CENTER for PRODUCE SAFETY

CPS 2021 RFP
FINAL PROJECT REPORT

Project Title
Microbial characterization of irrigation waters using rapid, inexpensive, and portable next generation sequencing technologies

Project Period
January 1, 2022 – December 31, 2023

Principal Investigator
Kerry Cooper, PhD
University of Arizona
School of Animal and Comparative Biomedical Sciences
1177 E. 4th St.
Tucson, AZ 85721-0090
T: 520-621-3342
E: kcooper@email.arizona.edu

Co-Principal Investigators
Kelly R. Bright, PhD
University of Arizona
Department of Environmental Science
Water & Energy Sustainable Technology Center
Tucson, AZ 85745
T: 520-626-8094
E: bright@email.arizona.edu

Walter Q. Betancourt, PhD
University of Arizona
Department of Environmental Science
Water & Energy Sustainable Technology Center
Tucson, AZ 85745
T: 520-621-6366
E: wbetancourt@email.arizona.edu

Channah Rock, PhD
University of Arizona
Maricopa Agricultural Center
Maricopa, AZ 85138
T: 480-628-9794
E: channah@cals.arizona.edu

Objectives

1. *Evaluate the detection limits of the iSeq100 and MinION sequencing technologies for three bacterial pathogens, two viral pathogens, and one protozoan pathogen in irrigation waters of varying quality.*
2. *Use shotgun metagenomics to characterize the microbial communities of irrigation waters from several Southwest regions using the "gold standard" of large amounts of Illumina sequencing and compare to the portable MinION technology.*
3. *Conduct whole genome sequencing, shotgun metagenomic, MinION, and iSeq100 workshops/trainings for the produce industry.*

Funding for this project was provided partly through the CPS Campaign for Research

FINAL REPORT

Abstract

New microbial detection approaches such as next generation sequencing (NGS) are being increasingly applied for tracing microbial contaminants entering the food chain. The produce industry can directly benefit from these powerful new approaches, such as shotgun metagenomics, which allow for the rapid identification of all the bacterial, viral, fungal, and protozoan pathogens in water, soil, or product samples in a single assay, thereby eliminating the need for many different detection assays. Ultimately, these novel approaches will be able to reduce the time and cost of not just food safety surveillance but also plant pathogen surveillance programs by combining everything into a single rapid assay. However, the application of these new NGS approaches need to be verified and validated for use by the produce industry. The goal of this project was to investigate two NGS technologies (Illumina iSeq100 and Oxford Nanopore MinION) that offer slightly different approaches for pathogen detection, and identify the benefits and limitations of each, verify the results, and validate the technologies for use by the produce industry. This study proposed to accomplish the following objectives: (1) Evaluate the detection limits of the iSeq100 and MinION sequencing technologies for three bacterial pathogens, two viral pathogens, and one protozoan pathogen in agricultural waters of varying quality. (2) Use shotgun metagenomics to characterize the microbial communities of agricultural waters from several Southwest regions using the “gold standard” of large amounts of Illumina sequencing and compare to the portable MinION technology. (3) Conduct whole genome sequencing, shotgun metagenomic, MinION and iSeq100 workshops/trainings for the produce industry.

Background

Agricultural water is obviously a potential critical source of contamination for the produce industry. However, with the introduction of next generation sequencing (NGS) technologies there are new tools and approaches that can dramatically improve pathogen detection and overall agricultural water monitoring programs. Currently, the detection of bacterial pathogens like Shiga toxin-producing *Escherichia coli* (STECs) from agricultural water that is approved by the FDA involves PCR or qPCR, isolating the pathogen and performing whole genome sequencing (WGS) on the isolate; however, the FDA itself estimates this can take up to 12 days from start to finish to accomplish [1]. Whole genome sequencing offers the ability to dramatically reduce the time it takes to screen agricultural water or even food products for pathogens and can also reduce costs by screening for multiple pathogens in a single assay. Shotgun metagenomics is a particular WGS approach that involves sequencing all the DNA present in a sample including all bacterial, viral, fungal, protozoan, etc., which provides a very detailed genomic picture of the microbial communities present [2]. The produce industry can directly benefit from this cutting-edge WGS approach as it potentially allows for the rapid identification of all pathogens in an agricultural water, soil, or plant sample in a single assay [3]. Furthermore, the same shotgun metagenomic assay for food safety can also be used for plant pathogen surveillance [4]. Shotgun metagenomics is being actively explored for detection of different pathogens in a number of different types of fresh produce [5] including bagged spinach [6], parsley [7], and lettuce [8], to name a few.

One major question that remains unanswered is whether the use of pathogen enrichment culturing is needed for detection using shotgun metagenomics, as this adds precious time and limits on-site pathogen detection of agricultural water or food products directly in the fields. Recently, the Center for Food Safety and Applied Nutrition (CFSAN) found a 24-hr enrichment of irrigation water followed by shotgun metagenomics could detect *E. coli* O157:H7 at 10^3 CFU/ml, which was significantly lower than the 10^5 CFU/ml for the standard qPCR protocol [1]. Overall, shotgun metagenomics is rapidly becoming the next evolution of food safety surveillance.

Besides the shotgun metagenomics approach, there has also been the advent of less expensive and compact sequencing technologies such as the Illumina iSeq100 system or the Oxford Nanopore MinION device, which serve to further improve the abilities of public health agencies, regulatory agencies, and potentially the produce industry to implement WGS and shotgun metagenomics into food safety surveillance programs. The compact Illumina iSeq100 System at a cost of ~$28,000 delivers a cost-effective solution for independent, small-scale, NGS applications that provide rapid detection of microorganisms for making real-time decisions to protect public health. The truly compact Oxford Nanopore Technology (ONT) MinION is also a less expensive sequencing alternative (as low as ~$1,000). The MinION technology is highly mobile due to its extremely small size (W 4.1" x H 0.9" x D 1.3"; total weight 87 grams) and provides real-time sequencing that begins to provide results within minutes thus allowing for on-site pathogen detection. Moreover, the new MinION Mk1c system contains the MinION sequencer attached to a touchscreen tablet that does all the basic computation requirements such as base calling and initial sequence analysis for the user. This interface can also be integrated with a database of various foodborne pathogens for detection, quantification of the detected pathogen, and can calculate if the pathogen is at an actionable risk level. The MinION Mk1c system allows for complete portability of the sequencer and analysis into the field for on-site sequencing with no specialized equipment and represents the future of food safety surveillance programs for the produce industry as a whole.

Nevertheless, like all new technologies and/or approaches, each must be tested and validated in the laboratory before applying them to the field, the benefits and limitations of the new technologies must be compared, and the results verified against the existing "gold standard". Currently for population-scale genomics like shotgun metagenomics, Illumina is considered the "gold standard" [9]. Despite the exciting potential for improving food safety provided by these new compact and less expensive technologies, neither has been evaluated for their potential benefits, limitations, or requirements for effective implementation by the produce industry. Overall, the goal of this project was to further evaluate these two technologies for use by the produce industry as part of food safety surveillance programs. The actual detection limits for several critical pathogens without pathogen enrichment culturing were determined, and issues were identified based on agricultural water quality or microbial community variation. The results were compared with those generated by the NGS "gold standard" of Illumina sequencing.

Research Methods and Results

<u>Objective 1.</u> Evaluate the detection limits of the iSeq100 and MinION sequencing technologies for three bacterial pathogens, two viral pathogens, and one protozoan pathogen in agricultural waters of varying quality.

Sampling and spiking methods. This objective assessed how sensitive iSeq100 and MinION technologies are for identifying pathogens in agricultural waters of varying quality by determining the detection limit of the iSeq100 and MinION systems for Gram-negative (*E. coli* O157:H7 and *Salmonella*) and Gram-positive (*L. monocytogenes*) bacterial pathogens, two viral pathogens [hepatitis A virus (RNA virus) and adenovirus (DNA virus)], and one eukaryotic pathogen (*Cryptosporidium parvum*). *C. parvum* oocysts were obtained from the University of Arizona's *Cryptosporidium* Production Laboratory, and the remaining pathogens were provided by the PI or co-PIs laboratories for the study. Agricultural water samples were collected by co-PI Rock's team from two different locations around Arizona (Yuma region and Maricopa region) during four different times of the year (winter, spring, summer, and fall). Agricultural water was collected in sterile 10 L containers along the main canal of either agricultural water site. At the point of each agricultural water collection the following data was collected: water pH, air temperature, water temperature, turbidity, total dissolved solids (TDS), and electrical conductivity. The water samples were transported to the Cooper laboratory at the University of Arizona in Tucson, AZ for pathogen spiking and processing. The agricultural water samples from the two locations were spiked with different concentrations (10^1, 10^2, 10^3, and 10^4 cells/ml) and different combinations of the six pathogens (**Table 1**) in three different water qualities (low, medium, or high turbidity). Agricultural water not spiked with any pathogens served as a control for each water quality from each location.

Spiked sample processing methods. After the sample was spiked with the appropriate pathogen inoculum, it was concentrated by membrane filtration involving vacuum filtration through a 0.22 µm pore size filter for bacterial and protozoan pathogen DNA extraction in the Cooper laboratory. For viral pathogens, a stepwise filtration as used through membrane filters of 0.8, 0.65, and 0.22 µm pore sizes followed by centrifugal ultrafiltration using a CentriconPlus-70 filter, 100 kDa cutoff. The concentrate was then used for viral RNA/DNA extraction by co-PIs Bright and Betancourt. DNA from bacterial and protozoan communities was extracted directly from the 0.22 membrane filters initially using the Qiagen PowerWater Kit per the manufacturer's instructions, while the viral concentrates were RNA/DNA extracted using the Qiagen PowerSoil Kits per the manufacturer's instructions. However, after initial spiking experiments it was determined that the Qiagen PowerSoil Kit provided better overall DNA extractions for the bacterial and protozoan pathogen sample too, and was thus used for the experiments for all pathogens during the study. Upon completion of DNA extraction, DNA from each sample was quantified using a Qubit fluorometer and sequencing libraries prepared for each sample for the two different sequencing technologies using the same DNA set from an individual sample for both library preparations.

Sequencing methods. Sequencing libraries for the MinION technology were prepared using the Oxford Nanopore PCR Barcoding Kit according to the manufacturer's instructions. Sequencing libraries were individually barcoded and then pooled together for sequencing for up to 72 hr on the ONT MinION Mk1c device. Samples that do not generate enough data were re-sequenced. All sequencing reads were filtered, processed, and base called by the MinKNOW software (Oxford Nanopore) installed on the MinION Mk1C device, and finalized sequencing reads used for data analysis as outlined below. Sequencing libraries for the Illumina iSeq100 device were prepared using the Illumina DNA Prep Tagmentation Preparation kit and barcoded with the Illumina DNA/RNA unique dual indexes (barcodes) according to the manufacturer's instructions. Barcoded Illumina libraries were pooled together in equal molar ratios and sequenced on the iSeq100 device. After completion of the sequencing by both technologies, the generated sequence data for each sample was randomly divided into sets of different amounts of data [200 megabases (Mb), 400 Mb, 600 Mb, 800 Mb, and 1,000 Mb] and evaluated for the amount of sequence data needed for proper detection each pathogen or combinations of pathogens under different concentrations, water qualities, and agricultural water source location for the two different technologies.

Data analysis methods. The different amounts of data for each of the agricultural water samples sequenced by the two technologies mentioned above were analyzed using two different methods. First, sequence reads from each of the samples were mapped using Geneious Prime software to a reference genome for the targeted pathogen to determine the number of pathogen "hits" or sequence reads matching the pathogen in the sequenced sample. Second, all the sequence reads of each sample were used to taxonomically identify all the members of the microbial community in the sample using Kraken2 software, and the number of "hits" to the targeted pathogens determined and compared to the mapping results. This analysis was repeated for all pathogen-spiked samples sequenced during the study at the different amounts of data. However, it became very clear after the initial sequencing runs and analysis that the Illumina iSeq100 would not generate enough data to effectively detect various pathogens at lower detection levels, therefore it was only used for the first set of samples in the project.

Pathogen detection results. The shotgun metagenomics approach was able to sequence at least 50% of the entire genome of each of the bacterial pathogens at 10^2 CFU/ml, when at least 1 gigabase (Gb) or 500,000 ONT sequencing reads/6.5 million Illumina sequence reads were generated from a sample. This was enough to detect each of the bacterial foodborne pathogens in spiked irrigation water samples consistently down to a detection level of 10^2 CFU/ml without any type of pathogen culture enrichment. The number of sequencing reads mapping or hitting the pathogen genome per 10,000 sequence reads of a sample for each of the bacterial pathogens, *E. coli* O157:H7 (**Figure 1**), *Salmonella* Typhimurium (**Figure 2**), and *L. monocytogenes* (**Figure 3**), decreases as the pathogen concentration decreases in the sample, but there are still pathogen hits even at the lowest level tested in the study (10^1 CFU/ml). Hits per 10,000 reads at each of the tested concentrations in the project indicate that bacterial pathogens can be detected at lower levels, but it will require deeper/more sequencing data to be generated on average for samples with extremely low concentrations (like real-world applications) to be effective at detecting pathogens. This is demonstrated in **Figure 4**, which shows the number of hits or reads that map to a reference *L. monocytogenes* genome when mapping 500,000 ONT sequencing reads or approximately 1 Gb to the reference. At high concentrations (10^4 CFU/ml) there are high numbers of *L. monocytogenes* hits, but even at 10^1 CFU/ml there are still hits mapping to the pathogen to potentially identify its presence in the sample. The study also found that shotgun metagenomics could identify *Cryptosporidium parvum* in the irrigation water samples initially at higher levels of 10^3 oocysts/ml or 10^4 oocysts/ml (**Figure 5**), but this detection level was lowered by conducting different types of DNA extractions with the oocysts (discussed below). While the number of hits per 10,000 reads for *C. parvum* looks similar to those for the bacterial pathogens, it should be noted the X-axis is 10 times lower in Figure 5, and thus the initial detection was not as effective as the bacterial detection. Unfortunately, none of the viral pathogens were effectively detected using shotgun metagenomics regardless of the viral pathogen concentration in the irrigation water sample; it has been reported previously that viral detection with shotgun metagenomics can be challenging. This suggests that additional steps, such as DNA-targeted enrichment or hybrid capture, would be required to detect viral pathogens, but would also require additional costs and time for those samples.

Initial comparison of sequencing technology results. Both ONT MinION and Illumina iSeq100 devices were able to detect bacterial and protozoan pathogens as described above when enough sequencing depth was generated for the sample. However, in order to reach the sequencing depth that is required to detect the various pathogens particularly at the lower detection levels, the Illumina iSeq100 device could only sequence one or two samples at a time, whereas ONT MinION could handle at least 12–24 samples. The number of samples sequenced by Illumina can be increased by using a targeted DNA enrichment step after DNA extraction or hybrid capture techniques to isolate specific target regions prior to sequencing but this has a number of issues: (1) additional costs; (2) additional steps in the process; and (3) only targets limited areas of the pathogen genome thus limiting the genomic information generated with the shotgun metagenomics. Therefore, as mentioned above, most of this study focused on the ONT MinION device for the experiments as it has the most immediate applicability to industry as a potential tool. Yet, it should be reiterated the Illumina "gold standard" technology is also just as effective at pathogen detection using shotgun metagenomics, and Illumina's larger sequencing devices would be more than effective at increasing sample number and thus lowering costs but at a loss of portability and at a much higher equipment cost.

Multiple pathogen detection results. One major potential advantage of shotgun metagenomics over traditional pathogen testing is the ability to theoretically screen and detect numerous different pathogens in a single assay. Shotgun metagenomics approach during this study was able to detect multiple combinations of bacterial and/or protozoan pathogens in a single sample at different concentrations, with the detection levels for each of the different pathogens being nearly identical when the pathogens were alone in the irrigation water sample (**Figure 6**). Again, the detection limit was approximately 10^2 CFU/ml for the bacterial pathogens and 10^3 oocysts/ml for the protozoan pathogen but could be reduced to as low as 10^1 CFU/ml for bacterial and 10^2 oocysts/ml for the protozoan pathogen with enough sequencing depth. Overall, the application of shotgun metagenomics could effectively detect three bacterial pathogens and a protozoan pathogen in a single spiked irrigation water sample by only filtering the water, DNA extracting, and then ONT sequencing.

<u>Effect of irrigation water sample location on pathogen detection results.</u> In order for this tool/technology to be effective for industry, it needs to be applicable in any growing region in the world. Therefore, the study included spiked irrigation water samples from two different locations in Arizona that are known to have different microbiomes (background microbes). The location of the irrigation water did not have an impact on the ability of shotgun metagenomics to detect either bacterial foodborne pathogens or protozoan pathogens. The number of hits per 10,000 sequence reads was very similar between the two locations (**Figure 7**) and demonstrates the detection limits were identical between the locations for all the pathogens. These results suggest that shotgun metagenomics can be applied to agricultural water by industry around the United States and potentially the world without losing any accuracy for pathogen.

<u>Influence of irrigation water turbidity on pathogen detection results.</u> The quality of agricultural water is quite variable, therefore different turbidity levels of the irrigation water from both test locations were utilized to assess the impact of turbidity on the ability for shotgun metagenomics to detect pathogens. The water turbidity levels used for the study were: (1) <50 NTUs (low); (2) 50–200 NTUs (medium); and (3) >200 NTUs (high), and it was found that low and medium turbidity had no impact on the ability to detect different pathogens as previously described. However, there was a major impact in high turbidity water even at the highest pathogen concentration (10^4 CFU/ml), and the overall detection limit in high turbidity increased at least two logs compared to low and/or medium turbidity water samples (**Figure 8**). The research team hypothesizes that the amount of humic acid and organic matter in the high turbidity irrigation water is interfering with the assays, as these have previously been found before to interfere with shotgun metagenomics and sequencing, but additional experiments outside the parameters of this study would need to be conducted to test this hypothesis.

<u>Impact of different DNA extraction method results.</u> Initially, all irrigation water samples were DNA extracted using the Qiagen PowerWater DNA extraction kit, but it was found that there were better DNA yields with the Qiagen PowerSoil DNA extraction kit and thus this kit was used for the remainder of the study for all samples except as outlined below. The PowerSoil can handle organic matter and other inhibitors better than other kits and, as mentioned above, higher turbidity levels with high levels of organic matter can interfere with pathogen detection using shotgun metagenomics. As previously described, the initial detection level for *C. parvum* was 10^3 oocysts/ml when using the Qiagen PowerSoil DNA extraction kit on the filtered sample, but the research team felt there was an issue of poor DNA extraction efficiency from the oocysts as they possess a thick lipid-rich wall that may require alternative extraction procedures to obtain good quality nucleic acid for sequencing. In consultation with a *C. parvum* research laboratory at the University of Arizona, it was determined that most effective DNA extraction method for oocysts is multiple freeze/thaw cycling with liquid nitrogen. All the samples spiked with *C. parvum* oocysts were redone with the liquid nitrogen method and compared against the previous Qiagen methods for the ability to detect *C. parvum*. The liquid nitrogen method dramatically improved the ability to detect *C. parvum* in the samples, and reduced the detection limit down to 10^2 oocysts/ml (**Figure 9**),

which reaches a detection level similar to those for bacterial pathogens. The results indicate that the method of DNA extraction needs to be standardized to find the most effective for all pathogens prior to implementation of shotgun metagenomics as a tool.

Sequence depth results. As previously mentioned, increasing the sequencing depth/amount of sequencing generated for a sample appeared to increase the ability to detect the pathogens in various samples, particularly those near or below the detection limit (e.g. 10^2 or 10^1). Examining the number of pathogen hits at different sequence depth levels [200 Mb (~100,000 reads), 400 Mb (~200,000 reads), 600 Mb (~300,000 reads), 800 Mb (~400,000 reads), and 1,000 Mb (~500,000 reads)] found that the number of hits per 10,000 reads did not significantly change for samples spiked at 10^2 CFU/ml or 10^1 CFU/ml (**Figure 10**), which is expected and a sign of equal distribution of sequencing across the samples. However, the overall number of reads hitting the pathogen genome increased as the total number of sequencing reads increased, for example, an average of 10 hits per 10,000 reads would give 100 hits at 100,000 reads but 500 hits at 500,000 reads (**Figure 11**). Therefore, sequencing depth or the amount of total sequencing for a sample is critical to effectively detecting a pathogen at lower concentrations and could be utilized to potentially help to improve detection limits using shotgun metagenomics.

Objective 2. Use shotgun metagenomics to characterize the microbial communities of agricultural waters from several Southwest regions using the "gold standard" of large amounts of Illumina sequencing and compare to the portable MinION technology.

Irrigation water microbiome methods. This objective characterized the microbial communities (microbiome) of agricultural water from six different locations in Arizona during four different seasonal time points. Community baselines were established using Illumina sequencing (current "gold standard" of shotgun metagenomics) on the more expensive non-portable Illumina NovaSeq, which produces significantly more sequence data than the iSeq100 at a higher price, and then were compared to the microbial communities identified from the same samples by MinION sequencing. Ten-liter (10 L) agricultural water samples were collected from a main canal at three areas in two Arizona growing locations during each of the four seasonal time points for both years of the study. As described above, the following data was collected for each sample at the point of collection: water pH, air temperature, water temperature, turbidity, total dissolved solids, and electrical conductivity. All agricultural water were transported back to the Cooper laboratory at the University of Arizona for processing. Processing, DNA extraction, Illumina barcoding, Illumina sequencing library preparation, MinION barcoding, MinION library preparation, and MinION sequencing of all samples was conducted as described above. Pooled barcoded Illumina sequencing libraries were sent to the North Carolina State University Genomic Sciences Center and sequenced on a single lane of an Illumina NovaSeq S4 flow cell. Barcoded MinION sequencing libraries were sequenced on the MinION Mk1C devices in the Cooper laboratory. Each sequenced sample for both technologies was analyzed for the microbial communities using Kraken2 software to identify the bacterial, fungal, viral, and protozoan communities that are present in the different agricultural waters, and also to provide relative abundance estimates for each microbial member. Comparisons of the irrigation water microbiome between the two sequencing technologies was conducted using the PhyloSeq package in R.

Irrigation water microbiome generated by different sequencing technologies results. To confirm that these two technologies identify the abundance and profile of microbial populations at similar accuracy, both technologies were used for assessing the irrigation water microbiome utilizing the same DNA from the sample. Interestingly, the technologies did not produce similar results, as the microbial communities for the different samples clustered together based on sequencing technology and not the individual sample, and it was consistent for both Yuma and Maricopa, AZ, location samples (**Figure 12**). Furthermore, examining the abundance of the microbial community members for the different samples between the two sequencing technologies revealed that there was quite a bit of variation between the technologies, which helps explain why the samples clustered by technology and not location (**Figure 13**). This indicated that there are issues with the technologies for shotgun metagenomics that need to be assessed, particularly if one technology is inaccurate or if there is another issue is going on between the technologies.

Positive control comparison. Although not part of the scope of this project, an American Type Culture Collection (ATCC) positive control microbiome sample that is composed of equal portions of a diverse group of 20 different bacteria (**Table 2**) was sequenced several times by both technologies to assess the accuracy of both technologies. The positive control is composed of 5% of each of the 20 strains, and both technologies were fairly accurate with the results. However, there were some bacterial members that both technologies had issues with in the ATCC positive control including *Schaalia odontolytica*, *Porphromonas gingivalis*, and *Cereibacter sphaeroides*, whereas the Illumina NovaSeq ("gold standard") also had issues with *Streptococcus agalactiae* and the ONT MinION with *Bifidobacterium adolescentis* (**Figure 14**). Overall, it appears that the issue between the technologies in identifying microbiome is predominately due to variation in the level of abundance of the different bacterial members of the community and not really omitting or missing a particular bacterial species. The ONT MinION had a better evenness of the positive control that was closer to the actual composition in comparison to the Illumina NovaSeq, but neither of the sequencing technologies were perfect match.

Objective 3. Conduct whole genome sequencing, shotgun metagenomic, MinION and iSeq100 workshops/trainings for the produce industry.

Genome workshop methods. To educate produce industry stakeholders about whole genome sequencing, new sequencing technologies, the results of this project, and how these technologies can be applied to food safety surveillance programs, throughout the project, PI Cooper and co-PI Rock conducted in-person workshops/trainings in Yuma, AZ, and Tucson, AZ, for stakeholders or other interested parties. The training perspective consisted of a hands-on demonstration of the MinION Mk1c and computational data analysis for pathogen detection. Industry training materials were generated and disseminated to each of the participants prior to the workshops.

"How do you genome" workshop results. In total, the project team conducted four in-person workshops that covered a variety of topics about whole genome sequencing, shotgun metagenomics, and the project in general. A summary of the "How do you genome" workshops is provided in **Table 3**, including dates, locations, and attendance numbers. Briefly, the first workshop covered general information about the study and whole genome sequencing, whereas the second and third workshops focused on hands-on training for industry about using the NCBI pathogen detection website for pathogen detection. The fourth workshop included a representative from Oxford Nanopore to address any industry questions about applying the technology that were outside the PI and co-PIs knowledge, and hands-on demonstrations of the MinION Mk1c device. After consultation with several members from industry and the CPS technical committee, it was decided to shift the last couple of workshops to focus more on agricultural water. Therefore, pre-recorded presentations were generated covering (1) a general introduction to whole genome sequencing; (2) application of the shotgun metagenomics for agricultural water testing; and (3) general findings of this study. These pre-recorded presentations are available from the PI and co-PI Rock upon request from industry representatives, and current efforts are working to get them advertised and posted to a broad range of industry members.

Outcomes and Accomplishments

Numerous irrigation water samples from two different regions in Arizona were collected, spiked with different combinations of pathogens, processed without any type of pathogen culture enrichment, and then sequenced on two different devices, ONT MinION and Illumina iSeq100. The ability of shotgun metagenomics to detect bacterial, viral, and protozoan pathogens from the spiked irrigation water samples was examined and the detection limit for the different pathogens was also assessed. The ability of shotgun metagenomics to detect multiple pathogens in a single assay were also assessed to confirm it can eliminate the need for testing for each individual pathogen in the future. The ability of shotgun metagenomics to perform pathogen detection in irrigation water from different locations and/or different turbidities was also evaluated, which allows for an initial understanding of any potential issues that could limit the ability of the technology to be employed by industry. Efficacy of several different DNA extraction kits from Qiagen were assessed to determine the best extraction approach for shotgun metagenomics from irrigation water samples. Additionally, due to the toughness of protozoan pathogen oocysts, alternative methods for DNA extraction from this type of pathogen were explored to determine the best method available. The amount of sequencing or sequencing depth needed to effectively identify various foodborne pathogens was examined particularly at the lowest detection limits identified, which allows for understanding if higher amounts of sequencing could improve the detection limits of shotgun metagenomics as a pathogen detection tool.

Additionally, the two sequencing technologies, ONT and Illumina, were compared for the ability of both to accurately identify the irrigation water microbiome from different locations and turbidities. After determining issues with the two technologies to identify the same microbial communities from irrigation water, an ATCC positive control was used to assess potential issues or inaccuracies by the different sequencing technologies. Finally, multiple workshops were held for industry to discuss whole genome sequencing basics, outline of this project, and hands-on training of different bioinformatic software/website for industry members to work with CDC and FDA data that is publicly available.

Summary of Findings and Recommendations

This study demonstrated the power of shotgun metagenomics to reduce sampling turnaround times by potentially eliminating the need for enrichments and other pathogen confirmation requirements. There are still significant hurdles that need to be overcome before it can be effectively applied to a company's food safety program for routine surveillance/testing of agricultural water or other types of samples. Below are the critical findings and the needed steps to move the process closer to a usable technology/tool for industry:

Positive Results:

- A sequencing depth of 1,000 Mb/1.0 Gb or approximately 500,000 ONT MinION sequencing reads was effective at identifying all tested bacterial pathogens (*Salmonella*, *L. monocytogenes*, and *E. coli* O157:H7) in irrigation water samples at 10^2 CFU/ml or 100,000 CFU/L.
- Shotgun metagenomics using ONT MinION device was highly effective at identifying multiple pathogens in a single assay, meaning a single sample can be processed once for all bacterial foodborne pathogens reducing costs and time.
- ONT MinION device can effectively confirm the presence of a pathogen in a sample in less than 96 hours total without culture enrichment using shotgun metagenomics.

- Shotgun metagenomics using ONT MinION effectively detected the protozoan pathogen (*Cryptosporidium parvum*) in irrigation water samples at 10^2 oocysts/ml or 10,000 oocysts/L, but methods for DNA extraction were critical to effectively detect *C. parvum* oocysts in the samples particularly at lower concentrations.
- Location of the irrigation water sample did not impact the ability to detect bacterial or protozoan pathogens using shotgun metagenomics, therefore indicating that this technology/tool has the potential to work in other growing regions.

Negative Results:

- Shotgun metagenomics failed to detect the selected viral pathogens and will thus require additional methods in the sample processing to effectively confirm the presence of viral foodborne pathogens.
- *Cryptosporidium* requires specialized DNA extraction methods to effectively be detected using shotgun metagenomics, which limits the ability to rapidly detect this specific pathogen with other types of foodborne pathogens.
- Illumina iSeq100 technology could effectively detect pathogens using shotgun metagenomics similar to ONT MinION, but the amount of sequencing needed to detect at even 10^4 CFU/ml would be 5–10x higher than for ONT MinION.
- The current detection limit of 10^2 CFU/ml or 100,000 CFU/L for bacterial foodborne pathogens using shotgun metagenomics is 3 or 4 logs higher than is required to meet real world application demands by industry. Nevertheless, improved concentration methods could help to lower this detection limit.
- Higher turbidity levels caused issues with the ability to detect any foodborne pathogens, although at $\geq$200 NTUs, these turbidity levels were significantly higher than typically seen in agricultural water.
- Based on ATCC positive control experiments, there were differences in the two sequencing technologies' ability to identify the abundance of various bacteria in irrigation water microbiome analysis samples, which raises concerns about this technology/tool, but does not appear to impact foodborne pathogen detection.

Next Step Recommendations:

- Develop methods that can reduce the detection limit to real world levels (<10 CFU/L) for bacterial foodborne and protozoan pathogens.
- Combine DNA extraction method/protocols for effectively extracting DNA from foodborne coccidia (e.g. *Cryptosporidium*, *Cyclospora*, Toxoplasma) and bacterial foodborne pathogens (e.g. *Salmonella*, *Listeria*, *E. coli* O157:H7, etc.) simultaneously to maximize the detection efficiency of these pathogens.
- Create a standardized method that will effectively identify any viral foodborne pathogens from agricultural water and can be integrated into a system with other types of foodborne pathogens to allow shotgun metagenomics to identify all types of foodborne pathogens from a single sample.
- Establish the exact number of sequencing reads/sequencing depth required to accurately identify any foodborne pathogen in an agricultural water sample at <10 CFU/L or lower every single time.
- Build a standardized computational pipeline that is simply to use to conduct the analysis needed to identify any foodborne pathogen from an agricultural water sample.
- Determine components of high-water turbidity that interfere with pathogen detection and generate standardized methods to overcome them.

- Explore the benefits of targeted sequencing technologies for deeper and better uniformity of coverage that can lead to greater analytical sensitivity, particularly for detection of foodborne viruses.

The ONT MinION device is handheld and portable for on-site testing of agricultural water or other sample types for foodborne pathogens, however it will probably never be directly employed by growers or other industry representatives. It can still be easily utilized by small local testing laboratories or others that currently conduct frequent pathogen testing for industry, but all the above-mentioned next steps should be addressed before it is utilized for testing samples as part of a food safety program. However, industry representatives should still understand the capabilities and limitations to understand how to best apply this tool to their company's food safety program when it does become feasible.

References cited

1. **Maguire M, Kase JA, Roberson D, Muruvanda T, Brown EW et al.** Precision long-read metagenomics sequencing for food safety by detection and assembly of Shiga toxin-producing *Escherichia coli* in irrigation water. *PLoS One* 2021;16(1):e0245172.
2. **Hoper D, Grutzke J, Brinkmann A, Mossong J, Matamoros S et al.** Proficiency Testing of Metagenomics-Based Detection of Food-Borne Pathogens Using a Complex Artificial Sequencing Dataset. *Front Microbiol* 2020;11:575377.
3. **Rusinol M, Martinez-Puchol S, Timoneda N, Fernandez-Cassi X, Perez-Cataluna A et al.** Metagenomic analysis of viruses, bacteria, and protozoa in irrigation water. *Int J Hyg Environ Health* 2020;224:113440.
4. **Boykin LM, Sseruwagi P, Alicai T, Ateka E, Mohammed IU et al.** Tree Lab: Portable genomics for Early Detection of Plant Viruses and Pests in Sub-Saharan Africa. *Genes (Basel)* 2019;10(9).
5. **Lewis E, Hudson JA, Cook N, Barnes JD, Haynes E**. Next-generation sequencing as a screening tool for foodborne pathogens in fresh produce. *J Microbiol Methods* 2020;171:105840.
6. **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):8183-8191.
7. **Fernandez-Cassi X, Timoneda N, Gonzales-Gustavson E, Abril JF, Bofill-Mas S et al.** A metagenomic assessment of viral contamination on fresh parsley plants irrigated with fecally tainted river water. *Int J Food Microbiol* 2017;257:80-90.
8. **Ogunremi D, Dupras AA, Naushad S, Gao R, Duceppe MO et al.** A New Whole Genome Culture-Independent Diagnostic Test (WG-CIDT) for Rapid Detection of *Salmonella* in Lettuce. *Front Microbiol* 2020;11:602.
9. **Greig DR, Jenkins C, Gharbia S, Dallman TJ**. Comparison of single-nucleotide variants identified by Illumina and Oxford Nanopore technologies in the context of a potential outbreak of Shiga toxin-producing *Escherichia coli*. *Gigascience* 2019;8(8).

APPENDICES

Publications and Presentations

Presentations:

Cooper, KK. AOAC International Annual Meeting, New Orleans, LA. “Microbial characterization of agricultural waters using the rapid, inexpensive, and portable Oxford Nanopore MinION” (August 2023).

Cooper KK. Southwest One Health Symposium, Flagstaff, AZ. “Utilizing genomics to address produce safety challenges” (November 2023).

Publications:

Two manuscripts are currently in preparation for publication.

Budget Summary

The research project was awarded $322,942, and the funds remaining (as of January 31, 2024) will used for the PI and co-PIs to attend the 2024 CPS Research Symposium in June 2024.

Tables 1–3 and Figures 1–14 (See below)

Table 1. Mixed pathogen spiking protocol used for irrigation water samples.

Bacterial Pathogen	**Viral Pathogen**	**Protozoan Pathogen**
Salmonella Typhimurium *Escherichia coli* O157:H7 *Listeria monocytogenes*	None	None
None	Hepatitis A virus Adenovirus	None
Salmonella Typhimurium *Escherichia coli* O157:H7 *Listeria monocytogenes*	None	*Cryptosporidium parvum*
None	Hepatitis A virus Adenovirus	*Cryptosporidium parvum*
Salmonella Typhimurium *Escherichia coli* O157:H7 *Listeria monocytogenes*	Hepatitis A virus Adenovirus	None
Salmonella Typhimurium *Escherichia coli* O157:H7 *Listeria monocytogenes*	Hepatitis A virus Adenovirus	*Cryptosporidium parvum*
Salmonella Typhimurium	None	None
Escherichia coli O157:H7	None	None
Listeria monocytogenes	None	None
Salmonella Typhimurium	None	*Cryptosporidium parvum*
Escherichia coli O157:H7	None	*Cryptosporidium parvum*
Listeria monocytogenes	None	*Cryptosporidium parvum*
Salmonella Typhimurium *Escherichia coli* O157:H7	None	None
Escherichia coli O157:H7 *Listeria monocytogenes*	None	None
Salmonella Typhimurium *Listeria monocytogenes*	None	None
Salmonella Typhimurium *Escherichia coli* O157:H7	None	*Cryptosporidium parvum*
Escherichia coli O157:H7 *Listeria monocytogenes*	None	*Cryptosporidium parvum*
Salmonella Typhimurium *Listeria monocytogenes*	None	*Cryptosporidium parvum*

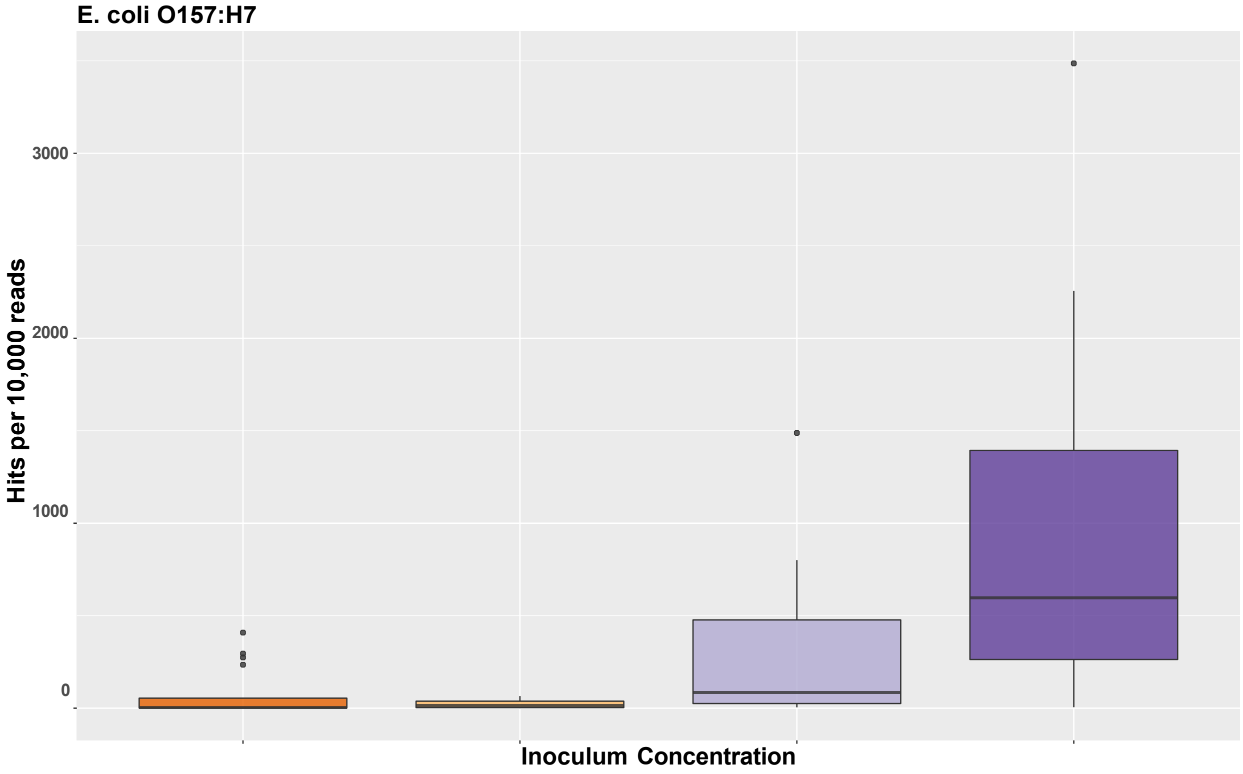

Figure 1. Number of sequence read hits to *Escherichia coli* O157:H7 for Oxford Nanopore MinION sequencing at different spiked inoculum concentrations in the irrigation water. Results are standardized for different amounts of sequencing, and thus represent the number of *E. coli* O157:H7 hits per 10,000 sequence reads from spiked irrigation water sample.

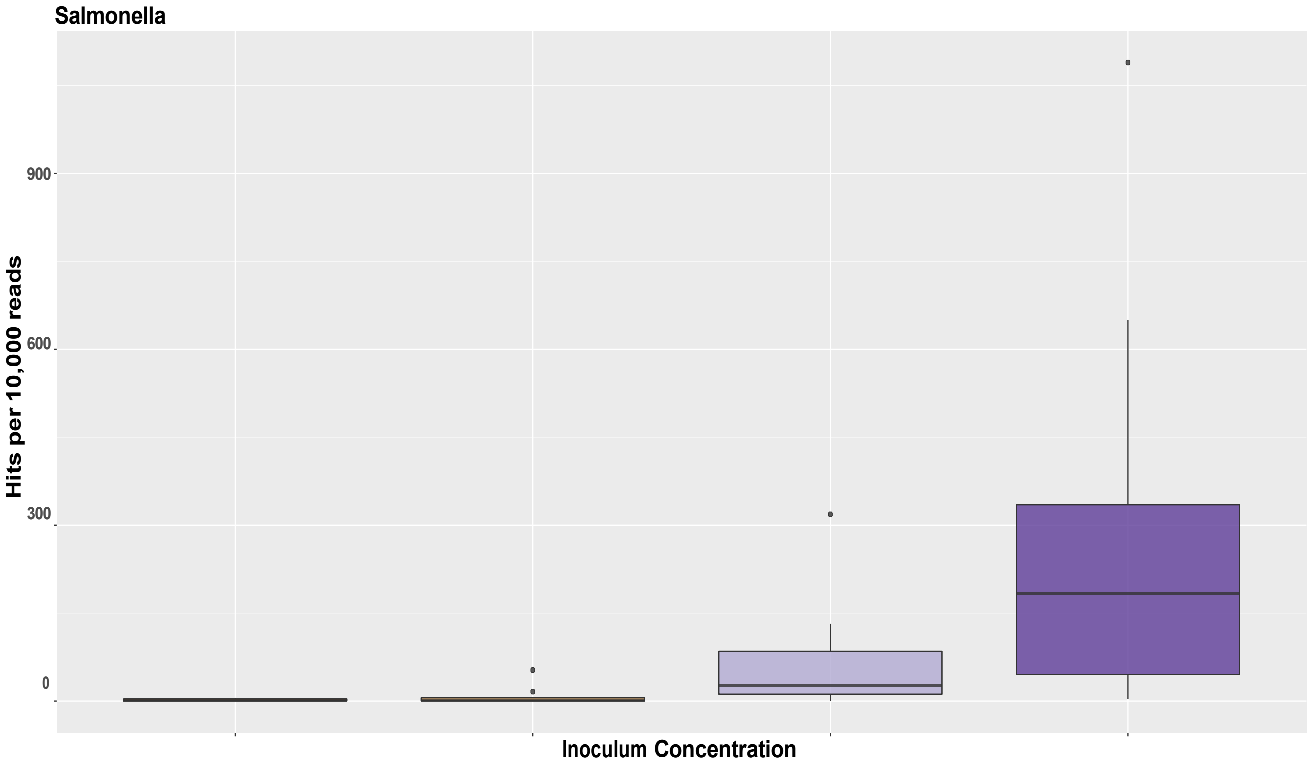

Figure 2. Number of sequence read hits to *Salmonella enterica* subsp. *enterica* serotype Typhimurium for Oxford Nanopore MinION sequencing at different spiked inoculum concentrations in the irrigation water. Results are standardized for different amounts of sequencing, and thus represent the number of *S.* Typhimurium hits per 10,000 sequence reads from spiked irrigation water sample.

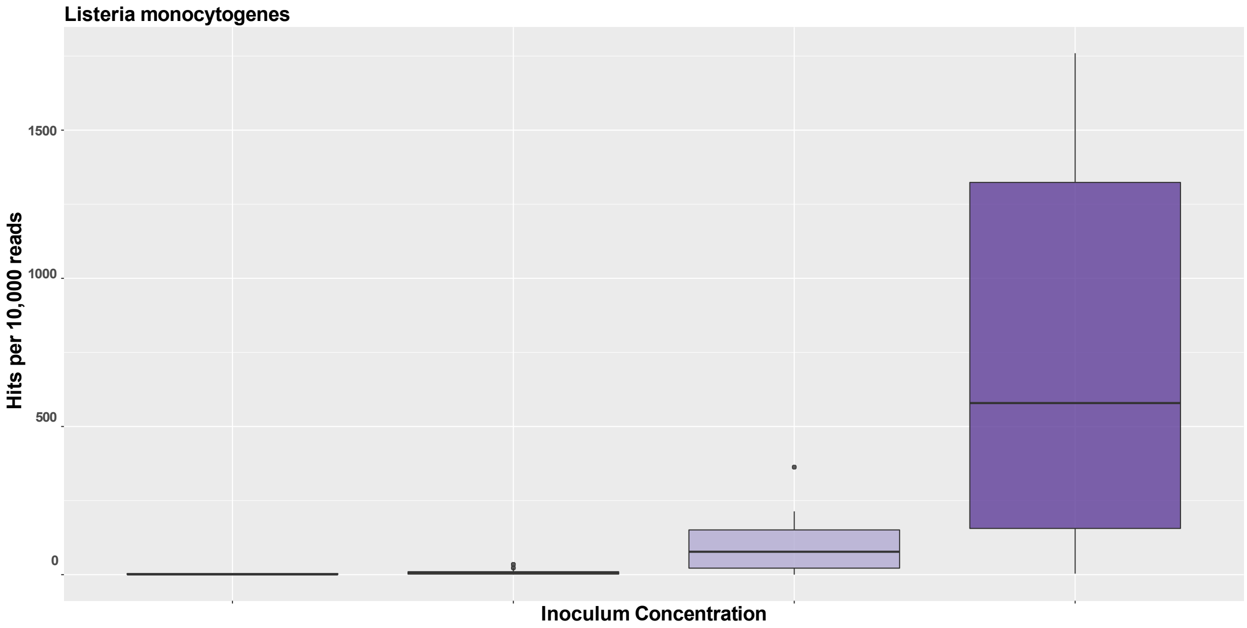

Figure 3. Number of sequence read hits to *Listeria monocytogenes* for Oxford Nanopore MinION sequencing at different spiked inoculum concentrations in the irrigation water. Results are standardized for different amounts of sequencing, and thus represent the number of *L. monocytogenes* hits per 10,000 sequence reads from spiked irrigation water sample.

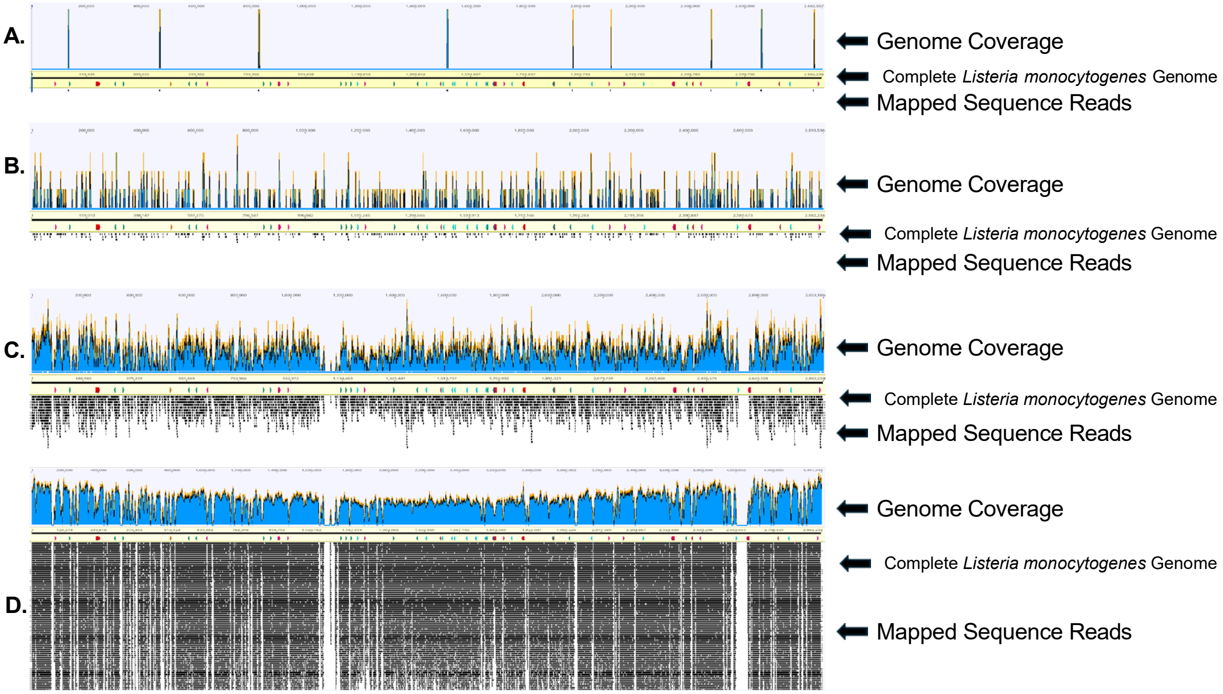

Figure 4. Genomic maps of sequence read hits to *Listeria monocytogenes* at different spiked inoculum concentrations, which demonstrates the amount of the genome that is covered by the sequencing at that particular concentration. A. Spiked 10^1 CFU/ml in 1 L of irrigation water; B. Spiked 10^2 CFU/ml in 1 L of irrigation water; C. Spiked 10^3 CFU/ml in 1 L of irrigation water; D. Spiked 10^4 CFU/ml in 1 L of irrigation water. Genome coverage indicates the amount of sequence reads mapping to that particular region, and is represented by the blue graph on each of the genome maps. Solid black line in the middle of each genome map represents the entire *L. monocytogenes* genome including all 2.8 Mb of the genome. Each black line in the region marked as mapped sequence reads represents a sequence read mapped to the genome, thus significantly more reads are mapped to the genome at 10^4 CFU/ml compared to 10^1 CFU/ml.

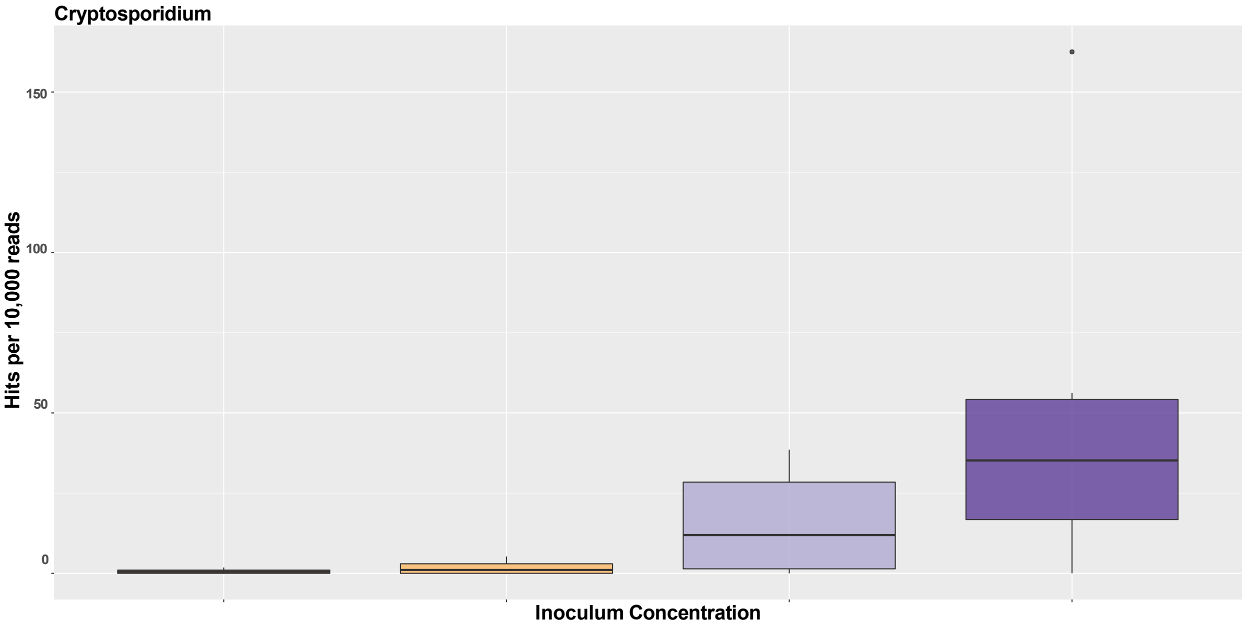

Figure 5. Number of sequence read hits to *Cryptosporidium parvum* for Oxford Nanopore MinION sequencing at different spiked inoculum concentrations in the irrigation water. Results are standardized for different amounts of sequencing, and thus represent the number of *C. parvum* hits per 10,000 sequence reads from spiked irrigation water sample. Note Y-axis amounts are reduced 10-fold for *C. parvum* compared to different bacterial pathogens.

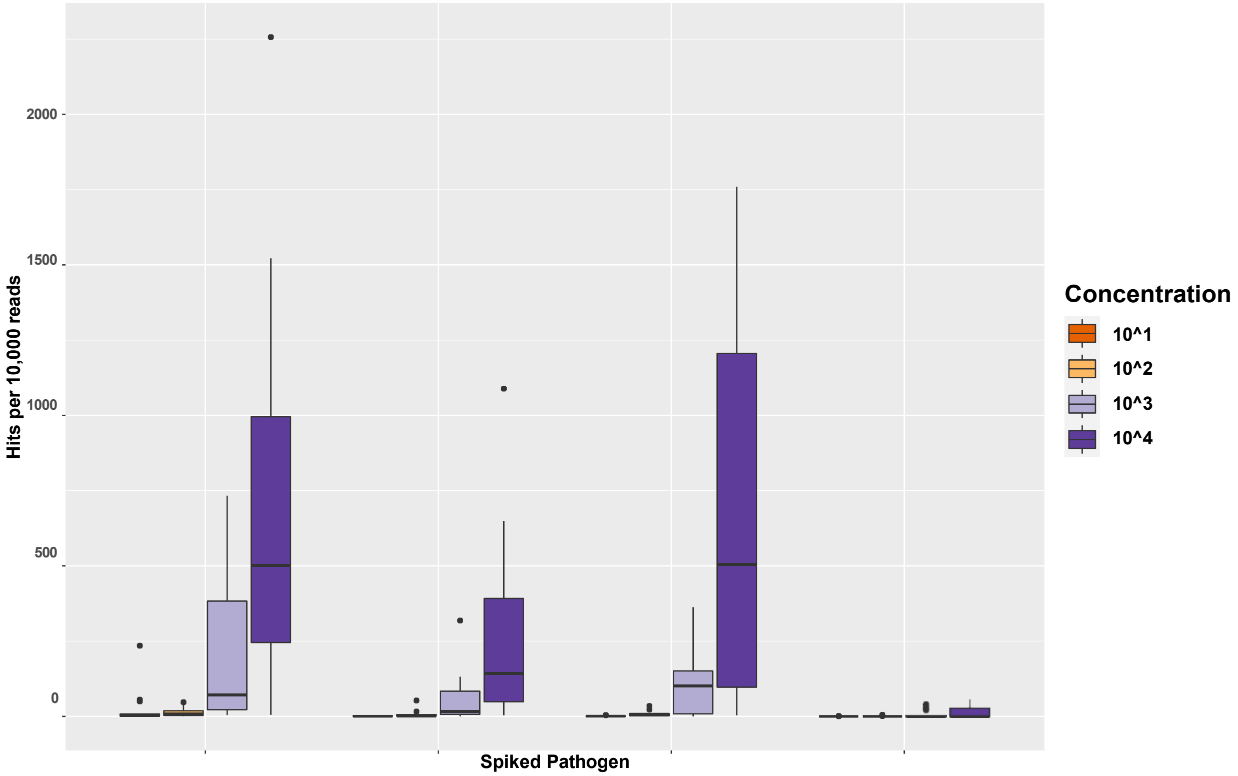

Figure 6. Ability to detect multiple foodborne pathogens from in a single assay. Number of different pathogens genome hits per 10,000 Oxford Nanopore MinION sequencing reads at different spiked inoculum concentrations. Each sample had all four pathogens at the same spiked concentration.

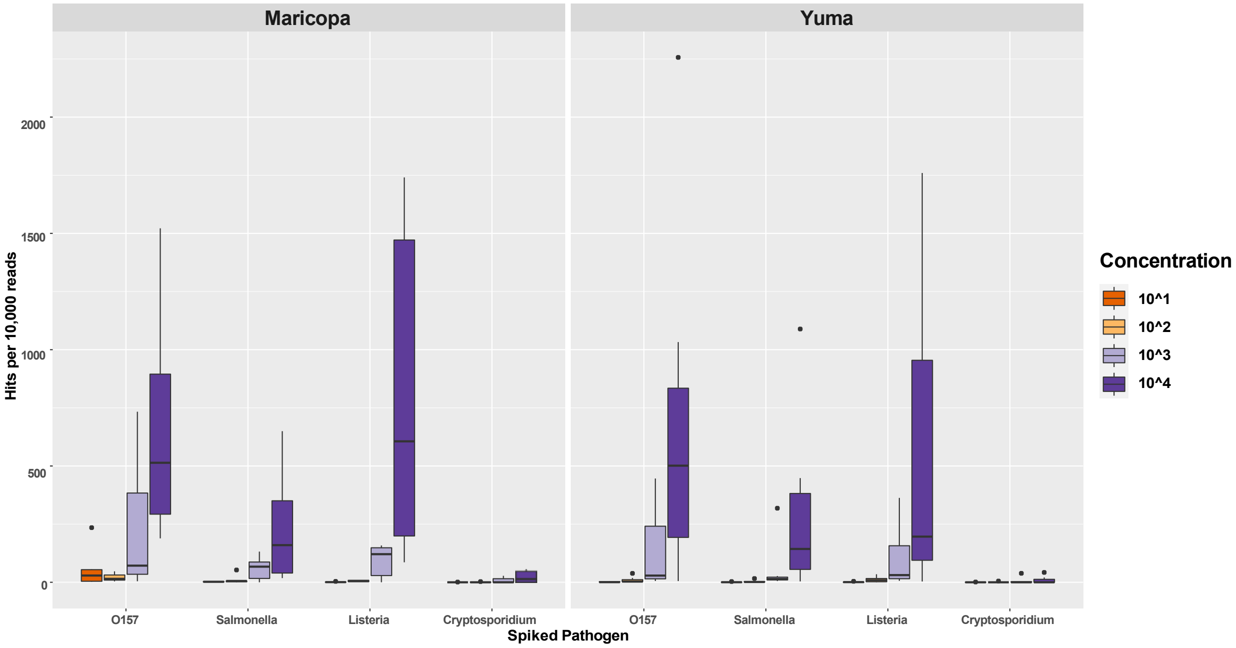

Figure 7. Effect of location on the ability to detect multiple spiked pathogens in a single assay for irrigation water using Oxford Nanopore MinION sequencing. Demonstrates that different irrigation water background microbiome does not interfere with ability to identify multiple pathogens in a single spiked sample. Each plot represents number of pathogen genome hits per 10,000 Oxford Nanopore sequencing reads. Each sample had all four pathogens at the same spiked concentration.

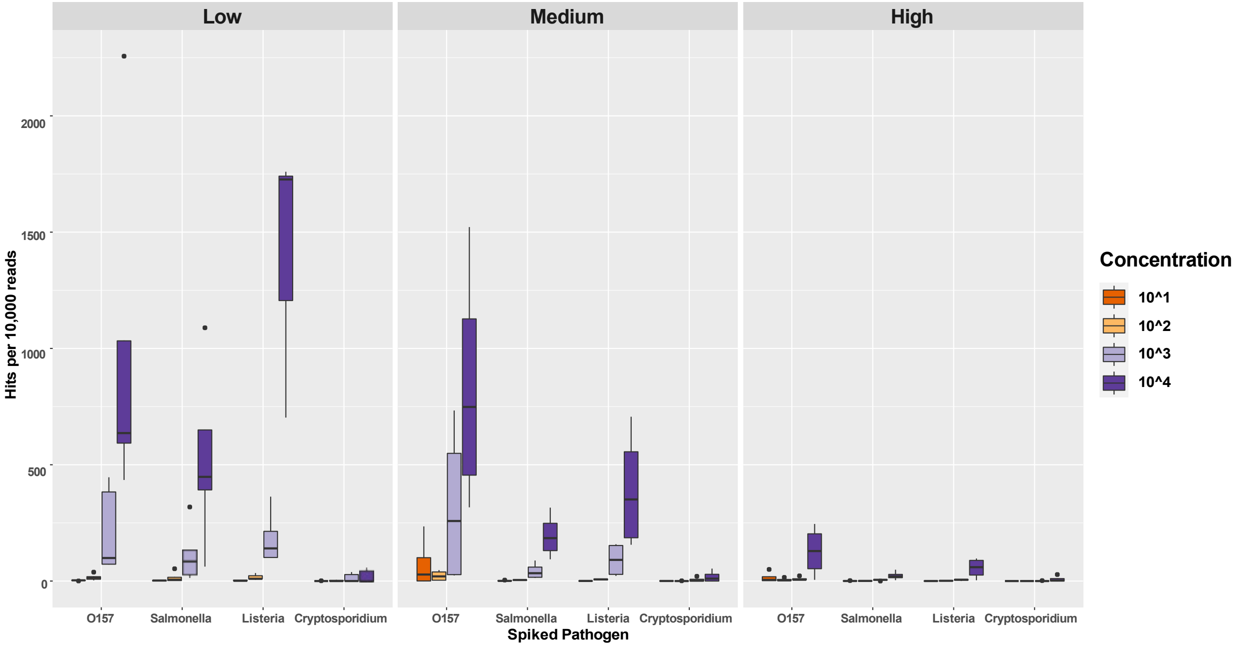

Figure 8. Effect of irrigation water turbidity on the ability to detect various spiked foodborne pathogens in samples using Oxford Nanopore sequencing. Low turbidity irrigation water (<50 nephelometric turbidity units (NTU)), medium turbidity irrigation water (51–200 NTU), and high turbidity irrigation water (>200 NTU). Irrigation water turbidity at levels low or medium does not interfere with pathogen detection using Oxford Nanopore sequencing from spiked irrigation water with multiple pathogens present, but high turbidity levels decrease the detection limit by two logs or higher to $>10^4$. Each plot represents number of pathogen genome hits per 10,000 Oxford Nanopore sequencing reads. Each sample had all four pathogens at the same spiked concentration.

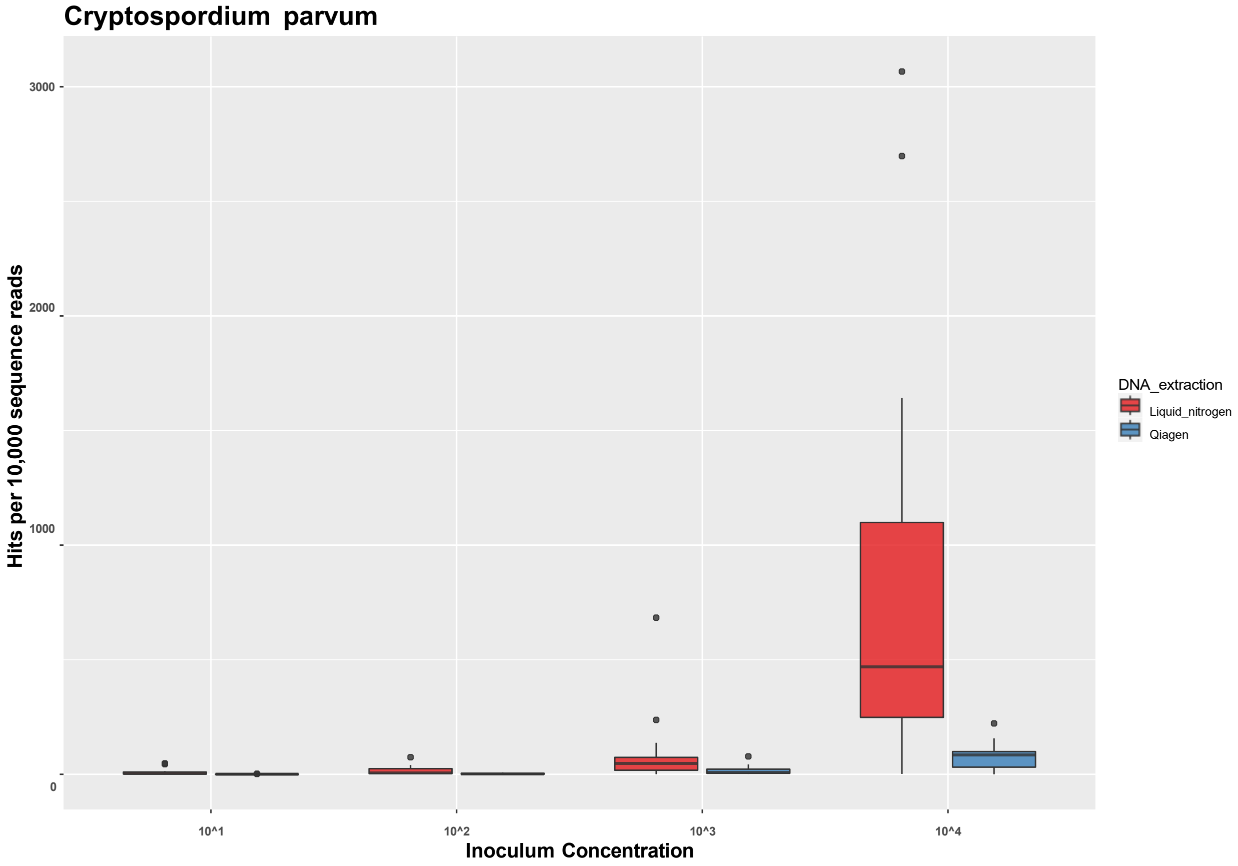

Figure 9. Different DNA extraction processes on *C. parvum* oocysts. Different processes resulted in difference in ability to detect the parasite *C. parvum* from spiked irrigation water samples. Utilizing commercial Qiagen PowerSoil or PowerWater kits resulted in about 10x lower Oxford Nanopore sequencing hits to the pathogen's genome compared to breaking oocysts open with liquid nitrogen freeze/thaw cycling. Each plot represents number of pathogen genome hits per 10,000 Oxford Nanopore sequencing reads.

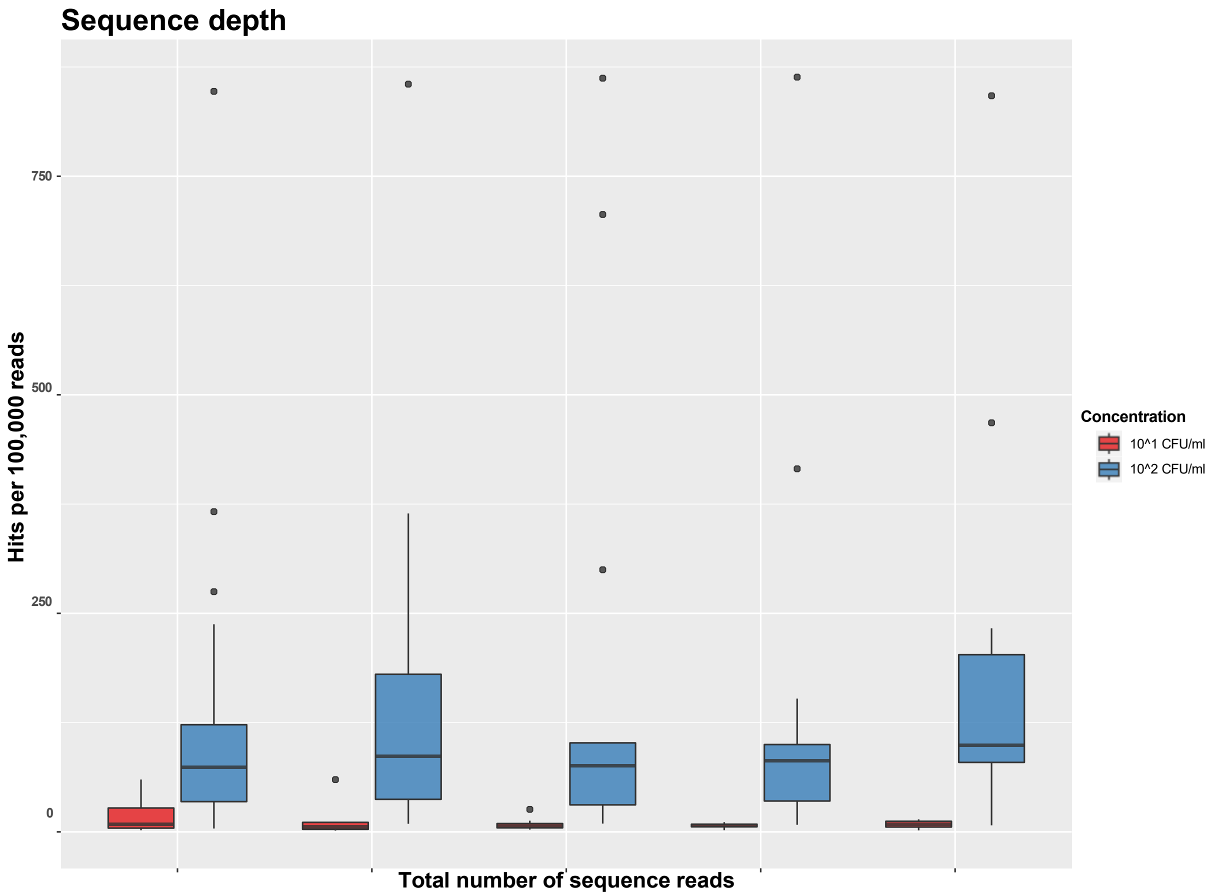

Figure 10. Effects of overall sequencing depth (number of sequence reads generated from a sample) on the ability to detect *Listeria monocytogenes* from spike irrigation water samples at various concentrations. Each plot represents number of pathogen genome hits per 100,000 Oxford Nanopore sequencing reads. Sequencing depth does not alter the standardized number of hits per 10,000 reads, but the increase in overall read numbers will result in an increase in total reads mapping to the pathogen genome.

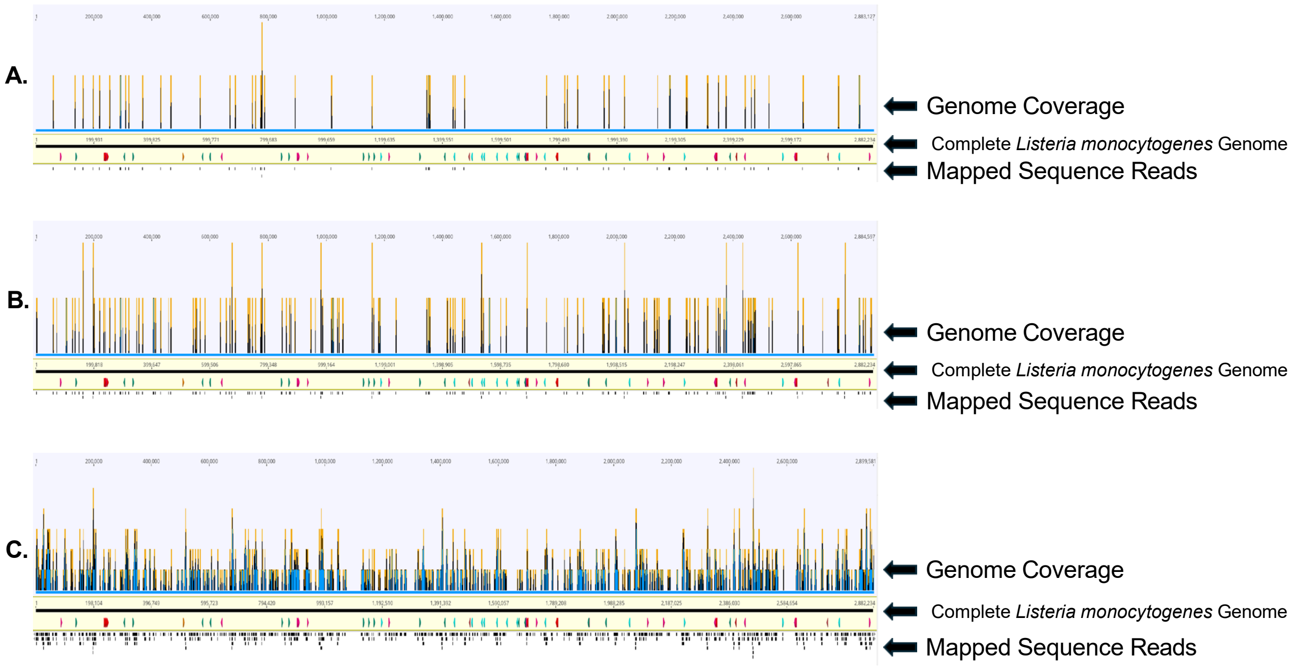

Figure 11. Genomic maps of sequence read hits at varying sequencing depths to *Listeria monocytogenes* at 10^2 CFU/ml spiked inoculum concentration, which demonstrates the more sequencing depth generated for a particular sample, the more genome hits for the pathogens you will get detected. A. 200 megabases (Mb) or approximately 100,000 sequence reads; B. 600 megabases (Mb) or approximately 300,000 sequence reads; C. 1,000 megabases (Mb) or approximately 500,000 sequence reads.

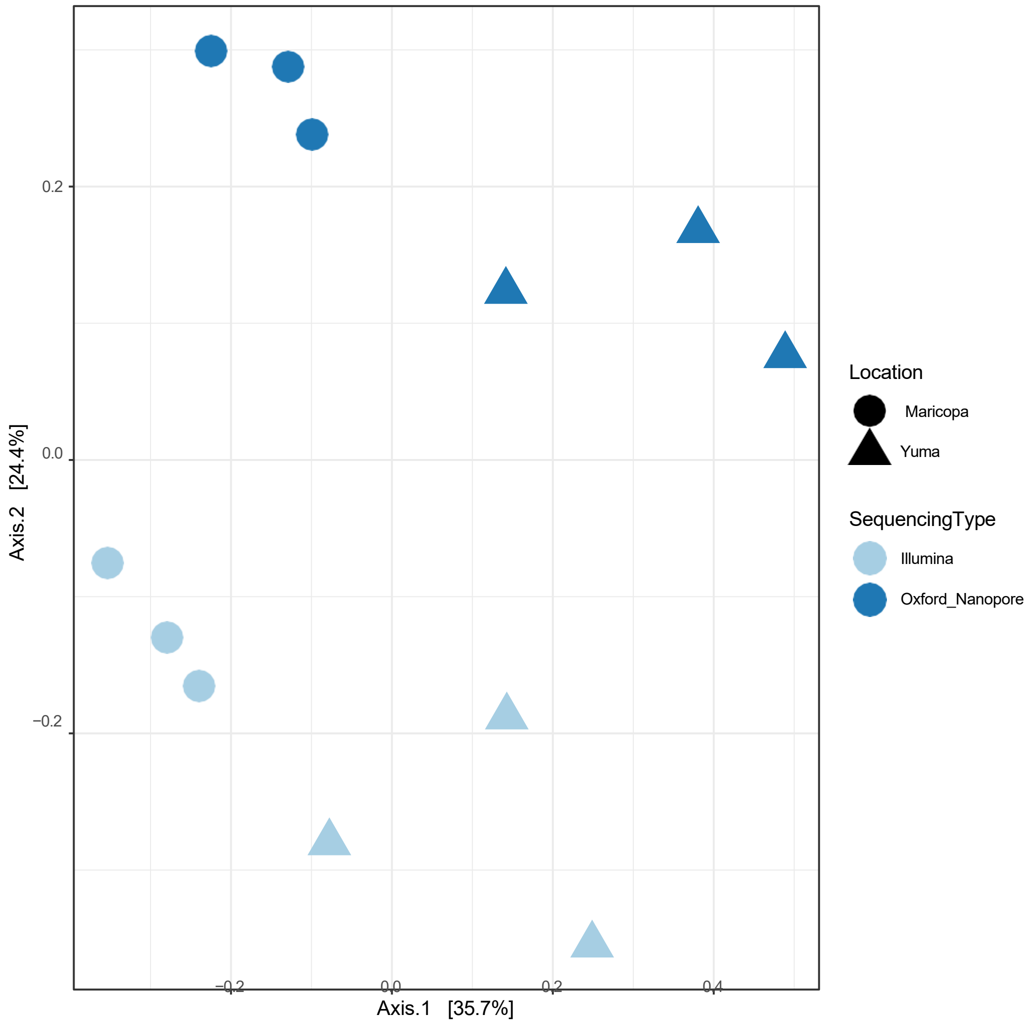

Figure 12. Ordination plot that demonstrates that different sequencing technologies can generate different microbiome results from irrigation water samples. The closer two symbols are located to each other the closer their microbial community's composition and abundance are to each other, therefore those clustered close together are very similar. Circles represent location of the sample, whereas the color of the symbol represents the sequencing technology. Each sample was divided into two equal parts and sequenced with both technologies, then for analysis each sample was randomly subsampled to 100,000 total sequencing reads to eliminate sequencing depth bias. Samples clustered more by sequencing technology type compared to location. Analysis represents only the six winter irrigation water samples.

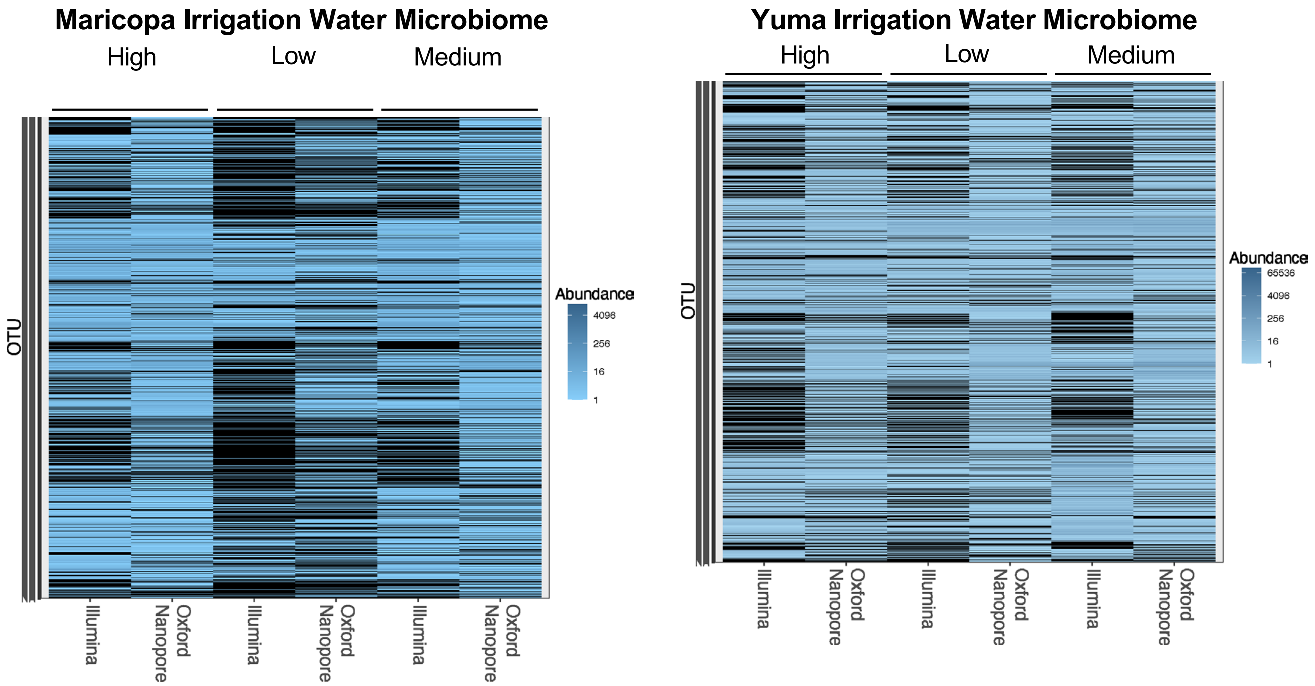

Figure 13. Heat map showing the abundance of different microbial species present in different turbidity levels of irrigation water samples collected from either Maricopa, AZ or Yuma, AZ. Each sample was divided into two equal parts and sequenced with both technologies, then for analysis each sample was randomly subsampled to 100,000 total sequencing reads to eliminate sequencing depth bias. Each blue mark represents the abundance of that particular microbial species in that sample, the darker the blue the more abundant and black means it is virtually absent from that sample. Turbidity differences for the samples is labeled along the top of the figure. Variation in abundance of different microbial species from same sample sequenced with different technologies indicates variation in the ability for different sequencing technologies to correctly identify the microbiome of environmental samples like irrigation water. Analysis represents only the six winter irrigation water samples.

Table 2. American Type Culture Collection (ATCC) 20 bacterial strain positive control.

Bacterial Strains
Acinetobacter baumannii
Schaalia odontolytica
Bacillus pacificus
Phocaeicola vulgatus
Bifidobacterium adolescentis
Clostridium beijerinckii
Cutibacterium acnes
Deinococcus radiodurans
Enterococcus faecalis
Escherichia coli
Helicobacter pylori
Lactobacillus gasseri
Neisseria meningitidis
Porphromonas gingivalis
Pseudomonas paraeruginosa
Cereibacter sphaeroides
Staphylococcus aureus
Staphylococcus epidermidis
Streptococcus agalactiae
Streptococcus mutans

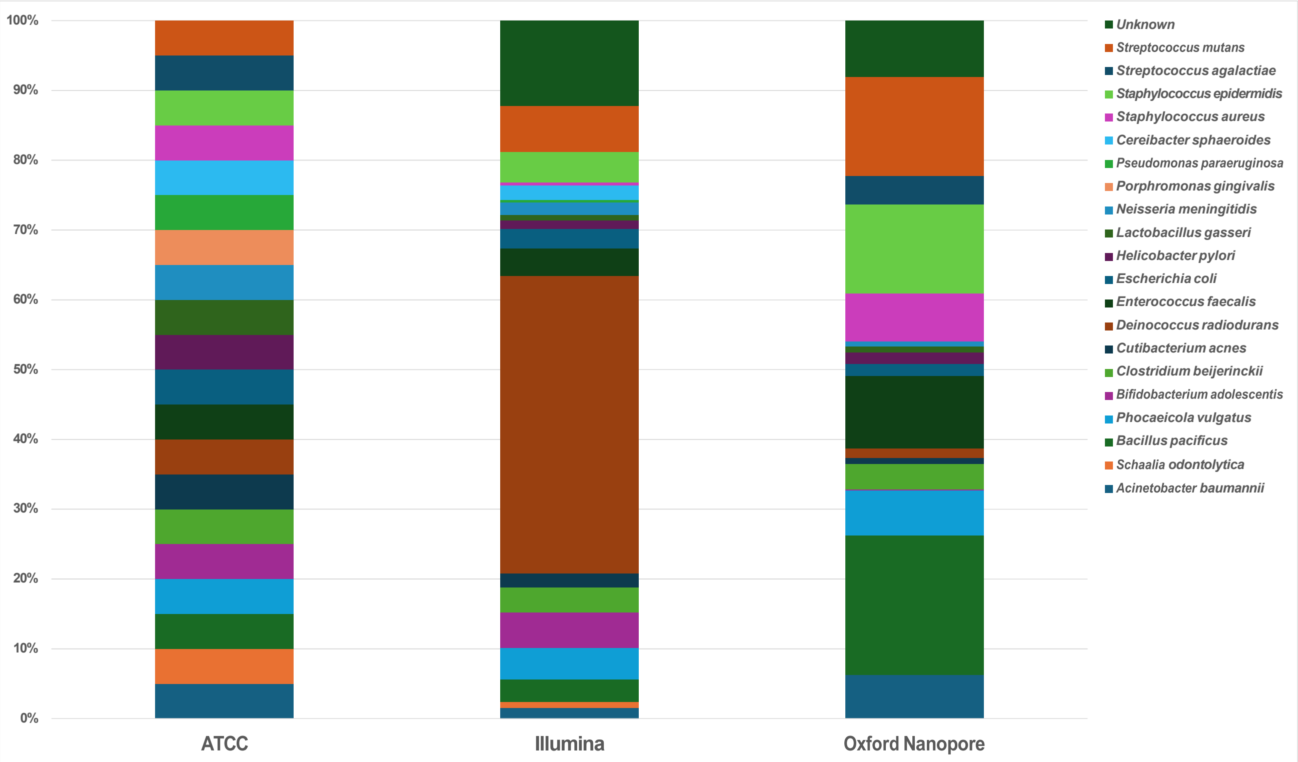

Figure 14. Comparison of the two sequencing technologies (Illumina NovaSeq and Oxford Nanopore MinION) to correctly identify the defined 20 bacterial strain positive control from the American Type Culture Collection (ATCC), which has equal amounts of the 20 strains in the sample or 5% of each bacterial strain. Legend indicates the bacteria species in the order of the stacked plot, and unknown represents the percentage of sequencing reads generated by each technology that did not match one of the 20 strains from the control.

Table 3. “How do you genome” Workshop Information.

	Location	**Date**	**Attendance**
Workshop #1	Yuma, AZ	December 12, 2022	13
Workshop #2	Tucson, AZ	October 27, 2023	5
Workshop #3	Yuma, AZ	November 14, 2023	25
Workshop #4	Yuma, AZ	December 13, 2023	8
Workshop #5	Pre-recorded	Available upon request	N/A
Workshop #6	Pre-recorded	Available upon request	N/A