological and/or metagenomic analysis during longitudinal surveys. While the most abundant species in the microbiome is likely to fluctuate seasonally, they may be an indicator to changing populations and importantly may serve as a means to monitor for deviations in water microbial quality.

The ability to detect STECs directly from agricultural water could decrease the time to detection by at least 24 hours since there is no sample enrichment step. Nanopore sequencing is capable of detecting species present in as few as 50 reads or the equivalent of 4 cells (38). Therefore, with the high quality and increased output gained from the ligation kit, we expected to be able to accurately detect the presence of *E. coli* and identify STEC strains. Nanopore sequencing is, however, not typically used as a screening tool due to its high cost compared to other means of detection like qPCR. We have previously established that the limit of complete, fragmented assembly for STECs by nanopore sequencing is 10^5 CFU/ml (49). This is achievable by sample enrichment where low levels of target are amplified many fold, but herewith we aimed to establish the extent to which virulotyping is possible in agricultural water. We obtained species-level detection of *E. coli* with less than 150 reads, but we were unable to make accurate strain-level and virulotype identification, which requires a complete genome (9, 32, 49).

While we applied a high standard to the agricultural water samples, the rapid and field sequencing runs produced data that could identify the microbiome community and better inform water resource managers and others that monitor agricultural water quality with regards to unexpected deviations once a baseline for their particular water source is established (Additional File 4). The ligation kit requires additional time and equipment but produces more output (Gb yield) and produced a higher number of quality reads. This amount of data, though, was unable to detect STECs in unenriched agricultural water probably due to low levels present. Depending on the desired outcome, nanopore technology can provide high quality, informative, long reads and provides access to tools that aid in fast comprehensive analysis through the EPI2ME cloud-based service.

Conclusions

Oxford Nanopore's LRK001 and RAD004 field and rapid sequencing kits can be appropriate for genus or species level identification of microorganisms that are highly abundant. However, the performance of both kits for microbiome characterization from field samples could be affected by the type of sample to be tested, resulting in low number of reads and low sequence quality. On the other hand, the LSK109 ligation kit provided adequate yield with deeper sequencing depth and better pore performance for assessing the metagenomic composition of agricultural water. We were unable to identify the presence of STEC in the sequencing reads which suggests a low *E. coli* concentration was present. The results from this pilot study provide preliminary evidence that MinION sequencing of agricultural water using the ligation kit has the potential to be used for rapid microbiome determination in the field with optimal results for water quality surveillance.

Methods

Agricultural water collection and concentration

An irrigation canal and other water (100L) was collected at multiple sites (9, 10, 11, 12, 17, and 26) (Fig. 2) in the Southwestern United States. Water samples were filtered and concentrated using a Rexeed 25S Ultrafilter (Dial

Medical Supply, Chester Springs, PA) with approximately 650 ml recovered upon backflush, according to the FDA BAM Dead-end Ultrafiltration method described in Chapter 19c (50).

DNA extraction

DNA was extracted directly from a total of 10 ml of concentrated agricultural water. Ten aliquots of 1 ml each were centrifuged at 10,000 x g for 3 minutes, the supernatant was discarded, and the first pellet was resuspended in 800 ml sterile water and used to combine and resuspend the remaining nine aliquots. DNA was extracted by either the ZymoBIOMICS DNA Miniprep kit (Zymo Research, Irvine, CA) (site 26) according to manufacturer's instructions or the Maxwell RSC Cultured Cells DNA kit with a Maxwell RSC Instrument (Promega Corporation, Madison, WI) (sites 9, 10, 11, 12, and 17) according to manufacturer's instructions for Gram-negative bacteria with additional RNase treatment. DNA concentration was determined by Qubit 4 Fluorometer (Invitrogen, Carlsbad, CA) according to manufacturer's instructions.

Test for accurate metagenomic identification of the ZymoBIOMICS Microbial Community DNA Standard using the Oxford Nanopore rapid sequencing kit RAD004

The ZymoBIOMICS Microbial Community DNA Standard (Zymo Research) is composed of 8 bacteria and 2 yeasts (<https://www.zymoresearch.com/collections/zymbiomics-microbial-community-standards/products/zymbiomics-microbial-community-dna-standard>). Representing microorganisms contain a wide range of GC content from 15% to 85%, which allows for assessing biases that could arise because GC content variation. The organisms in this community are distributed equally (12%), except the 2 yeasts (each present at 2%). Four microliters (400 ng) of this ZymoBIOMICS Microbial Community DNA Standard (Zymo Research) was used for preparing the DNA library for sequencing in a MinION device using the rapid sequencing kit RAD004 according to manufacturer's instructions. The library was run in FLO-MIN106 (R9.4.1) flow cells, according to the manufacturer's instructions for 48 hours (Oxford Nanopore Technologies). The runs were live base called using Guppy v3.2.10 included in the MinKNOW v3.6.5 (v19.12.6) software (Oxford Nanopore Technologies). The run was analyzed by the "What's in my pot" (WIMP) workflow contained in the EPI2ME cloud service (Oxford Nanopore Technologies) at 5, 25, and 48 hours.

Metagenomic sequencing, contig assembly, and annotation

DNA recovered from the agricultural water samples underwent a 0.7X (v/v) Agencourt Bead clean-up (Beckman Coulter, Indianapolis, Indiana). DNA was sequenced using a MinION nanopore sequencer (Oxford Nanopore Technologies, Oxford, UK). The sequencing libraries were prepared using either the Rapid Sequencing (SQK-RAD004), the Field Sequencing Kit (SQK-LRK001) or the Genomic DNA by Ligation kit (SQK-LSK109) and run in FLO-MIN106 (R9.4.1) flow cells, according to the manufacturer's instructions for 48 hours (Oxford Nanopore Technologies). The runs were live base called using Guppy v3.2.10 included in the MinKNOW v3.6.5 (v19.12.6) software (Oxford Nanopore Technologies). The initial classification of the reads for each run was done using the "What's in my pot" (WIMP) workflow contained in the EPI2ME cloud service (Oxford Nanopore Technologies). Reads were assessed for quality including a minimum 5,000 bp length filter.

BAM STEC detection

The presence of STEC was determined according to the protocols in Chapter 4A of the FDA Bacterial Analytical Manual (BAM) (<https://www.fda.gov/food/laboratory-methods-food/bam-diarrheagenic-escherichia-coli>). Briefly, 225 ml of each agricultural water sample was enriched by adding an equal volume of 2X modified Buffered Peptone Water with pyruvate (mBPWp) and incubated at 37°C static for 5 hours. Antimicrobial cocktail [Acriflavin-Cefsulodin-Vancomycin (ACV)] was added and incubated at 42°C static overnight (18-24 h). DNA supernatants recovered from boiled samples were analyzed by qPCR detecting *stx1*, *stx2*, and *wzy*.

Metagenomic data accession numbers. The metagenomic sequence data from this study are available in GenBank under BioProject number PRJNA751542.

List Of Abbreviations

STEC

Shiga toxin-producing *Escherichia coli*

HUS

hemolytic uremic syndrome

stx

Shiga toxin gene

LEE

locus of enterocyte effacement

eae

intimin gene

TTSS

type 3 secretion system

WGS

whole genome sequencing

MAG(s)

metagenome assembled genome(s)

WIMP

What's In My Pot taxonomic classification program in Oxford Nanopore's EPI2ME cloud service

Declarations

Ethics approval and consent to participate. Not applicable

Consent for publication. Not applicable

Availability of data and materials. The metagenomic sequence data from this study are available in GenBank under BioProject number PRJNA751542.

Competing interests. The authors declare that they have no competing interests.

Funding. NGE research was supported by funding from the MCMi Challenge Grants Program Proposal #2018-646 and the FDA Foods Program Intramural Funds. MM was supported by funding from the MCMi Challenge Grants Program.

Authors' contributions. EWB, MA, SMM, and NGE designed the study. JK provided agricultural water samples and performed BAM analysis. All library preparation and sequencing were performed by NGE. NGE and MM analyzed the data and wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements. The study was supported by funding from the MCMi Challenge Grants Program Proposal #2018-646 and the FDA Foods Program Intramural Funds. Dr. Maguire acknowledges a Research Fellowship Program (ORISE) for the Center for Food Safety and Applied Nutrition administered by the Oak Ridge Associated Universities through a contract with the U.S. Food and Drug Administration.

Tables

Table 1. Summary of the WIMP output for the nanopore sequencing of the ZymoBIOMICS microbial community DNA standard using the RAD004 kit at different time intervals (5, 24 and 48 hours). Total number of reads identified for each microorganism and their expected percentage distribution in the sample.

Organism	Reads			% expected distribution	% observed distribution 5 hours	% observed distribution 25 hours	% observed distribution 48 hours
	5 hours	25 hours	48 hours				
<i>Pseudomonas aeruginosa</i>	44,852	250,652	356,955	12	9.4	9.7	9.9
<i>Escherichia coli</i>	42,402	234,272	327,769	12	8.9	9.0	9.1
<i>Salmonella enterica</i>	57,801	308,965	426,178	12	12.1	11.9	11.8
<i>Lactobacillus fermentum</i>	65,916	342,047	470,972	12	13.8	13.2	13.0
<i>Enterococcus faecalis</i>	65,494	351,481	484,785	12	13.7	13.6	13.4
<i>Staphylococcus aureus</i>	49,431	278,968	396,196	12	10.3	10.8	11.0
<i>Listeria monocytogenes</i>	66,411	365,095	510,984	12	13.9	14.1	14.1
<i>Bacillus subtilis</i>	64,242	344,286	486,524	12	13.4	13.3	13.5
<i>Saccharomyces cerevisiae</i>	10,539	56,087	76,771	2	2.2	2.2	2.1
<i>Cryptococcus neoformans</i>	10,644	57,929	79,383	2	2.2	2.2	2.2
Total	477,732	2,589,782	3,616,517	100	100.0	100.0	100.0

^aexpected distribution for the ZymoBIOMICS microbial community DNA standard.

Table 2. Taxonomic rank diversity of the metagenomic analysis at each sample site.

Sample Site	Phyla	Class	Order	Family	Genera
9	50	94	196	423	1,469
10	52	94	197	419	1,464
11	55	98	201	431	1,544
12	55	96	200	427	1,516

References

1. Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, et al. Foodborne illness acquired in the United States—major pathogens. *Emerg Infect Dis*. 2011;17(1):7–15.
2. Gonzalez-Escalona N, Meng J, Doyle MP. Shiga Toxin-Producing *Escherichia coli*. *Food Microbiology: Fundamentals and Frontiers*, 5th Edition: American Society of Microbiology; 2019.
3. Beutin L, Martin A. Outbreak of Shiga toxin-producing *Escherichia coli* (STEC) O104:H4 infection in Germany causes a paradigm shift with regard to human pathogenicity of STEC strains. *J Food Prot*. 2012;75(2):408–18.
4. Mellmann A, Bielaszewska M, Köck R, Friedrich AW, Fruth A, Middendorf B, et al. Analysis of collection of hemolytic uremic syndrome–associated enterohemorrhagic *Escherichia coli*. *Emerging Infectious Diseases*. 2008;14(8):1287–90.
5. Tarr PI, Gordon CA, Chandler WL. Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome. *The Lancet*. 2005;365(9464):1073–86.
6. Allos BM, Moore MR, Griffin PM, Tauxe RV. Surveillance for sporadic foodborne disease in the 21st century: the FoodNet perspective. *Clin Infect Dis*. 2004;38 Suppl 3:S115-20.
7. Mead PS, Slutsker L, Dietz V, McCaig LF, Bresee JS, Shapiro C, et al. Food-related illness and death in the United States. *Emerg Infect Dis*. 1999;5(5):607–25.
8. National Advisory Committee on Microbiological Criteria for Foods. Response to questions posed by the Food and Drug Administration regarding virulence factors and attributes that define foodborne Shiga toxin-producing *Escherichia coli* (STEC) as severe human Pathogens. *J Food Prot*. 2019;82(5):724–67.
9. Gonzalez-Escalona N, Kase JA. Virulence gene profiles and phylogeny of Shiga toxin-positive *Escherichia coli* strains isolated from FDA regulated foods during 2010–2017. *PLoS One*. 2019;14(4):e0214620.
10. Kaper JB, Nataro JP, Mobley HL. Pathogenic *Escherichia coli*. *Nat Rev Microbiol*. 2004;2(2):123–40.
11. Garmendia J, Frankel G, Crepin VF. Enteropathogenic and enterohemorrhagic *Escherichia coli* infections: translocation, translocation, translocation. *Infection and Immunity*. 2005;73(5):2573–85.
12. Fischer N, Bourne A, Plunkett D. Outbreak alert! 2015: a review of foodborne illness in the US from 2004–2013. Center for Science in the Public Interest, Washington, DC. 2015.
13. Olaimat AN, Holley RA. Factors influencing the microbial safety of fresh produce: a review. *Food Microbiol*. 2012;32(1):1–19.
14. Tack DM, Ray L, Griffin PM, Cieslak PR, Dunn J, Rissman T, et al. Preliminary incidence and trends of infections with pathogens transmitted commonly through food – Foodborne Diseases Active Surveillance Network, 10 U.S. Sites, 2016–2019 [Available from: https://www.cdc.gov/mmwr/volumes/69/wr/mm6917a1.htm?s_cid=mm6917a1_w#F1_down].
15. Steele M, Odumeru J. Irrigation water as source of foodborne pathogens on fruit and vegetables. *J Food Prot*. 2004;67(12):2839–49.

16. Uyttendaele M, Jaykus L-A, Amoah P, Chiodini A, Cunliffe D, Jacxsens L, et al. Microbial hazards in irrigation water: standards, norms, and testing to manage use of water in fresh produce primary production. *Comprehensive Reviews in Food Science and Food Safety*. 2015;14(4):336–56.
17. Monaghan JM, Hutchison ML. Distribution and decline of human pathogenic bacteria in soil after application in irrigation water and the potential for soil-splash-mediated dispersal onto fresh produce. *J Appl Microbiol*. 2012;112(5):1007–19.
18. Oliveira M, Vinas I, Usall J, Anguera M, Abadias M. Presence and survival of *Escherichia coli* O157:H7 on lettuce leaves and in soil treated with contaminated compost and irrigation water. *Int J Food Microbiol*. 2012;156(2):133–40.
19. Allende A, Monaghan J. Irrigation Water Quality for Leafy Crops: A Perspective of Risks and Potential Solutions. *Int J Environ Res Public Health*. 2015;12(7):7457–77.
20. Environmental assessment of factors potentially contributing to the contamination of romaine lettuce implicated in a multi-state outbreak of *E. coli* O157:H7 [Available from: <https://www.fda.gov/food/outbreaks-foodborne-illness/environmental-assessment-factors-potentially-contributing-contamination-romaine-lettuce-implicated>].
21. Investigation Summary: Factors potentially contributing to the contamination of romaine lettuce implicated in the fall 2018 multi-state outbreak of *E. coli* O157:H7 [Available from: <https://www.fda.gov/food/outbreaks-foodborne-illness/investigation-summary-factors-potentially-contributing-contamination-romaine-lettuce-implicated-fall>].
22. FDA announces new protocol for the development and registration of treatments for preharvest agricultural water. [Available from: <https://www.fda.gov/news-events/press-announcements/fda-announces-new-protocol-development-and-registration-treatments-preharvest-agricultural-water>].
23. Leafy greens STEC action plan [Available from: <https://www.fda.gov/food/foodborne-pathogens/leafy-greens-stec-action-plan>].
24. Gonzalez-Escalona N, Allard MA, Brown EW, Sharma S, Hoffmann M. Nanopore sequencing for fast determination of plasmids, phages, virulence markers, and antimicrobial resistance genes in Shiga toxin-producing *Escherichia coli*. *PLoS One*. 2019;14(7):e0220494.
25. Hoffmann M, Luo Y, Monday SR, Gonzalez-Escalona N, Ottesen AR, Muruvanda T, et al. Tracing origins of the *Salmonella* Bareilly strain causing a food-borne outbreak in the United States. *J Infect Dis*. 2016;213(4):502–8.
26. Gonzalez-Escalona N, Toro M, Rump LV, Cao G, Nagaraja TG, Meng J. Virulence gene profiles and clonal relationships of *Escherichia coli* O26:H11 isolates from feedlot cattle as determined by whole-genome sequencing. *Appl Environ Microbiol*. 2016;82(13):3900–12.
27. Brown E, Dessai U, McGarry S, Gerner-Smidt P. Use of Whole-Genome Sequencing for Food Safety and Public Health in the United States. *Foodborne Pathog Dis*. 2019;16(7):441–50.
28. BAM Chap. 4A: Diarrheagenic *Escherichia coli* [Available from: <https://www.fda.gov/food/laboratory-methods-food/bam-chapter-4a-diarrheagenic-escherichia-coli>].
29. Loman NJ, Constantinidou C, Christner M, Rohde H, Chan JZ, Quick J, et al. A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic *Escherichia coli* O104:H4. *JAMA*. 2013;309(14):1502–10.
30. Huang AD, Luo C, Pena-Gonzalez A, Weigand MR, Tarr CL, Konstantinidis KT. Metagenomics of two severe foodborne outbreaks provides diagnostic signatures and signs of coinfection not attainable by traditional methods. *Appl Environ Microbiol*. 2017;83(3).

31. Leonard SR, Mammel MK, Lacher DW, Elkins CA. Application of metagenomic sequencing to food safety: detection of Shiga Toxin-producing *Escherichia coli* on fresh bagged spinach. *Appl Environ Microbiol*. 2015;81(23):8183–91.
32. Leonard SR, Mammel MK, Lacher DW, Elkins CA. Strain-level discrimination of Shiga toxin-producing *Escherichia coli* in spinach using metagenomic sequencing. *PLoS One*. 2016;11(12):e0167870.
33. Lusk Pfefer T, Ramachandran P, Reed E, Kase JA, Ottesen A. Metagenomic description of preenrichment and postenrichment of recalled Chapati Atta flour using a shotgun sequencing approach. *Genome Announc*. 2018;6(21).
34. Ottesen A, Ramachandran P, Chen Y, Brown E, Reed E, Strain E. Quasimetagenomic source tracking of *Listeria monocytogenes* from naturally contaminated ice cream. *BMC Infect Dis*. 2020;20(1):83.
35. Kovac J, Bakker Hd, Carroll LM, Wiedmann M. Precision food safety: A systems approach to food safety facilitated by genomics tools. *TrAC Trends in Analytical Chemistry*. 2017;96:52–61.
36. Gigliucci F, von Meijenfeldt FAB, Knijn A, Michelacci V, Scavia G, Minelli F, et al. Metagenomic characterization of the human intestinal microbiota in fecal samples from STEC-infected patients. *Front Cell Infect Microbiol*. 2018;8:25.
37. Moss EL, Maghini DG, Bhatt AS. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. *Nat Biotechnol*. 2020.
38. Nicholls SM, Quick JC, Tang S, Loman NJ. Ultra-deep, long-read nanopore sequencing of mock microbial community standards. *Gigascience*. 2019;8(5).
39. Boykin LM, Sseruwagi P, Alicai T, Ateka E, Mohammed IU, Stanton JL, et al. Tree Lab: Portable genomics for early detection of plant viruses and pests in sub-saharan Africa. *Genes (Basel)*. 2019;10(9).
40. Bertrand D, Shaw J, Kalathiyappan M, Ng AHQ, Kumar MS, Li C, et al. Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. *Nat Biotechnol*. 2019;37(8):937–44.
41. McIntyre ABR, Ounit R, Afshinnekoo E, Prill RJ, Hénaff E, Alexander N, et al. Comprehensive benchmarking and ensemble approaches for metagenomic classifiers. *Genome Biology*. 2017;18(1):182.
42. Faria NR, Sabino EC, Nunes MRT, Alcantara LCJ, Loman NJ, Pybus OG. Mobile real-time surveillance of *Zika* virus in Brazil. *Genome Medicine*. 2016;8(1):97.
43. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for *Ebola* surveillance. *Nature*. 2016;530(7589):228–32.
44. Gowers GF, Vince O, Charles JH, Klarenberg I, Ellis T, Edwards A. Entirely off-grid and solar-powered DNA sequencing of microbial communities during an ice cap traverse expedition. *Genes (Basel)*. 2019;10(11).
45. Johnson SS, Zaikova E, Goerlitz DS, Bai Y, Tighe SW. Real-time DNA sequencing in the antarctic dry valleys using the Oxford nanopore sequencer. *J Biomol Tech*. 2017;28(1):2–7.
46. Edwards A, Soares A, Rassner SME, Green P, Félix J, Mitchell AC. Deep Sequencing: Intra-terrestrial metagenomics illustrates the potential of off-grid Nanopore DNA sequencing. *bioRxiv*. 2017:133413.
47. Wnuk E, Wasko A, Walkiewicz A, Bartmiski P, Bejger R, Mielnik L, et al. The effects of humic substances on DNA isolation from soils. *PeerJ*. 2020;8:e9378.
48. Lakay FM, Botha A, Prior BA. Comparative analysis of environmental DNA extraction and purification methods from different humic acid-rich soils. *J Appl Microbiol*. 2007;102(1):265–73.
49. Maguire M, Kase JA, Roberson D, Muruvanda T, Brown EW, Allard M, et al. Precision long-read metagenomics sequencing for food safety by detection and assembly of Shiga toxin-producing *Escherichia coli* in irrigation

50. BAM 19c: Dead-end ultrafiltration for the detection of *Cyclospora cayetanensis* from agricultural water [Available from: <https://www.fda.gov/media/140309/download>].

Figures

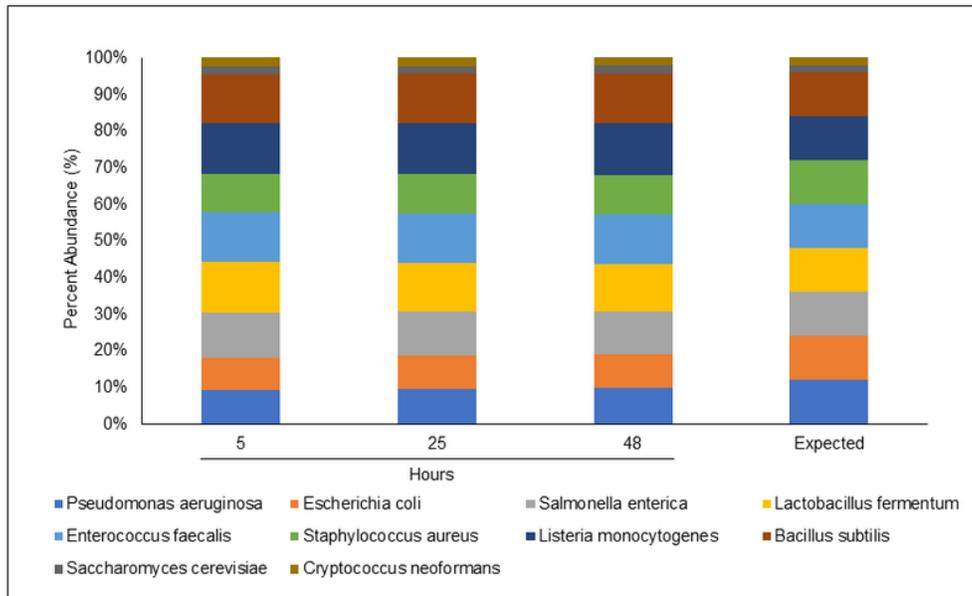


Figure 1

Percentage of microbial composition observed by WIMP from nanopore sequencing of the ZymoBIOMICS microbial community DNA standard using the RAD004 kit at different time intervals (5, 24 and 48 hours) showing that the composition remained stable across those tested time frames and in a similar composition to the expected proportions.

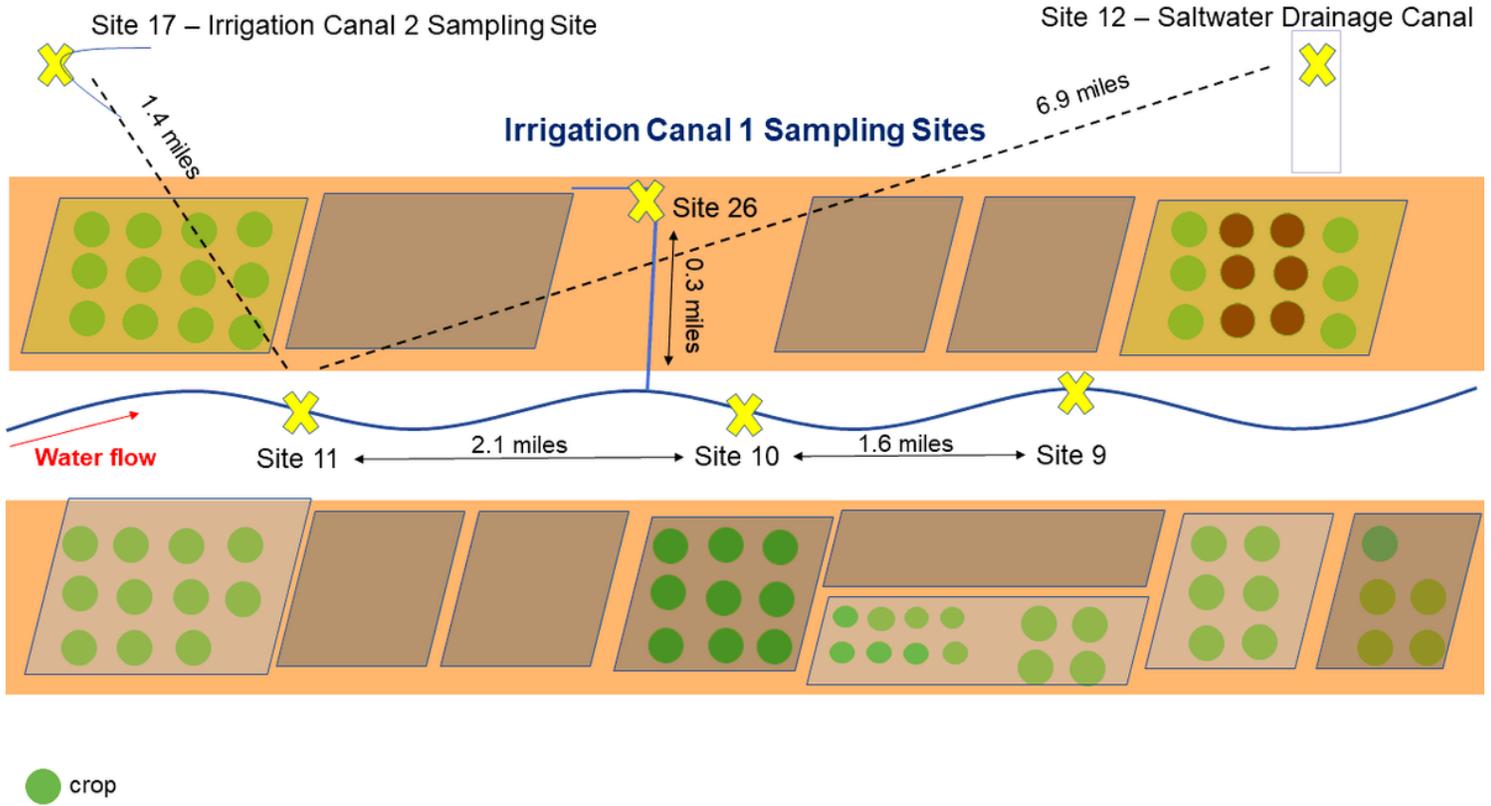
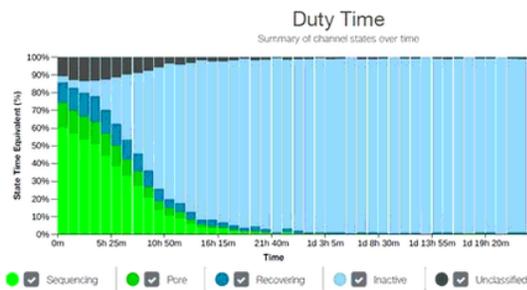


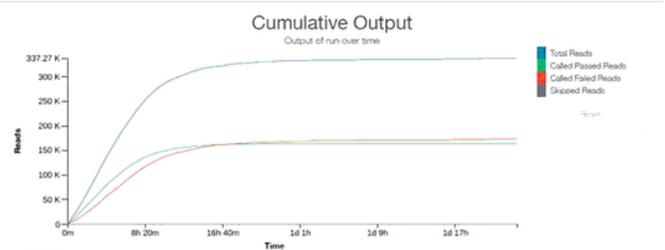
Figure 2

Map of sampling sites. Agricultural water samples were collected along irrigation canals and a saltwater drainage canal in the Southwestern United States. The relative (solid) and direct (dotted) distance between sampling sites 9, 10, 11, 12, 17, and 26 are shown.

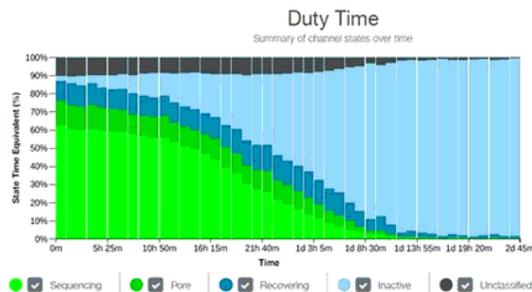
A. LRK001 Field Sequencing Kit – 0.3 M reads



B.



C. RAD004 Rapid Sequencing Kit – 1.7 M reads



D.

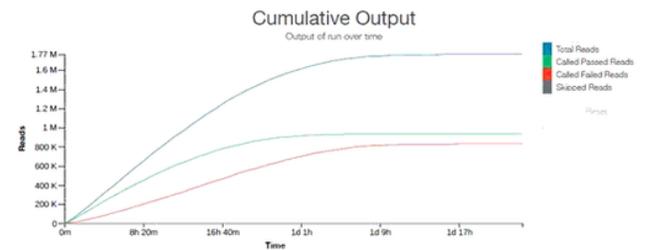
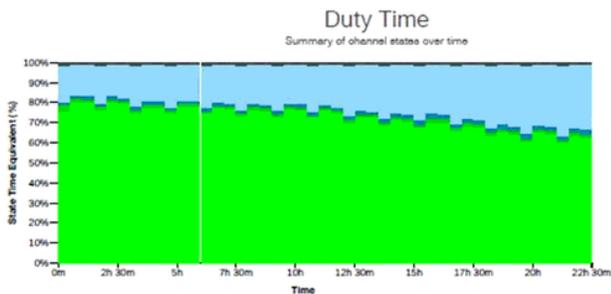


Figure 3

Nanopore sequencing outputs for sample 26 using the LKR001 field sequencing kit and the RAD004 rapid sequencing kit. A) Pore sequencing and availability showing low pore availability and pore death after 16 hours. B) Cumulative output of the LRK001 kit showing that reads passing filter were equal to the reads not passing filter indicative of problems with the sequencing. C) Pore availability for the RAD004 kit showing that the pores decayed by 20 hours of sequencing. D) Cumulative output for the RAD004 kit, with a similar problem with the read quality with 50% passing filter.

A. LSK109 Ligation Kit – 2.2 M reads



B.

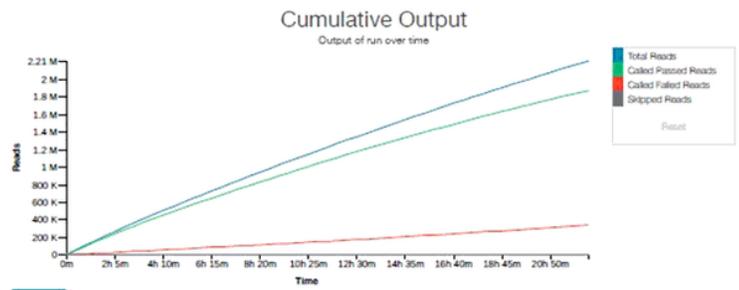


Figure 4

Nanopore sequencing output for sample 9 using the LSK109 ligation kit. A) Pore sequencing and availability remained above 60% for the 24-hour sequencing run. B) Cumulative output of the LSK109 kit showed more than 85% of total reads passed the quality filter.

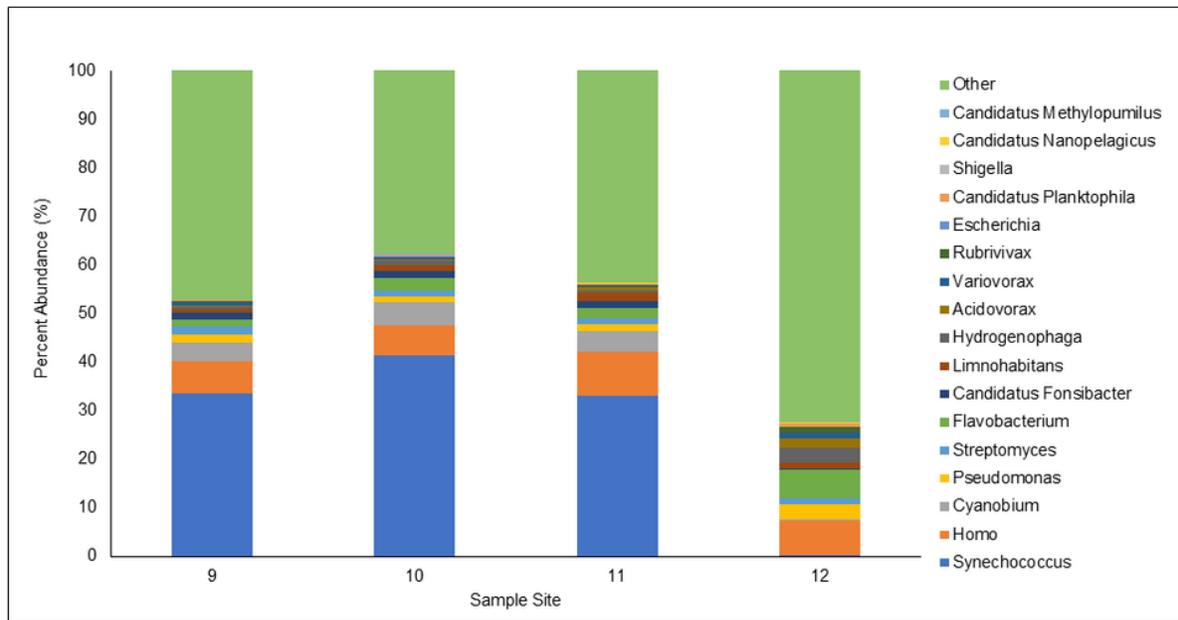


Figure 5

Relative abundance of WIMP identified genera for each agricultural water site. Reads were analyzed by the EPI2ME WIMP workflow. Bacterial genera contributing more than 1% of the classified reads are shown and the sum of the remaining genera identified are included as "Other."

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AdditionalInformation.docx](#)