Entry to the Stockholm Junior Water Prize 2024:

Is the water safe to drink? The rapid test is the missing link!

Noah Bryan

Bayview Secondary School, Richmond Hill, Ontario, Canada

2. Preliminary Matters

2a. Abstract

Globally, contaminated drinking water causes 2.2 billion illnesses and 1 million deaths annually. Lack of access to clean water also leads to chronic health issues in children, such as stunting and malnutrition. Current water tests are time-consuming, require laboratory settings, and focus on detecting fecal indicator bacteria like *Escherichia coli*, which does not reliably indicate water safety. To address these gaps, I developed a field-based water testing approach using three portable devices: a water concentrator, a Bento Lab, and an Oxford Nanopore MinION handheld sequencer. In this method, ten liters of water are processed in the concentrator, where bacteria bind to paramagnetic affinity capture beads. After isolating the bead/bacteria complexes, bacterial DNA is extracted, amplified via polymerase chain reaction, and sequenced in real-time to identify all present bacteria, including pathogens. The test, performed at directly the water source, takes nine hours. Analysis of two Ontario lakes (Lake Wilcox and Guelph Lake) confirmed the presence of multiple pathogenic bacterial species, with two pathogens verified by culture-based methods. This innovative test offers a potentially transformative solution to the global water contamination challenge.

2b. Table of Contents

2. Preliminary Matters ……1
2a. Abstract ……1
2b. Table of Contents ……2
2c. Key Words ……3
2d. Abbreviations and Acronyms ……...3
2e. Acknowledgements ……3
2f. Biography ……3
3. Introduction ……4
4. Methods and Materials ……6
5. Results ……...10
6. Discussion ……...13
7. Conclusions ……...16
8. References ……...17

2c. Key Words

Drinking water contamination, waterborne illnesses and deaths, inadequate testing methods, *Escherichia coli*, rapid, field-based testing, bacterial concentration, polymerase chain reaction, real time DNA sequencing

2d. Abbreviations and Acronyms

E. coli (*Escherichia coli*), 16s rRNA gene (16s ribosomal ribonucleic acid gene), PCR (Polymerase Chain Reaction), ng (nanogram), g (grams), µg (microgram), µl (microliter), SM (saline-magnesium), TSA (tryptic soy agar), mL (millilitre), kbp (kilobase pair)

2e. Acknowledgements

I would like to express my deepest appreciation to Dr. Lawrence Goodridge and his laboratory staff, who provided the necessary guidance and laboratory supplies required for the completion of my project. Specifically, I would like to thank Dr. Valeria Pinto, and Dr. Opeyemi Lawal, who taught me how to do Oxford Nanopore sequencing and the downstream analysis of the sequencing data, and provided quality assurance and quality control of the sequencing data to ensure there were no errors in analysis. I would also like to thank Dr. Goodridge for his unbending belief in me, mentorship (I have spent the past two years completing science fair projects in his lab, which have led to several publications and conference abstracts at international science conferences), and for treating me as a regular member of his lab group, even though I am still in high school.

2f. Biography

Noah Bryan is a grade 11 IB student from Canada who is committed to bettering the lives of marginalized communities through prevention of adverse health outcomes. Inspired by his grandfather's stroke caused by high blood pressure, Noah founded the Make a Move Foundation launching a Hypertension 101 initiative to fight this "Silent Killer", providing free Home Blood Pressure Kits to those that can't afford them through a partnership with a medical supply company. He has been featured on CBC Kids and was a guest on CBC Radio's The Morning Show for his work in this area. At his first STEM competition in grade 8, he noticed a severe lack of minority representation which inspired him to join the Canadian Black Scientist Network. When he noticed "no swimming signs" at his community lake, he launched a science project to investigate the lake's safety for recreation by detecting bacteriophages to evaluate water quality.

These efforts won him the Platinum Award for best discovery project at the 2023 CWSF (Canada Wide Science Fair) and two first author publications. Noah was the youngest ever presenter of his abstract related to water quality at the 2023 International Association for Food Protection (IAFP) conference. Most recently, Noah answered the UNICEF challenge to develop a rapid water quality test to determine bacterial pathogens tackling a leading cause of mortality in African children under 5. For this effort he won the Sanofi Biogenius Canada Award, The Challenge Award, The Youth Can Innovate Top Prize and the Stockholm Junior Water Prize at CWSF 2024. Noah is a published poet, Grade 8 Royal Conservatory classically trained pianist and now composes his own music. He is a past nationally ranked tennis player and certified tennis coach. He is also a back-to-back National Speaker's Idol Finalist, speaking against hate and human rights violations. Noah aspires to become a medical doctor with a research interest in infectious diseases and genomics.

3. Introduction

Access to safe drinking water is a fundamental human right and a critical component of public health. However, a staggering 2.2 billion people worldwide [1] primarily in Central and Southern Asia and Sub-Saharan Africa [2], still lack access to this essential resource. The consequences of this global crisis are dire. Each year, approximately 1 million people succumb to diseases caused by contaminated water sources [3], including 255,500 children under the age of five [1]. The impacts extend beyond acute illnesses and deaths; chronic health complications such as stunting and malnutrition are prevalent, and many children miss school due to waterborne diseases, significantly hindering their prospects [2].

The importance of clean water cannot be overstated. Safe drinking water is vital for maintaining hygiene, preventing the spread of infectious diseases, and ensuring the overall well-being of communities. Yet, despite significant efforts and investments in water infrastructure, the challenge of ensuring safe drinking water for all remains daunting. Rapid urbanization, population growth, and climate change exacerbate the situation, placing additional pressure on already scarce water resources and infrastructure [2].

Building and maintaining drinking water infrastructure in developing countries face significant challenges, including limited financial resources, inadequate technical expertise, and often, political and social instability [4]. These factors can lead to poorly constructed systems, frequent breakdowns, and lack of consistent maintenance, resulting in unreliable access to clean water. Additionally, the rural and remote

locations of many communities complicate infrastructure development and maintenance efforts. In this context, the development of effective and affordable tests to assess drinking water quality becomes vital. These tests provide critical information about water safety, helping to identify contamination quickly and prevent waterborne diseases. They empower local communities to monitor their water sources, ensuring prompt intervention when problems arise, thereby safeguarding public health and promoting sustainable development.

Current drinking water testing methods have several significant limitations that impede progress [5,6]. These tests often require at least 24 hours to produce results and depend on laboratory equipment, making them unsuitable for remote or resource-limited settings. Furthermore, these tests typically focus on detecting fecal indicator bacteria, such as non-pathogenic *Escherichia coli* [6].

Non-pathogenic (generic) *E. coli* is often chosen as an indicator of drinking water quality because it can be a reliable marker for the presence of fecal contamination. *E. coli* is commonly found in the intestines of humans and warm-blooded animals, making its presence in water a strong indication of potential contamination by pathogens that can cause waterborne diseases [7]. Additionally, *E. coli* is relatively easy to detect and quantify using standard microbiological methods, allowing for 24-48 h, cost-effective water quality assessments [7]. Its presence in drinking water serves as a warning sign for possible health risks, prompting necessary interventions to ensure water safety.

While *E. coli* detection is useful, it is not a comprehensive measure of water safety, and a growing body of scientific literature indicates that *E. coli* does not account for the full spectrum of pathogens, including bacteria, viruses and parasites, that can contaminate water sources [8]. For example, outbreaks due to contaminated drinking water have continued to occur, even in developed countries where robust water quality monitoring approaches using *E. coli* are routinely employed [9]. This means that a negative *E. coli* water test should not be used to indicate that there are no pathogenic bacteria or viruses and parasites that are present in the water. The diversity of viruses and parasites, coupled with the difficulty in detecting these pathogens because they cannot be cultured and are often present in water below detectable levels, means that there are no available water tests for these pathogens.

Thus, the narrow focus on indicator bacteria, and the other inadequacies of current water testing methods hinder efforts to monitor and improve water quality. These limitations affect the ability to accurately understand and communicate risks, initiate community-led water quality monitoring programs, and integrate water testing into large-scale national household surveys. These challenges are most pronounced in regions that urgently need these interventions to safeguard public health. Collectively, current water testing limitations represent a major gap in public health that needs to be addressed.

Recognizing these critical issues, the United Nations International Children's Emergency Fund (UNICEF) has launched a global innovation project challenge that seeks to develop an easy-to-use, rapid detection method or portable kit capable of accurately identifying fecal contamination in drinking water within 10 hours [1]. A major limitation of the challenge is that it still focuses on detection of *E. coli*, which as stated is not a reliable indicator of the presence and diversity of pathogens that may be present in the water.

Therefore, the main objective of this study was to create a portable, field-ready test that can deliver results within 10 hours, addressing some of the UNICEF Innovation project objectives. Additionally, this new approach aimed to directly detect bacterial pathogens rather than relying solely on fecal indicator bacteria, which is a main gap in current water testing approaches. By overcoming the limitations of current testing methods, this innovation holds the promise of significantly enhancing the ability to ensure safe drinking water, particularly in underserved regions.

4. Materials and Methods

Portable lab setup and overview of the testing approach. The field-based portable testing lab for microbial water quality analysis (Figure 1) is designed to be both efficient and comprehensive. It consists of three main components: a 10-liter water sampling and bacterial concentration device, the Bento Lab portable laboratory, and an Oxford Nanopore MinION sequencer attached to a laptop. The process begins with the 10-liter water sampling device, which collects water and concentrates any bacteria present by binding them to paramagnetic affinity capture beads. This concentrated sample is then processed in the Bento Lab (https://bento.bio/), a compact and portable laboratory setup that enables on-site DNA extraction, polymerase chain reaction (PCR) amplification of the bacterial 16S rRNA gene. Finally, the amplified DNA is analyzed using the Oxford Nanopore MinION sequencer, which, when connected to a laptop, provides real-time sequencing data, data processing and bioinformatic analysis to identify all bacterial

species present in the sample, including potential pathogens. This portable lab setup allows for rapid, on-site microbial water quality analysis, delivering results within nine hours.

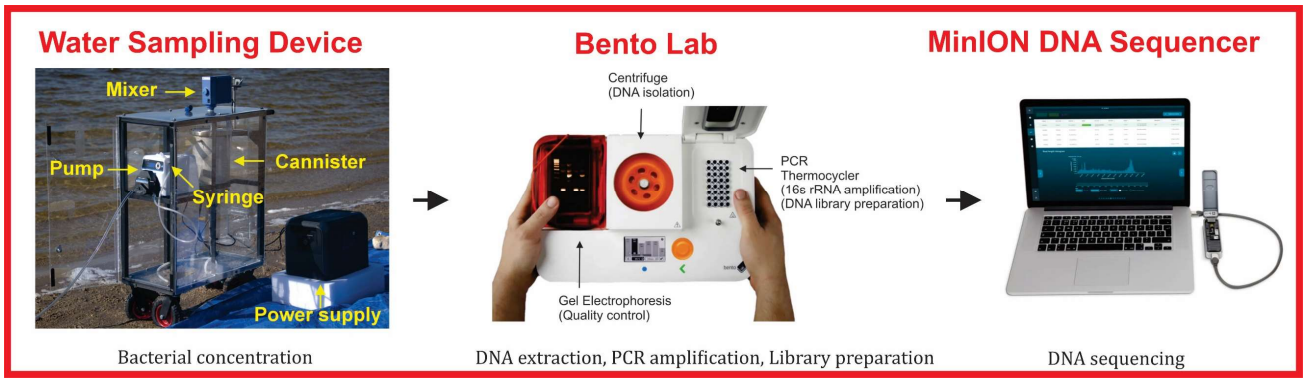

Figure1. The portable lab set up (https://bento.bio/story/rapid-water-testing-with-portable-technology/).

Evaluation of bacterial concentration methods. For all experiments, water was sampled from Lake Wilcox and Guelph Lake between June 2023 and April 2024. To determine the most effective method for concentrating bacteria from 10 liters of water, I designed an experiment comparing three different approaches: polyethylene glycol 6000 (PEG 6000) precipitation, anion exchange resin beads, and Nanotrap Microbiome B Particles. Each method was evaluated based on its efficiency in recovering bacterial cells from lake water samples. For the PEG 6000 method, I collected 10 liters of water using the water sampling device and adjusted the water samples to a final concentration of 10% weight/volume PEG 6000 (Millipore Sigma, Oakville, Ontario, Canada), followed by incubation at room temperature for 4 hours, and collection of the PEG/bacteria complexes in a 5-micrometer sock filter. The anion exchange resin beads were utilized by collecting 10-liter water samples in the sampling device, adding 0.5 gram of Amberlite™ IRA-900 chloride form resin beads (Millipore Sigma, Oakville, Ontario, Canada), followed by stirring the water for 30 minutes, and collection of the resin beads in a 5-micrometer sock filter. The Nanotrap Microbiome B Particles (1.2 mL) (Ceres Nanosciences, Manassas, VA, USA) were added directly to the water samples, and the sample was stirred for 30 minutes, followed by collection of the resin beads in a 5-micrometer sock filter. Each experiment was conducted twice. Total DNA was isolated from the individual PEG 6000, Amberlite™ IRA-900 chloride form resin beads and Nanotrap Microbiome B Particle samples, using the Qiagen DNA Blood and Tissue Kit (Qiagen, Hilden, Germany) followed by PCR of the 16s rRNA gene, and DNA sequencing to evaluate the bacterial concentration efficiency of the respective methods.

PCR amplification of the 16s rRNA gene. 16S barcoding PCR was performed using the Oxford Nanopore kit SQK-16S024 (Oxford Nanopore Technologies, UK). The primers provided in the kit include the unique barcode sequence attached to the universal 16S forward and reverse primers [10]. The entire 16S rRNA gene PCR amplification was performed following Oxford Nanopore instructions. Briefly, 10 ng genomic DNA was subjected to PCR amplification with the LongAmp Hot Start PCR Master Mix (M0533S, New England Biolabs, USA). The amplification conditions were as follows: initial denaturation at 95°C for 1min; 25 cycles of 95°C for 20s, 55°C for 30s, and 65°C for 2min; final extension at 65°C for 5 min. The Bento Lab contains 32 PCR wells, which provided ample space to run multiple samples during development of the assay, including positive and negative controls (I used amplification of the 16s rRNA gene from *E. coli* as a positive control, and sterile water as a negative control). The amplicons (1.5 Kb size) were analyzed by electrophoresis on the Bento Lab using 1.5% (w/v) agarose gel and visualization was by EZ-Vision DNA dye staining (VWR, ON, Canada).

DNA Library Preparation and Sequencing of the 16s rRNA gene. The barcoded amplicons were purified using the AMPure XP beads (Beckman Coulter, ON, Canada) as per Oxford Nanopore's instructions. Samples were then quantified using the Quantus Fluorometer. Final Library preparation for the sequencing was performed using the recommended protocol for ONT kit SQK-16S024. The library was sequenced using the Flongle flow cell R9.4.1 (FLO-FLG001, Oxford Nanopore Technologies) on a MinION Mk1B nanopore sequencer. Sequencing continued until the number of output reads reached a plateau (approximately 4 hours).

Statistical Analysis of DNA Sequence Data. Following DNA sequencing, sequence reads were analyzed using statistical approaches via the FastQC pipeline (https://github.com/s-andrews/FastQC) to determine the quality of the sequencing reactions. The sequence length of the reads obtained were assessed with the read count to determine the overall quality of the sequences. In addition, the per sequence quality scores were assessed to obtain a summary of the overall quality of the reads in the sequence data. Finally, a quality assessment of the data based on the GC content was conducted to measure the quality and level of contamination in the DNA sequences.

Bioinformatic Analysis of DNA Sequence Data. MINKNOW software v.1.11.5 (Oxford Nanopore Technologies) was used for data acquisition. The two ends of the raw reads obtained were trimmed and

filtered using a quality score threshold of 10 and reads shorter than 100 bp were removed. Thereafter, the high-quality reads were mapped to a host sequence database to remove host DNA. For taxonomic classification and assignment, the non-host reads were mapped to the comprehensive and up-to-date Kraken database using kraken2 (GitHub - DerrickWood/kraken2: The second version of the Kraken taxonomic sequence classification system). The taxonomic profiles of the samples which indicated which bacterial species were present, were visualized using the Krona pipeline (q2-krona: Plugin for creating Krona plots - Community Contributions / Plugins - QIIME 2 Forum).

Validation of water test results. To validate the results of the water test, I used bacterial cultural approaches followed by whole genome DNA sequencing of isolated bacterial colonies to confirm the presence of two bacterial pathogens that were consistently identified in the water samples including *Bacillus anthracis* and *Vibrio cholerae*. Water samples (1 L) were collected from shallow regions of Lake Wilcox using a sterile bottle and kept at 4°C until processing. Water aliquots (1 mL) were serially diluted in 9 mL of modified saline-magnesium (SM) buffer and plated onto tryptic soy agar (TSA). The plates were incubated for 24 h at 37°C. Bacterial colonies of differing morphologies were subcultured on TSA. Genomic DNA was extracted from a pure colony using the DNeasy blood and tissue kit (Qiagen, Hilden, Germany) and sequenced using the Illumina DNA library prep tagmentation kit and IDT for Illumina DNA/RNA unique dual (UD) indexes as previously described [11].

Sensitivity of the Water Test. My data showed that the lowest concentration of DNA that was detectable in the water test was 1 ng (1×10^{-9} g) (Figure 2). To calculate the lowest number of bacterial cells detected in the water test, I used a formula that consisted of multiplying the average weight of a base pair of DNA by the number of base pairs in the 16 s rRNA gene (1500 bp) to obtain the average weight of the 16s rRNA gene. Next, the weight (in g) was divided by 1 ng (1×10^{-9} g) to obtain the number of 16s rRNA gene copies detected by the assay. This number was then divided by the average number of 16s rRNA genes in a bacterial cell, (estimated to be 10) to calculate the total number of 16s rRNA genes

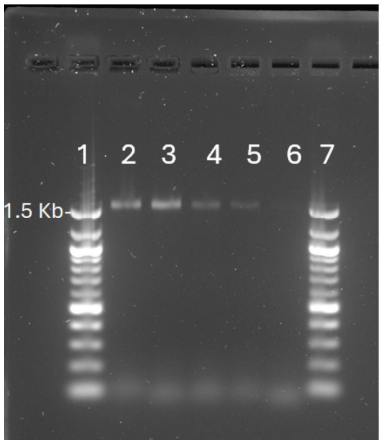

Figure 2. Agarose gel (1.5%) showing the sensitivity of the 16s rRNA water test. 1 ng of nucleic acid which corresponds to a detection limit of 6.17×10^3 bacterial cells/mL, was detectable.

detected per cell. Finally, since I concentrated 10 L (10,000 mL) of water prior to starting the test, the total number of 16s rRNA genes detected per cell was divided by 10,000 to determine the number of bacterial cells/mL that were detectable.

5. Results

An overview of my project is presented on the Bento Lab website (https://bento.bio/story/rapid-water-testing-with-portable-technology/).

Evaluation of bacterial concentration methods. In order to choose the best and most practical method for use in concentrating bacteria from large volumes of water, we evaluated three (PEG 6000, anion exchange beads, and Nanotrap Microbiome B Particles). for their ability to efficiently capture total bacteria from 10 litres of water, as determined by DNA quantification, 16s rRNA gene amplification and sequencing efficiency. Table 1 shows the results of the bacterial capture experiments. DNA was obtained from two of the approaches (PEG 6000, Nanotrap Microbiome B Particles). However, PCR amplification and DNA sequencing was only successful with the Nanotrap Microbiome B Particle samples (Figures 3, 4 a, b).

Table 1. Results (DNA concentration, 16s rRNA PCR and DNA sequencing) from the bacterial concentration experiment.

Concentration Method	Average recovered DNA Concentration (2 replicates)	16s rRNA PCR	16s rRNA sequencing
Nanotrap Microbiome B Particles	21.5 ng/µg	Yes	Yes
PEG 6000	13.0 ng/µg	No	N/A
Amberlite Resin Beads	0.0 ng/µg	N/A	N/A

PCR amplification of the 16s rRNA gene. PCR amplicons of the same size as the 16s rRNA gene were consistently amplified from all Lake Wilcox and Guelph Lake samples following concentration using the Nanotrap Microbiome B Particles (Figure 4b). The amount of amplified DNA varied depending on the concentrations of bacteria in the water samples.

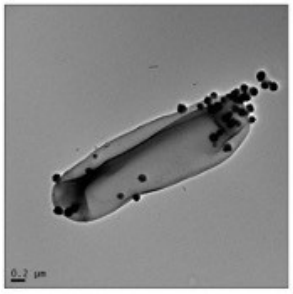

Figure 3. Transmission Electron Micrograph of Nanotrap beads attached to a bacterial cell.

Statistical Analysis of DNA Sequence Data. The sequence length of the reads, the per sequence quality scores and the GC content statistical analysis are shown in Figure 5. Overall, the sequence data from both lakes had an average minimum sequence length of ~1500 bp which implies that the sequence datasets were of high-quality. The difference between the read counts relative to the sequence length distribution are not statistically

significant ($p > 0.05$) (Figure 5a). Similarly, the mean sequence quality scores indicated that the sequence datasets from both lakes were of high quality (Figure 5b). The sharp monomodal peak of samples from both lakes in the GC content data (Figure 5C) are an indication of high-quality reads devoid of contamination.

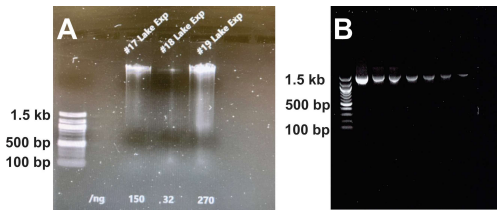

Figure 4. Agarose gels showing the results of the bacterial concentration experiments. **A.** Agarose gel of isolated DNA from 3 Lake Wilcox samples following Nanotrap bead concentration. The DNA ladder is in lane 1. The DNA concentration in ng is shown beneath each lane (lanes 3-5). **B.** Agarose gel showing 16s rRNA amplicons from different samples of water from Lake Wilcox (lanes 3-8). The DNA ladder is in the lane 1. The positive control (*E. coli*) is in lane 2.

Bioinformatic Analysis of DNA Sequence Data. PCR amplification of the 16s rRNA gene and DNA sequencing led to identification of 544 bacterial species in the Lake Wilcox samples and 635 bacterial species in the Guelph Lake samples (Figure 6a, b). Although the bacterial species identified in the two lakes were similar, the relative abundances differed greatly. For Guelph Lake, out of the 635 bacterial species found, only 27 had ≥0.5% relative abundance. *Guyparkeria halophila* had the highest abundance (11.29%). *Salmonella enterica* was found alongside *Bacillus cereus*, *Staphylococcus aureus*, and *Clostridioides difficile*. These pathogen bacteria are all known to be disseminated by food and water. For Lake Wilcox, a higher number of bacterial species (n=39) had ≥ 0.5% relative abundance, with *Comamomas* sp. having the highest abundance (6.5%) in the sample. Similar to Guelph, *Salmonella enterica*, *Bacillus cereus*, and *Clostridioides difficile* were among the top 20 bacterial species identified.

Figure 5. Statistical analysis of the 16s rRNA DNA sequence reads from the water quality test. **A.** Length distribution of sequence reads obtained from the lake. The sequence length of the reads obtained were assessed with the read count to determine the overall quality of the sequences. The difference between the read counts relative to the sequence length distribution are not statistically significant ($p > 0.05$). **B.** The Per Sequence Quality Scores is the summary of the overall quality of the reads in the sequence data. It shows the distribution of mean quality scores across all sequences obtained from sequencing the full length 16S rRNA gene from the two lakes. The curve on the plot shows the number of reads that fall into each quality score. A normal distribution (bell curve) suggests high-quality datasets. Although the differences between the quality scores relative to the number of reads between the two lakes was statistically significant ($p < 0.05$), the sequence datasets from both lakes were of high quality. **C.** The GC content analysis of the sequence datasets measures the quality and level of contamination in the sequences. The sharp monomodal peak of samples from both lakes is an indication of high-quality reads devoid of contamination.

Validation of the Water Test results. I used standardized bacterial cultural approaches to determine if I could isolate any bacterial species that were detected by the water test, as confirmation that the test was

accurately detecting what was present in the lake samples. Following dilutions of the Lake Wilcox samples, and spread plating on TSA, I randomly selected two bacterial colonies from different water samples (obtained in different seasons) and had them sequenced in Dr. Goodridge's laboratory. The results

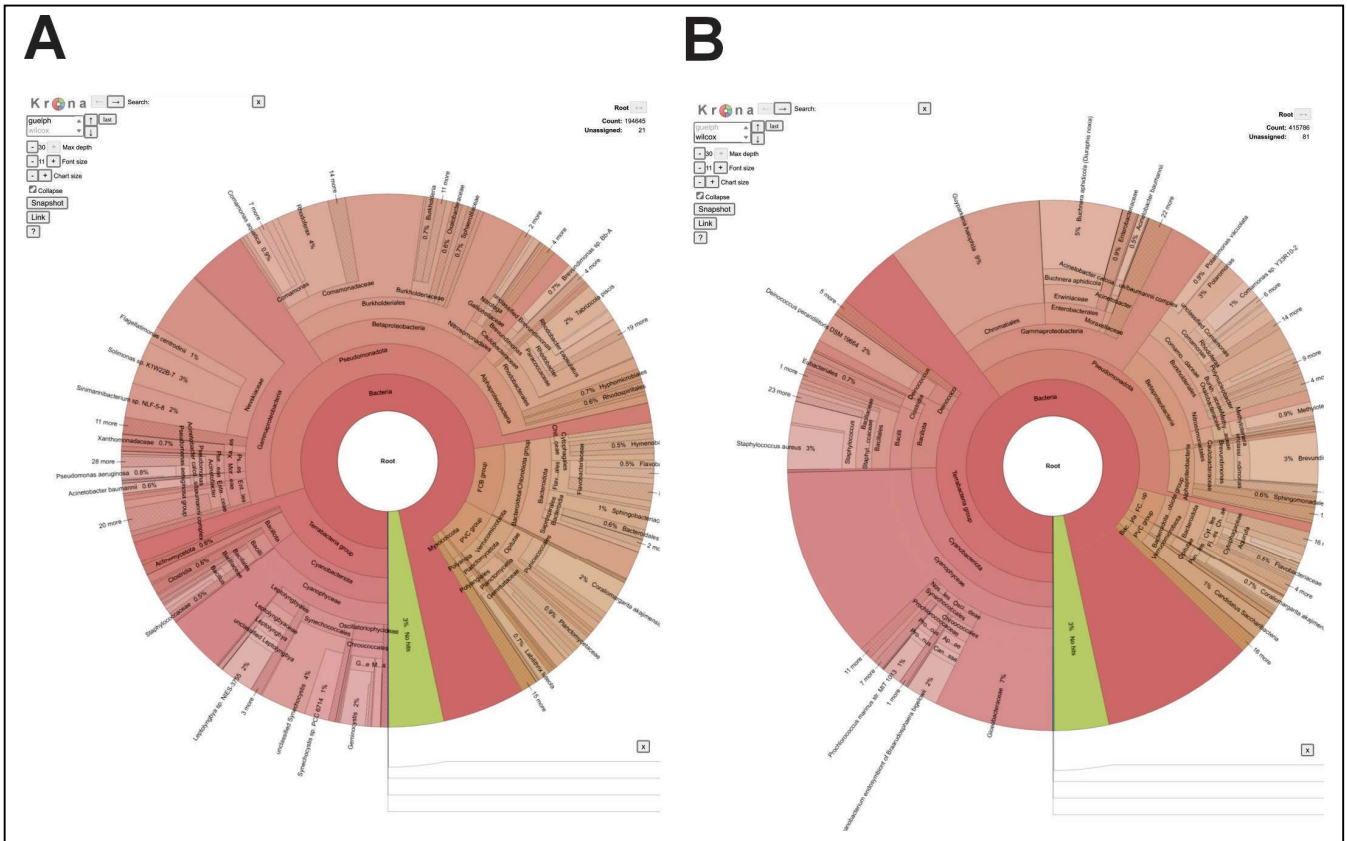

Figure 6. Bacterial diversity as determined by bioinformatic analysis of 16s rRNA gene sequences following the water test. **A.** Bacterial diversity in Guelph Lake. **B.** Bacterial diversity in Guelph Lake. Using a visual program called Krona, the data can be searched to identify bacterial pathogens from a web browser. To search the data from this Figure, go to https://www.dropbox.com/scl/fo/2vfolsjrtkt5kqahxe6s5/AIabwMwy5VNZ5h81OoLSo2A?rlkey=h7ey5891dlrfynsmxo77gtz46&st=tgpoi5ix&dl=0 and download the file named Guelph-Wilcox.html. Then enter the names of bacterial species such as *Escherichia coli*, *Salmonella enterica*, *Vibrio cholerae* in the search box and press enter.

indicate the presence of two bacterial pathogens (*Bacillus anthracis*, *Vibrio cholerae*) that were consistently found in the water samples. I published two journal articles detailing the isolation and DNA sequence analysis of these pathogens [12, 13].

Sensitivity of the water test. To calculate the lowest number of bacterial cells detected in the water test, I first converted the base pairs in the full length 16s rRNA gene to molecular weight. The average molecular weight of a base pair (bp) of double-stranded DNA is approximately 650 Daltons ($1.66053906660 \times 10^{-24}$ grams) [14]. Therefore, the molecular weight of the 16S rRNA gene is:

$$1{,}500 \text{ bp} \times 650 \text{ Daltons/bp} = 975{,}000 \text{ Daltons}$$

Next, I converted Daltons to grams. Since 1 Dalton is approximately $1.66053906660 \times 10^{-24}$ grams, the molecular weight of the 16S rRNA gene in grams is:

$$975{,}000 \text{ Daltons} \times 1.66053906660 \times 10^{-24} \text{ grams/Dalton} \approx 1.619 \times 10^{-18} \text{ grams}$$

Then, I calculated the number of 16S rRNA gene copies in 1 nanogram of DNA. 1 nanogram (ng) is 1×10^{-9} grams. Therefore, the number of 16S rRNA genes in 1 nanogram of DNA is:

$$1 \times 10^{-9} \text{ grams} / 1.619 \times 10^{-18} \text{ grams} \approx 6.17 \times 10^{8} \text{ copies}$$

So, there are approximately 6.17×10^{8} (or 617 million) full-length 16S rRNA gene copies in 1 nanogram of DNA. The next step was to divide this value by the average number of 16s rRNA genes in a bacterial cell, which I estimated to be 10. This provided the number of bacterial cells detectable in 10 L of water. Since the sensitivity of a water test is typically reported per mL or (sometimes per 100 mL), I divided 6.17×10^{7} bacterial cells by 10,000 mL to obtain the sensitivity of the assay in mL, which is:

$$6.17 \times 10^{7} \text{ bacterial cells} / 10{,}000 \text{ mL} = 6.17 \times 10^{3} \text{ bacterial cells/mL}$$

This value agrees favorably with the reported concentrations of bacteria observed in freshwater lakes. For example, Shao *et al*. (2021) reported that bacterial concentrations in a freshwater lake ranged from 10^{3} to 10^{6} bacterial cells per milliliter of water [15].

6. Discussion

Diseases linked to unsafe drinking water and poor sanitation represent a major health challenge globally. These illnesses vary from mild conditions to severe diseases like typhoid fever and cholera, which affects 69 countries [2]. Diarrhea, often caused by contaminated water, can lead to dangerous dehydration and loss of essential salts, posing a significant health risk. This burden is especially severe for children under five, making diarrhea the fourth leading cause of death in this age group worldwide [2].

By 2020, safely managed drinking water services were reported in 138 countries, covering 45% of the global population [2]. Starting in 2012, the Joint Monitoring Programme, in collaboration with UNICEF's Multiple Indicator Cluster Survey program, developed a standardized water quality module. This module has become practical due to the increased availability of affordable, precise water testing methods suitable for household surveys [2]. By early 2022, water quality data from over 50 nationally or sub nationally representative household surveys were available from more than 40 countries [16]. Despite these advancements, current testing procedures often require laboratory analysis, which can be a limitation. Additionally, these tests are based on detection of non-pathogenic *E. coli*, and as described by Nowicki *et al*. (2021), the use of *E. coli* to monitor water quality in developing countries involves questions surrounding the extent that *E. coli* collected at the source is linked to recent fecal contamination, and also questions regarding the health hazard implications of changes in *E. coli* concentration between the source and point of use [17]. These questions are not easily resolved, which explains in my opinion, why accurate and rapid testing methods based on direct pathogen detection need to be developed. Also, the development of field-based tests, like the one I developed in this study, which are designed to be used right at the water source, aims to enable more frequent and consistent testing of drinking water in areas with common contamination issues.

A major limitation of current drinking water tests is the need to enrich or grow the bacteria to a high enough concentration that they are detectable. This means that current tests take at least 24 hours to deliver a result [4, 5]. To develop a rapid field-based test, I had to develop an approach to concentrate the bacteria from the water rapidly, to decrease the time to detection. To develop the concentration method, I chose two methods traditionally used to concentrate bacteria from liquid samples including PEG 6000 and anion exchange resin beads, and a newer approach based on the use of Nanotrap paramagentic beads that have affinity for bacteria. Polyethylene glycol (PEG) 6000 is a high molecular weight polymer that is often used for the concentration of bacteria from water samples through a process called precipitation [18]. The method involves adding PEG 6000 to the water sample, which induces the precipitation of bacteria by reducing the solubility of proteins and other macromolecules. PEG acts by creating an osmotic imbalance, causing water to move out of bacterial cells and promoting the aggregation and precipitation of the cells [18]. This method is advantageous because it is relatively simple and cost-effective, making it suitable for field applications and large volume samples.

Anion exchange resins are another effective tool for concentrating bacteria from water sample [19]. These resins are composed of positively charged groups that attract and bind negatively charged bacterial cell surfaces [19]. By passing water through a column packed with anion exchange resin, or by allowing the resin to circulate freely in a water sample, bacteria are adsorbed onto the resin surface. The bound bacteria can then be eluted with a high salt buffer or another eluent that disrupts the ionic interactions [19].
Nanotrap microbiome beads are an innovative technology designed to capture, concentrate, and preserve microbial organisms and their associated molecules from complex biological samples such as water, blood, or environmental samples. Nanotrap beads are made of hydrogel particles that contain a network of nanopores [20]. These pores are functionalized with specific chemical ligands that can selectively bind to and trap target molecules, including proteins, and whole microorganisms [20]. The beads work by leveraging electrostatic interactions, hydrophobic interactions, to capture a wide range of microbial components.

My results indicated that the Nanotrap beads worked the best at concentrating bacteria from the water samples prior to DNA extraction, PCR amplification and sequencing. Compared to the PEG 6000 samples, higher concentrations of DNA were obtained from the Nanotrap beads. The DNA was also of higher quality, since only DNA from the Nanotrap beads were amplifiable by PCR. Reasons that the PEG 6000 approach may not have worked include the fact that to be efficient, samples to be concentrated using PEG should be refrigerated [20]. In our study, due to the field-testing requirements, refrigerating the samples and incubating for a long time were not feasible. Reasons that the Anion exchange beads may not have worked include the fact that bacterial cells surface charges are highly variable [21], and environmental material present in the lake water may have outcompeted the bacteria for binding sites on the resin beads.

Following concentration of bacteria from the water samples, I detected the presence of bacterial species in the water using the 16s rRNA gene. The 16S ribosomal RNA (rRNA) gene is a component of the 30S small subunit of prokaryotic ribosomes [22]. This gene is approximately 1,500 base pairs long and is highly conserved across different species of bacteria and archaea. It plays a crucial role in the synthesis of proteins by facilitating the binding of mRNA and transfer RNA (tRNA) during translation. Full-length sequencing of the 16S rRNA gene is widely used in bacterial taxonomy because it allows for the precise identification and classification of bacteria [22]. The sequence data from the conserved regions help in aligning the sequences from different bacteria, while the variable regions provide the necessary divergence

to differentiate species. As demonstrated in my study, The 16S rRNA gene sequence can be used to construct phylogenetic trees, which illustrate the evolutionary relationships among different bacterial species. By comparing the sequences of the 16S rRNA gene from different organisms, researchers can infer evolutionary distances and identify new species. In my study, analysis of two lakes, Lake Wilcox and Guelph Lake, led to the identification of 544 and 635 bacterial species respectively. These values compare well with other bacterial taxonomy studies in freshwater lakes using the 16s rRNA gene. For example, Zwart et al. (2002) [23] analyzed 16S rDNA sequences from freshwater plankton in various global locations, including new sequences from Parker River (USA), Lake Soyang (South Korea), and Lake IJssel (Netherlands). In total, 689 bacterial and 75 plastid sequences from North America, Europe, and Asia were examined. Most bacterial sequences were closely related to other freshwater clones, indicating habitat-specific clustering. This suggests these sequences represent species indigenous to freshwater. Thirty-four phylogenetic clusters were identified, with 23 containing no cultivated organisms. These clusters span several bacterial groups, showing that freshwater ecosystems harbor distinct bacterial communities not found in soil or marine environments.

7. Conclusions

I have developed a rapid field-based test that can detect the presence of multiple bacterial species in water within 9 hours. The test has been validated on two lakes previously shown to be impacted by fecal contamination. In addition to the rapid nature of my test, which decreases the detection time over commonly used tests by at least 15 hours, as well as the fact that my test has been designed to be used directly at the water source, a major aspect of my test is the fact that it directly detects pathogens, as opposed to fecal indicator bacteria, which do not reliably indicate the presence of all pathogenic microorganisms that may be present in a water sample. Directly identifying pathogens in drinking water tests is superior to detecting fecal indicators because it provides specific information about the presence of harmful microorganisms, enabling more accurate risk assessment and targeted public health interventions. While fecal indicators like *E. coli* are useful for suggesting potential contamination, they do not identify specific pathogens that cause diseases. This lack of specificity can lead to either overestimation or underestimation of the actual health risk.

By directly identifying pathogens, public health officials can implement more precise and effective measures to prevent waterborne diseases. For instance, knowing the exact pathogens present in water can

guide specific treatment processes, such as the use of disinfectants or filtration methods tailored to remove those pathogens. Additionally, rapid and accurate pathogen identification can enhance outbreak response times, leading to quicker containment and treatment strategies, thereby reducing the incidence of illness and improving overall public health outcomes.

The development of the innovative testing method I describe in this study is crucial for improving water safety globally. It enables more effective monitoring and quicker response times, ultimately reducing the incidence of waterborne diseases and improving the quality of life for millions. The water test utilizes portable technology to conduct rapid and accurate assessments of water quality. This method involves sampling, bacterial concentration, DNA isolation, and sequencing. Given its simplicity and portability, individuals in developing countries with rudimentary scientific knowledge can be trained to perform these tests, ensuring sustainable and community-driven water quality monitoring. Efforts are currently underway to secure funding from entities like the World Bank to purchase the necessary equipment and provide training, aiming to roll out this testing method in areas impacted by poor water quality. As the global community continues to strive towards achieving Sustainable Development Goal 6—ensuring availability and sustainable management of water and sanitation for all [24]—such advancements are indispensable. The successful implementation of these innovations will not only save lives but also contribute to healthier, more resilient communities worldwide.

8. References

1. United Nations International Children's Emergency Fund (UNICEF). 2024. Rapid Water Quality Testing. Available at: https://www.unicef.org/innovation/rapid-water-quality-testing [Accessed 9 June 2024].

2. UNICEF, World Health Organization, and World Bank. 2022. State of the World's Drinking Water. Available at: https://www.unicef.org/media/129381/file/State%20of%20drinking%20water%20report.pdf [Accessed 9 June 2024].

3. World Health Organization. 2023. Drinking-water. Available at: https://www.who.int/news-room/fact-sheets/detail/drinking-water [Accessed 9 June 2024].

4. Onyebuchi Okafor, C., Ibiam Ude, U., Ngozi Okoh, F., et al. 2024. Safe Drinking Water: The Need and Challenges in Developing Countries. Water Quality - New Perspectives. IntechOpen. Available at: http://dx.doi.org/10.5772/intechopen.108497.

5. Ramírez-Castillo, F.Y., Loera-Muro, A., Jacques, M., Garneau, P., Avelar-González, F.J., Harel, J., and Guerrero-Barrera, A.L. 2015. Waterborne pathogens: detection methods and challenges. Pathogens, 4(2), pp.307-334. doi: 10.3390/pathogens4020307.

6. Holcomb, D.A., and Stewart, J.R. 2020. Microbial Indicators of Fecal Pollution: Recent Progress and Challenges in Assessing Water Quality. Curr Environ Health Rep, 7(3), pp.311-324. doi: 10.1007/s40572-020-00278-1.

7. Khan, F.M. and Gupta, R. (2020). *Escherichia coli* (*E. coli*) as an Indicator of Fecal Contamination in Groundwater: A Review. In Jeon, H.-Y. (Ed.), Sustainable Development of Water and Environment. Springer International Publishing, pp. 225-235. Proceedings of the ICSDWE2020.

8. Food and Agricultural Organization of the United Nations, World Health Organization. 2019. Safety and Quality of Water Used in Food Production and Processing - Meeting report Microbiological Risk Assessment Series no. 33. Rome.

9. Reynolds, K. 2012. Water Quality Monitoring: How to Miss an Outbreak. Available at: https://wcponline.com/2012/08/05/water-quality-monitoring-miss-outbreak/ [Accessed 9 June 2024].

10. Frank, J.A., et al. 2008. Critical Evaluation of Two Primers Commonly Used for Amplification of Bacterial 16S rRNA Genes. Appl Environ Microbiol, 74, pp.2461-2470.

11. Bryan, N., Anderson, R., Lawal, O.U., Parreira, V.R., and Goodridge, L. 2023. Draft Genome Sequence of *Exiguobacterium* sp. Strain N5, Isolated from a Recreational Freshwater Kettle Lake in Ontario. Microbiol Resour Announc, 12(4), e0126122. doi: 10.1128/mra.01261-22.

12. Bryan, N., Anderson, R., Lawal, O.U., Parreira, V.R., and Goodridge, L. 2023. Draft Genome Sequence of *Bacillus anthracis* N1, Isolated from a Recreational Freshwater Kettle Lake in Ontario, Canada. Microbiol Resour Announc, 12(4), e0126222. doi: 10.1128/mra.01262-22.

13. Lawal, O.U., Bryan, N., Soni, M., Chen, Y., Precious, M., Parreira, V.R., and Goodridge, L. 2024. Whole genome sequence of *Vibrio cholerae* NB-183 isolated from freshwater in Ontario, Canada harbors a unique gene repertoire. BMC Genom Data, 25(1), 18. doi: 10.1186/s12863-024-01204-2.

14. Stephenson, F.H. 2003. Quantitation of Nucleic Acids. in Calculations for Molecular Biology and Biotechnology, pp.90-108.

15. Shao, K., Yao, X., Wu, Z., et al. 2021. The bacterial community composition and its environmental drivers in the rivers around eutrophic Chaohu Lake, China. BMC Microbiol, 21, 179. doi: 10.1186/s12866-021-02252-9.

16. United Nations Children's Fund and World Health Organization. 2020. Integrating water quality testing into household surveys: Thematic report on drinking water. New York. Available at: https://www.who.int/publications/i/item/9789240014022 [Accessed 9 June 2024].

17. Nowicki, S., deLaurent, Z.R., de Villiers, E.P., Githinji, G. and Charles, K.J. 2021. The utility of Escherichia coli as a contamination indicator for rural drinking water: Evidence from whole genome sequencing. PLoS One, 16(1), e0245910. doi: 10.1371/journal.pone.0245910. PMID: 33481909; PMCID: PMC7822521.

18. Weiss, K. and Cameron, C. 1971. Precipitation of bacterial cells with polymers. United States Patent US3798320A.

19. Daniels, S.L., and Kempe, L.L. 1967. The Separation of Bacteria by Adsorption onto Ion Exchange Resins: II. Resolution of Binary Mixtures. In: Hershey, D. (ed.) Chemical Engineering in Medicine and Biology. Boston, MA: Springer. doi: 10.1007/978-1-4757-4748-5_12.

20. Shafagati, N., Fite, K., Patanarut, A., Baer, A., Pinkham, C., An, S., Foote, B., Lepene, B., and Kehn-Hall, K. 2016. Enhanced detection of respiratory pathogens with nanotrap particles. Virulence, 7(7), pp.756-769. doi: 10.1080/21505594.2016.1185585.

21. Gianazza, E., and Arnaud, P. 1982. Chromatography of plasma proteins on immobilized Cibacron Blue F3-GA. Mech Mol Interact Biochem J, 203, pp.637-641. doi: 10.1042/bj2030637.

22. Regueira-Iglesias, A., Balsa-Castro, C., Blanco-Pintos, T., and Tomás, I. 2023. Critical review of 16S rRNA gene sequencing workflow in microbiome studies: From primer selection to advanced data analysis. Mol Oral Microbiol, 38(5), pp.347-399. doi: 10.1111/omi.12434.

23. Zwart, G., Crump, B.C., Kamst-van Agterveld, M.P., Hagen, F., and Han, S.K. 2002. Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. AME, 28, pp.141-155. doi: 10.3354/ame028141.

24. United Nations. Undated. Goal 6: Ensure access to water and sanitation for all. Available at: https://www.un.org/sustainabledevelopment/water-and-sanitation/ [Accessed 9 June 2024].