Project Title:

Metagenomic analysis of drinking well water for rural and marginalized populations in regional Ontario—implications for testing, including human, public and environmental health.

Applicants: Dr. Anna Majury, PhD, DVM, Department of Biomedical and Molecular Sciences, Division of Microbiology and Immunology and Schools of Environmental Sciences and of Public Health Sciences, Queen's University at Kingston, and Regional Microbiologist, Public Health Ontario. Co-Investigators: Gerald Evans MD, Paul Belanger PhD, Xudong Liu PhD, Queen's University.

Project Proposal

Background:

As part of growing national and international awareness on the need for protection of freshwater sources, which provide water for irrigation, and human and livestock consumption and upon which 30% of Ontarians rely as their primary drinking water source, this area has become a growing subject of research. Within these complex freshwater systems, are active and poorly understood microbial ecosystems. Evidence shows the existence of anthropogenic impacts to microbial ecosystems, but little research has studied the impact of said groundwater microbial communities on the dependent human populations. Traditionally, research utilized cultivation-based techniques to determine most probable number and general activity estimates in drinking water; these methods are only successful for a small portion of microbes and limit study into the diversity of the populations. Recently, molecular methods have become available and have had limited application to groundwater communities, but studies into similar ecosystems, such as soil, have suggested that microbial diversity is much greater than initially anticipated.

One of the most recent technological advancements is the application of metagenomics to groundwater ecosystems; metagenomics promise a greater understanding of bacterial community metabolism and, ultimately, the elucidation of core functions. The basic premise of the metagenomic approach is to treat the collective genomes of a bacterial community as a single large genome, or metagenome. Metagenomics bypasses the need for isolation or cultivation of microorganisms through extraction of total community DNA followed by construction of genetic libraries and subsequent high throughput sequencing or library screening. Metagenomic methods can be used to determine both taxon and functional diversity in microbial communities.

The application of this technology, due to the susceptibility of microbial populations to environmental fluctuations, has largely been limited to contaminated groundwater sites (i.e. nuclear, heavy metal, dense non-aqueous phase liquids), with a goal of understanding the processes by which groundwater microbes can aid in remediation or be used as an indicator in assessing the extent of contamination. So far, the information on bacterial communities in ‘pristine’ and moderately contaminated shallow aquifers is scarce and numerous researchers have identified this gap and called for the application of metagenomic technologies in studies on these systems. Notably, Griebler and Lueders (1), in their review on the history and current understanding of groundwater microbial communities, note that “an ‘endemic’ groundwater microbiota has not been described. … Extensive research is still needed to understand the main factors controlling microbial biodiversity in pristine ground waters”. They further highlight the potential impact of this research: “Our current understanding of key biogeochemical processes and diversity of the microbes involved in and responsible for ecosystem functioning is still

insufficient for most groundwater ecosystems. Sound predictions of how changes in environmental conditions, especially as a result of anthropogenic impacts, may affect microbial diversity and how changes in biodiversity will affect biogeochemical processes and the flow of energy and matter in individual habitats are urgently required"(1). Furthermore, Smith *et al*. (2) connect the variations and on-going perturbations in microbial communities and the impact on human populations. Terrestrial subsurface environments, including groundwater, accommodate the largest reservoir of microbes in the biosphere, with an estimated 3.8-6.0 X 10^{30} cells. These communities also influence the purity of groundwater and subsequent availability of potable drinking water.

A major goal in the study of groundwater microbiology is to determine what the effects of these shifts in microbial ecology have on water quality, and ultimately on the populations and environments impacted by them. Obtaining a broader understanding of the structure and function of microbial communities inhabiting different groundwater systems is particularly important given the increased need for managing groundwater reserves of potable water. Given this identified gap in literature and study on the diversity and structure of microbial communities in pristine and moderately contaminated sites, we propose advancing the knowledge by providing the metagenomes of rural groundwater systems in Eastern Ontario. This work will provide a greater human and ecological context for the metagenomes by considering the impact of variation in microbial community taxon and functional composition on the human populations dependent on the groundwater systems.

Hypothesis/Rationale:

Current estimates indicate that more than 99% of the microorganisms present in many natural environments are not readily culturable and therefore not accessible for biotechnology or basic research. In fact, most of the species in many environments have never been described. Metagenome-based approaches offer a more comprehensive view of the genetic complexity of natural and engineered microbial communities, allowing us to better assess the microbial taxonomic diversity and metabolic potential within any given community. The number of metagenomic studies has increased in recent years due to the availability of next generation sequencing technologies. The discoveries have ranged from novel photosynthetic pathways to genetic pathways that are important in host-microbial interactions, findings that would have been difficult using conventional phylogenetic analyses. Comparison of different metagenomes has further enhanced our understanding of processes unique to some microbiomes and provided the genetic information to track multiple populations carrying a variety of functions. Interestingly, in spite of the public health relevance of drinking water (DW), little to no information is available on DW metagenomes. The goal of this study is to pilot metagenomic analysis of drinking well waters in rural and regional Ontario. Current culture based methods lack in terms of providing detailed information regarding the microbial communities in these varied water sources. Previous studies in our laboratory, using traditional PCR and culture, indicate that the diversity, and hence, the potential health impacts as a consequence of exposure to this water source do exist. Moreover, wells in rural Ontario have been shown to harbour STEC, *C. difficile*, VRE, Norovirus and other potentially infectious organisms (Majury *et al*., unpublished). Further, we propose that the drinking well biotype has an impact on the human microbiome, and vice versa, and that these, consequently, exert overall health and environmental impacts.

Research Plan/Methods:

Using geospatial analytic tools our team previously identified 4 regions at increased risk of *E. coli* contamination, based on current standard water testing methods using culture of *E. coli* (3). Moreover, using microbial source tracking, the primary source of contamination derives from human sources, even in cattle dense regions (4, unpublished data).We propose testing wells from these regions using standard culture techniques for *E. coli* and total coliforms, and by qPCR for *E. coli* and Bacteroidales and using metagenomics, the objective of which will be, using statistical analyses, to determine the relationship between and among the various microbial indicators, and the actual microbiome, as determined by metagenomics, in order to inform risk. Briefly, samples from 25 wells in each region, will be tested. a small volume will be tested using traditional culture methods, and a larger portion reserved for qPCR and metagenomic analyses. Findings will be correlated across the different technologies and regions and with previous and current studies of our team examining microbial sources, and potential waterborne pathogens.

Raw bacteria samples will be collected using standard filtration method with a 0.45 um filter. Bacterial DNA is to be extracted using PowerWater® DNA Isolation Kit, for the isolation of genomic DNA from membrane filtered water samples.

Metagenomics sequencing and bioinformatics analysis will be carried out at Queen's Genomics Laboratory under the direction of Dr. Liu. To explore richness and diversity of the bacteria community in the samples, we will 1) use the fast and cost-effective Ion Torrent's Semi-Conduct Next-Generation Sequencing with Personal Genome Machine platform (http://www.lifetechnologies.com) and, 2) adopt a targeted metagenomics approach, using Ion 16S™ Metagenomics Kit which targets 7 of 9 the hypervariable regions of the 16S rRNA gene (https://www.lifetechnologies.com/order/catalog/product/A26216).

'Ion' sequencing available with the Ion Torrent Personal Genome Machine (PGM) (LifeTechnologies) is a new Next Generation Sequencing technology, which has proven to be an emerging alternative to the previous technologies in NGS field due to its fast turning-out and much lower running cost along with its rapidly growing read length. Several research teams, including Whiteley *et al*., 2012, (5) have recently published microbial metagenomic sequencing studies that used Ion Torrent PGM, showing that the platform is a cost-effective alternative to longer-read sequencers for this application. It is now available longer read length sequencing kit (400bp) and can provide capacity similar to Roche 454 platform annotating the reads down to the species level.

The Ion 16S™ Metagenomics kit is an AmpliSeq based kit designed for rapid, comprehensive and broad-range analyses of mixed microbial populations using the Ion Torrent™ semiconducter sequencing workflow. The kit includes 2 sets of primers that can be used to amplify the corresponding hypervariable regions of the 16S rDNA gene in bacteria: Primer set V2-4-8 and Primer set V3-6,7-9, allowing 7 of 9 hypervariable regions (V2-3-4-6-7-8-9) of the 16S rRNA to be amplified by PCR from bacteria DNA samples. The pooled amplified fragments can then be ligated to Ion Adaptors to construct sequencing libraries. The libraries can then be sequenced using the Ion PGM™ Sequencing 400 Kit on the Ion PGM™ platform on "316" (100 Mb) or "318" (1 Gb) chips, with barcoded adaptors used in the library construction step allowing multiple samples to be sequenced on one sequencing run to increase the efficiency and lower the cost. The sequencing run will result in an average read length of > 350 bp.

Compared with most currently used strategies, which use single or only 2-3 variable regions of 16S rRNA gene with either 454 pyrosequencing or Illumina-based technology, the strategy we will adopt has an obvious advantage with respect to discrimination and sequence assignment in metagenomic analysis.

After sequencing, the data will be analyzed using either the METAP pipeline developed in Dr. Liu's Lab or the Ion 16S™ metagenomics analyses module within the Ion Reporter™ software.

Significance:

Currently, little is known regarding the contribution of private well waters to waterborne illness in Ontario, and current microbial indicators and the methods (culture) utilized to detect and report upon, are poorly understood, such that risk cannot be appropriately determined. Moreover, 1/3 of Canadians are known to rely on groundwater/well water as a primary drinking water source, including an estimated 98% of rural Ontarians. Newer technologies, including qPCR and metagenomics, are now available to shed light on this important issue. As such, this work will serve as a pilot to future studies and the development of an assessment tool explicitly designed to address risk associated with the consumption of private well water in Ontario, and by geographic region. This tool could be adopted by health authorities, both in Ontario, and elsewhere in Canada, in order to better inform testing related practices, including public health and environmental interventions.

References:

1. Griebler C, and Lueders T. Microbial biodiversity in groundwater ecosystems. Freshwater Biology (2009) 54, 649-677.
2. Smith RJ, Jeffries TC, Roudnew B, Fitch AJ, Seymour JR, Delpin MW, Newton K, Brown MH, and Mitchell JG. Metagenomic comparison of microbial communities inhabiting confined and unconfined aquifer ecosystems. Environmental Microbiology (2011) doi: 10.111/j.1462-2929.2011.02614.
3. Krolik J, Maier A, Evans G, Belanger P, Hall G, Joyce A, and Majury A. A spatial analysis of private well water E. coli contamination in southern Ontario. Geospatial Health 8(1), 2013, pp. 66-75.
4. Krolik J, Evans G, Belanger P, Maier A, Joyce A, Guimont S, Pelot A, and Majury A. Microbial source tracking and spatial analysis of *E. coli* contaminated private wells in southeastern Ontario. Journal of Water and Health, 12(2), 2014, doi: 10.2166/wh/2013.192
5. Whiteley AS, Jenkins S, Waite I, Kresoje N, Payne H, Mullan B, Allcock R, O'Donnell A. Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. J Microbiol Methods. 2012 Oct;91(1):80-8.