Drivers of Microbial Community Assembly within an Extent of Fractured Crystalline Rock Relevant to a Geologic Repository

by

Danielle Beaton

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements for the
Ph.D. degree in
Chemical and Environmental Toxicology
Ottawa-Carleton Institute of Biology
and
University of Ottawa
Ottawa, Canada

Thèse soumise à la
Faculté des études supérieures et postdoctorales
en vue de l'obtention du doctorat en
Toxicologie Chimique et Environnementale
L'Institut de biologie d'Ottawa-Carleton
and
Université d'Ottawa
Ottawa, Canada

This thesis is dedicated to my husband and to

the fond memory of my supervisor, Jack Cornett

Enjoy the Sunshine

Abstract

The objective of this thesis was to characterize the microbiology of subsurface fracture water at a location within the Canadian Shield considered as a prospective host site for a geologic repository. Repository performance and long-term safety predictions for geologic confinement of radioactive waste rely on understanding the natural microbial processes that occur within a host formation; however, the inaccessibility of the crystalline terrestrial subsurface means that this habitat is difficult to explore and as such is largely unknown.

The area characterized is located within the boundary of the Chalk River Laboratories site, situated within the Central Gneissic Belt of the Grenville province (formed 1.5 and 1.0 ba years before present); the site is also situated within the Ottawa-Bonnechere graben (formed 0.5 ba before present). Fracture water was accessed from one 34 m deep open drill hole and six cased and sealed drill holes; the Westbay Multilevel Groundwater Monitoring system preserved the natural fracture flow enabling discrete fracture water sampling at multiple depths via each of the sealed drill holes.

This thesis combined multiple datasets in an exploratory analysis for drivers of subsurface microbial assembly within an ecological framework of selection, dispersal, drift and diversification across two spatial extents: 25 km^3 and 1 km^3. The outcomes of the multivariate analyses, null models and generalized linear models identified prospective source waters and distributional relationships of total and viable cell counts with fracture water sulfate and manganese.

Random processes (dispersal, drift and diversification) explain close to 50% of the variance of the phylogenetic beta diversity among the component taxa. Selection associated with differential abundance of the 16S rRNA gene V4 region and spatial and environmental factors identified sulfate and organic carbon plus manganese and a spatial coefficient. Selection by differential abundance was not a major driver of community assembly; accounting for ~4% each of the total abundances linked to sulfate and manganese. At a spatial scale less than 1 km^3, it may be possible to identify either greater significance for sulfate and manganese or to identify additional environmental factors linked to selection.

Demonstrated metabolism included nitrate reduction (common across all sampling locations and at each sampling campaign) and sulfate reduction (observed at one sampling campaign). The distributions of total and viable cell counts correspond with the distributions of sulfate and manganese, respectively. The fracture water was a source of sulfate and manganese, but not a source of nitrate. A limited analysis of rock porewater identified sulfate and nitrogen compounds (ammonia, nitrite and nitrate) in parts per million concentrations, suggesting that the rock is a source of these compounds, and this finding warrants an assessment of rock weathering as a driver of selection at these sampling locations.

Phylogenetic relationships across sampling locations showed that the component taxa were more closely related than expected by chance; this pattern suggests that, at the spatial scale of the analysis, competitive exclusion was not a driver of subsurface community assembly. Co-existence of close relatives may be biologically relevant—and thus be a sign of diversification—or this pattern may reflect the 16S rRNA gene copy number combined with intra-genomic heterogeneity greater than the 97% sequence similarity threshold for binning sequencing reads into organizational taxonomical units (OTUs).

The distribution of genes for energy metabolism was uniform across all sampling locations. A metatranscriptome assessment would help differentiate between the genes present from the genes expressed. Testing for a wider range of demonstrated metabolic capabilities would support RNA-level gene expression analyses.

Overall, applying the ecological framework of four main drivers of community assembly show that, at the spatial scale of the sampling, up to 50% of the variance among community dynamics reflect randomness. Approximately ~1% of the total abundance was linked to measured metabolism; at smaller spatial scales, ~8% of the total 16S rRNA gene abundance was linked to differential abundances across—potentially connected--sampling locations. By sampling at smaller spatial scales, therefore, it may be possible to discern additional metabolic and selective processes. These data will inform models for the performance and long-term safety of geological repositories.

Résumé

L'objectif de cette thèse était de caractériser les microorganismes présents dans l'eau d'un aquifère fracturé situé sur le socle précambrien, lequel est un site envisagé pour le stockage des déchets nucléaires.

Les performances du stockage et les prédictions à long terme de la stabilité de ces sites de stockage dépendent cependant de nos connaissances des processus microbiens qui peuvent exister dans ces sites. Ces sites profonds sont toutefois rarement accessibles ce qui limite nos connaissances.

Le site étudié est situé au Laboratoires Nucléaires de Chalk River, dans la province géologique du Grenville (formée entre 1.5 et 1 milliard d'années) et dans la formation graben Ottawa-Bonnechere (formée il y a 0.5 milliards d'années). L'eau présente dans les fractures a été échantillonnée dans un puit foré de 34 mètres de profondeur et dans 6 autres puits scellés. Un système de mesure appelé "Westbay Multilevel Groundwater Monitoring system" a permis de préserver l'écoulement naturel de l'eau et l'échantillonnage à plusieurs profondeurs.

Cette thèse a combiné plusieurs ensembles de données dans une analyse exploratoire des facteurs affectant l'assemblage microbien souterrain dans un cadre écologique de sélection, de dispersion, de dérive et de diversification à travers deux étendues spatiales: 25 km^3 et 1 km^3. Les résultats des analyses multivariées, des modèles nuls et des modèles linéaires généralisés ont permis d'identifier les sources d'eau potentielles et les relations distributionnelles des dénombrements cellulaires totaux et viables avec le sulfate présent dans l'eau des fractures et le manganèse.

Les processus aléatoires (dispersion, dérive et diversification) expliquent près de 50% de la variance de la diversité bêta phylogénétique parmi les taxons constitutifs. La sélection associée à l'abondance différentielle de la région du gène V4 de l'ARNr 16S et les facteurs spatiaux et environnementaux ont identifiés le sulfate et le carbone organique plus manganèse et un coefficient spatial. La sélection par l'abondance différentielle n'était facteur importe l'assemblage communautaire; représentant environ 4% de chacune des abondances totales liées au sulfate et au manganèse. À une échelle spatiale inférieure à 1 km^3, il peut être possible

d'identifier soit une plus grande signification pour le sulfate et le manganèse, soit d'identifier des facteurs environnementaux supplémentaires liés à la sélection.

Le métabolisme identifié dans cette étude inclut la réduction des nitrates (commune à tous les sites d'échantillonnage et à chaque campagne d'échantillonnage) et la réduction des sulfates (observée lors d'une campagne d'échantillonnage).

Les distributions des comptes de cellules totaux et viables correspondent respectivement aux distributions de sulfate et de manganèse. L'eau de fracture était une source de sulfate et de manganèse, mais non pas une source de nitrate. Une analyse limitée de l'eau interstitielle a identifié des composés sulfatés et azotés (ammoniac, nitrite et nitrate) présents en partie par million (ppm), suggérant ainsi que la roche est une source de ces composés, et cette constatation justifie une évaluation de l'altération des roches en tant que facteur de sélection de ces sites d'échantillonnage.

Les relations phylogénétiques entre les sites d'échantillonnage ont montré que les taxons étaient plus étroitement liés que prévu par hasard; ce modèle suggère que, à l'échelle spatiale de l'analyse, l'exclusion compétitive n'était pas un moteur de l'assemblage de la communauté souterraine. La coexistence de parents proches peut être biologiquement pertinente - et donc être un signe de diversification - ou ce modèle peut refléter le nombre de copies du gène 16S ARNr combiné à une hétérogénéité intra-génomique supérieure au seuil de similarité de séquence de 97% pour le séquençage des unités taxonomiques (OTU).

La distribution des gènes pour le métabolisme énergétique était uniforme dans tous les sites d'échantillonnage. Une évaluation du métatranscriptome a permis de différencier les gènes présents des gènes exprimés. Le test d'une plus large gamme de capacités métaboliques démontrées soutiendrait les analyses d'expression génétique au niveau de l’ARN.

Dans l'ensemble, l'application du cadre écologique des quatre principaux facteurs de l'assemblage communautaire, à l'échelle spatiale de l'échantillonnage, jusqu'à 50% de la variance entre les dynamiques communautaires, reflète un caractère aléatoire. Approximativement 1% de l'abondance totale était liée au métabolisme mesuré; à des échelles spatiales plus petites, ~8% de l'abondance totale du gène de l'ARNr 16S était liée à des

abondances différentielles dans des emplacements d'échantillonnage potentiellement connectés. En échantillonnant à des échelles spatiales plus petites, il est donc possible de discerner des processus métaboliques et sélectifs supplémentaires. Ces données informeront les modèles sur la performance et la sécurité à long terme des sites de stockage géologiques.

Co-Authorship

Danielle Beaton is the primary author of this thesis and all enclosed documents. Chapters 2 to 5 were written as independent manuscripts in which the co-authors provided intellectual supervision and editorial comment, except where noted below.

Chapter 2 is published in Frontiers in Microbiology, 2017 and is co-authored by Dr. Ioana Gurban, Dr. Karen J. King-Sharp and Dr. Marilyne Stuart. Chapter 3 is also published in Frontiers in Microbiology and is co-authored by members of Dr. Bradley S. Stevenson's laboratory at the University of Oklahoma, Dr. Blake W. Stamps and Dr. Heather S. Nunn. Chapter 4 and Chapter 5 are also co-authored by members of Dr. Bradley S. Stevenson's laboratory at the University of Oklahoma. Dr. Sim Stroes-Gascoyne is a co-authored of Chapter 5.

Acknowledgements

I would like to thank my supervisor Jack Cornett for his support, mentoring and for his leading by example. I will remember Jack's passion for science and for life as shown by his stance of helping others achieve their goals. His mentorship encouraged me to explore creative options for communicating ideas and to seek ways that I too could help others. Enjoy the sunshine.

A special thank you goes to Danielle Fortin who took over as my supervisor and who, as a member of my committee, has given me scientific and moral support throughout my research years and the completion of this thesis. I am also grateful to my other committee members, Dr. Dmitry Klokov, DR. Ruth Wilkins and DR. Richard Goulet; without their advice and guidance, this thesis would not have been possible.

I would also like to thank the Director of Waste Management and Decommissioning, Dr. Steve Liblong and Tracy Sanderson from Canadian Nuclear Laboratories, who enabled me to continue my work on the Geological Waste Management Facility Project by completing this thesis. I thank Dr. David Rowan and Dr. Marilyne Stuart, Dr. Ian Castillo and the Director of Environmental, Radiological and Chemical Sciences, Dr. Joanne Ball, also from Canadian Nuclear Laboratories, who enabled me to start a new area of research. I also thank Jennifer Olfert, Penny Neal, Peter Angell and Doug Rodgers.

My understanding of the subsurface habitat benefited from interactions with my many fellow colleagues who worked on the multidisciplinary Geologic Waste Management Facility Project. Some these colleagues were Jamil Sidiqqi, Karen Sharp, Ed Kozak, Robert Thivierge, Robert Sikorski, Kevin Campbell and Diego Pucciarelli Ayllon.

Within Research and Development, sampling and analysis of the deep subsurface would not have been possible without the contributions made by Dan Festarini, Lee Bellan, Stephen Rose, Amy Festarini, Mandy Serran, Hui Qun (Grace) Chen, Kirsten Scott and many more.

The Geological Waste Management Facility Project was funded by Natural Resource Canada.

Last, but not least, I would like to thank my husband for his years of support and encouragement. You made it possible and worthwhile.

TABLE OF CONTENTS

TABLES

FIGURES

Abbreviations

ATP	Adenosine triphosphate
CNSC	Canadian Nuclear Safety Commission
CRG	Chalk River Geologic
CRL	Chalk River Laboratories
DGR	Deep geologic repository
GWMF	Geologic waste management facility
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	KEGG Orthology
L/ILW	Low-to-intermediate waste
MEM	Moran's eigenvector map
MG-RAST	Metagenome Rapid Annotation using Subsystem Technology
NRB	Nitrate reducing bacteria
NUB	Nitrate utilizing bacteria
NWMO	Nuclear Waste Management Office
PCA	Principal Component Analysis
PCoA	Principal Coordinate Analysis
PCR	Polymerase chain reaction
QIIME	Quantitative Insights into Microbial Ecology
SRB	Sulfate reducing bacteria

1. INTRODUCTION

1.1 Thesis Rationale

Low and intermediate level radioactive waste requires long-term management and confinement for approximately a few hundred to tens of thousands of years, depending on half-lives of the component radionuclides. High-level fuel waste requires much longer confinement of hundreds of thousands of years before the radionuclide inventory decays to the point when activity levels would be equivalent to the radioactivity from natural uranium oxide (NWMO backgrounder report, 2016). Within Canada, the fundamental rationale for responsible long-term management of such waste is protection of human health and the environment for present and future generations. As such, geological disposal is proposed for safe long-term confinement. Locating a repository, therefore, requires characterizing the prospective subsurface volume for its suitability for confining and isolating waste over the respective period of radioactive decay. Understanding the natural processes occurring within a subsurface volume--including microbial processes and weathering over geological time—are central to responsible long-term management of geologically confined waste. Before attempting to predict how *in situ* microbial processes could affect the performance and long-term safety of a repository, this thesis aimed to first characterize the subsurface microbiology of a prospective host site by analyzing site data: first by comparing observed total abundances and viability with the observed geochemistry as in indicator of prospective source waters and as the medium for prospective electron donor/acceptor pairs; then by comparing 16S rRNA gene assemblages—who is there, in what abundance and how are they related--with the corresponding geochemical and spatial datasets; then by comparing the 16S rRNA gene and metagenome assemblages of genes for energy metabolism with the geochemical and spatial datasets—including rock porewater. Growth on selective media provided evidence for gene expression. Comparisons were made by multivariate analysis, null models and generalized linear modeling. Outcomes from these analyses were interpreted under an ecological framework with four main drivers of community assembly and co-existence: selection, dispersal, drift and diversification.

1.1.1 Geological Confinement of Radioactive Waste

Based on the current nuclear generating capacity, an estimated 5.2 million spent CANDU nuclear fuel bundles will be stored at Canadian reactor sites[1]. In 1978, Atomic Energy of Canada Limited (AECL) received direction from the governments of Canada and Ontario to develop the concept of deep geological disposal of current and anticipated nuclear fuel wastes (Seaborn Panel report, 1998) and in 1988, the Canadian government referred the concept for public review. The outcome of this referral is documented the Seaborn Panel Review (1998). Included in the review was an independent scientific review group to examine the safety and scientific acceptability of the concept. Following from this review, in accordance with the Nuclear Fuel Waste Act[2], Canada's nuclear electricity producers established the Nuclear Waste Management Office, NWMO, in 2002. Under this act, the NWMO mandate was to assess methods for the management of nuclear fuel waste, of which each of the following methods must be the sole basis of at least one approach:

(a) deep geological disposal in the Canadian Shield;

(b) storage at nuclear reactor sites; and

(c) centralized storage, either above or below ground.

The NWMO is also responsible for designing and implementing Canada's plan for the safe, long-term management of used nuclear fuel. Their Adaptive Phased Management plan requires used fuel to be contained and isolated in a deep geological repository. As part of a related assessment for the management of low-to-intermediate level waste, Ontario Power Generation evaluated the Bruce Nuclear site as a location for hosting a low-to-intermediate level waste deep geologic repository; and thus, addressed part of the NWMO adaptive phased management, approach (b). The Bruce Nuclear site is located within a sedimentary formation. Over the intervening years, NWMO has evaluated numerous prospective host sites within the

[1] https://www.nwmo.ca/en/Canadas-Plan/Canadas-Used-Nuclear-Fuel/How-Much-Is-There

[2] http://laws-lois.justice.gc.ca/eng/acts/N-27.7/

Canadian Shield; on the list for further study are the Ontario locations: Hornepayne, Manitouage and Ignace.

Separate from these assessments, in 2007 Natural Resources Canada funded Atomic Energy of Canada Limited to evaluate the Chalk River Laboratories (CRL) subsurface as a prospective host site for managing low-to-intermediate level waste. The microbiological assessment described herein was part of this assessment.

1.2 Characterizing the Subsurface at Chalk River Laboratories

Chalk River Laboratories in Ontario is a licensed waste management operator situated within the Grenville Province of the Canadian Shield; the oldest landmass in North America. From 2007 to 2014, an approximately 25 km^2 area within the Chalk River Laboratories boundary was evaluated for its geological suitability for hosting a low-to-intermediate level waste repository. Characterization of the geology and fracture density of the rock mass involved drilling a series of boreholes to depths of 1.2 km to obtain rock cores. Characterization of the fracture water involved first casing the resulting boreholes and sealing them to isolate discrete fracture zones--and thus to preserve, as much as possible, natural fracture water flow (Neymark et al., 2013; King-Sharp et al., 2016). The installed Westbay™ multilevel groundwater monitoring system enabled sampling of the fracture water for hydrology, geochemistry and microbiology; a description of this sampling system is in the Methods Sections of the following chapters.

1.2.1 Inferring Long-Term Subsurface Microbial Processes

Although much has been learned since confirming that there are microbial processes occurring within terrestrial subsurface sediments (Lovley and Chapelle, 1995) and fractures within granitic rock (Pedersen, 1997), opportunities for accessing the terrestrial subsurface are limited; we therefore lack data and specific information at prospective host sites on the relevant geochemical and microbiological dynamics. Where researchers have studied subsurface environments, the application of non-culturing techniques allowed for the identification of a greater microbial diversity than what was previously possible. What this diversity means when seeking to site a repository is a new area of research. Specific question to address include:

- What taxa are present: who is there and in what abundance?
    - What proportion of the component taxa display viability and is this viability correlated with environmental factors?
    - What proportion of the component taxa display differential abundance and is this difference correlated with environmental factors?
- At the whole community level, how do local and regional diversities compare; and what do these metrics tell us about how these communities formed?
    - Can local and regional environmental or spatial factors explain the observed diversity?
    - By making diversity the independent variable, can we predict diversity driven processes over the timeframe for the radioactivity within a repository inventory to decay to background?
- How do the local and regional diversity compare with the corresponding diversity of functional genes?

Non-culturing techniques such as those that rely on the polymerase chain reaction (Mullis et al., 1986) reveal that there is more microbial diversity present within environmental samples than is represented in culture collections. Non-culturing techniques have the advantage of side-stepping our inability to culture most microbial taxa; but this approach can introduce bias by our tendency to infer that the detected but uncultured taxa possess similar characteristics as those of their known cultivated relatives. A second bias when applying non-culturing techniques to the study of microbial processes in natural environments is the assumption that the detected taxa are extant and thus contribute, in some way, to ecosystem functions. In fact, many of the identified taxa in an environment are in low abundance; and, as discussed in Section 1.3.1 on drivers of community assembly across spatial scales, the component taxa detected at a sampling location may reflect randomness.

Related to this bias is the possibility of detecting remnant DNA--or preserved DNA from the surrounding environment that is not associated with intact microbial cells. The presence of remnant DNA would have more impact on data obtained from, for example, shotgun metagenome sequencing that employs random primers to sample the range of genes present

across multiple genomes--rather than techniques that employ primers targeting specific genes. Given the propensity of bacteria to incorporate fragmented and damaged DNA from the environment (Overballe-Petersen, 2013) it may be difficult to detect the contribution of remnant DNA in such datasets. Furthermore, bacterial gene transfer, either from lysed microbial cells or by lateral transfer, would suggest that the traits represented within a sampled community might not be phylogenetically conserved.

Lastly, attempts to make connections between the subsurface geochemistry and the subsurface microbial communities is limited by the geochemistry that was opted to measure; these choices can also introduce bias by effectively blinding us to processes that can influence *in situ* redox conditions, and therefore the fate and mobility of introduced radioactive compounds.

To address uncertainties of correlating community patterns with the environment, I explored the application of an ecological framework of community assembly to evaluate the observed patterns obtained from culturing, microscopy and sequencing of microbial (mainly bacterial) DNA from fracture water sampled over a spatial extent within the Canadian Shield relevant to a geological repository. By targeting the 16S rRNA gene, observed patterns include data on taxa identities, their relative abundances and their shared history. By employing shotgun metagenome analysis, observed patterns of prokaryotic functional genes allowed for a comparison between phylogenetically conserved patterns across sampling locations and the abundances of genes for energy metabolism. An assumption inherent in this framework is that community patterns reflect phylogenetically conserved traits. Such hierarchical relationships assume gene transfer from parent to the resulting daughter cells upon cell division.

1.3 Everything is Everywhere: but the Environment Selects

Ideas do not arise in a vacuum.

The first part of the 1934 statement by Baas-Becking: “everything is everywhere: but the environment selects” asserts that, due to unlimited dispersal, no place on the Earth’s surface is devoid of microbial life. Two reviews by O’Malley in 2007 and 2008 document the history of this maxim in the context of nineteenth and twentieth century philosophical concepts of plant and animal ecology—concepts that sought to explain why everything was not everywhere--or

biogeography--and the influence these concepts had on the relatively new science of microbiology. In her review, O'Malley notes that the empirical and theoretical history of plant and animal biogeography formed from multiple "conceptual and epistemological tensions". Amongst these tensions were divine vs. physical forces; ecological determinism vs. contingent historical events; environmental determinism vs. random dispersal; constant dispersal vs. contingent processes affecting landmasses occurring over geological time.

Around this same period, the practical study of microbiology—motivated by apparent spontaneous generation leading to fermentation, putrefaction and disease—transformed into a science when Robert Koch postulated the germ theory of disease transmission following from developments in pure culture techniques, sterilization and re-transmission of disease via 'infusium' (Brock 1961). These developments enabled Martinus Beijerinck to demonstrate chemolithotrophy of nitrate dependent sulfur oxidation in 1904 (Brock 1961)—prompting him to propose that consistent microbial capabilities within environmental samples could be demonstrated by defining the physical and nutritional conditions.

Also around this period, the tensions between proponents for constant dispersal over stable landmasses and vicariance--or historical contingency by events resulting in geographical separation--were influencing Darwin's theory of evolution (O'Malley, 2007; O'Malley, 2008). In keeping with the dispersalist thinking of the time, Beijerinck proposed that the presence or absence of microbes was predictable: instead of spontaneous generation, microbial life was everywhere and it was evident wherever there were appropriate conditions. In support of this statement, Baas-Becking published empirical measurements for the redox and pH conditions at multiple aqueous environments and plotted the corresponding cultivated microbes from these locations onto the stability diagram for water (Baas Becking et al., 1960). These diagrams highlight the second part of this statement, that abiotic conditions constrain the microbes such that only the most abundant phenotypes would be detected. Yet natural environments display diverse microbial species, most which at any given time are of low abundance.

1.3.1 Assembly of Subsurface Microbial Communities

Many processes are involved in determining which species coexist and assemble into communities. Three major factors influencing community composition and diversity are niche-based biotic and abiotic competition (Keddy, 1992; Chesson, 2000; Adler et al., 2007; Mayfield and Levine, 2010), neutral assembly of ecologically equivalent component taxa (Hubbell, 2001), and prior events such as dispersal (Ricklefs, 1987).

Competition and co-existence driven by niche differences are among the most commonly studied outcomes of these processes. For example, microbial ecology focuses on niche differences linked to dissimilatory metabolism between electron donor and electron acceptor pairs; these pairs and the resulting free energy differences drive ATP biosynthesis and form much of the basis for microbial taxonomic classification. A common concept about the role of competition on community assembly is that closely related taxa compete more intensely than more distantly related taxa thus limiting their ability to co-exist. With the application of DNA sequence-based phylogenetic methods to characterize microbial communities, patterns of relatedness and co-existence have emerged that are described by ``environmental filtering``. Although this statement is akin the Bass-Becking statement, filtering is invoked when the component taxa are found to be more closely related than what would be predicted by either chance or competitive exclusion (Horner-Devine and Bohannan, 2006; Kraft et al., 2015; Cadotte and Tucker, 2017). To reconcile the roles of multiple mechanisms, a more nuanced explanation has been proposed: that competition alone can explain both dispersion of related taxa (by competitive exclusion) and clustering of related taxa (by competitive ability of shared traits that enable co-existence) (Mayfield and Levine, 2010). When applied to microbial communities, this explanation may provide an alternative hypothesis to competitive exclusion when testing for the genetic characteristics that enable microbes to persist within a subsurface.

Another observed pattern that has emerged with the application of DNA-based phylogenetic methods to microbial communities in natural environments is ``decay of similarity with geographic distance`` (Martiny et al., 2011; Horner-Devine and Bohannan, 2006; Lear et al., 2014) that spans from meters to kilometers. Distance decay of similarity is a concept in

biogeography (Nekola and White, 1999) and is used to explain both differences in species along geographic gradients—a niche based process--and difference in dispersal abilities.

1.3.1.1 Vellend's Ecological Framework of Community Assembly: Four Fundamental Ecological Processes

Table 1-1 provides a list of key terms related to community assembly and Figure 1-1 from Zhou and Ning (2017) depicts the ecological framework proposed by Vellend (Vellend, 2010; Vellend, 2016). This framework draws from the main processes applied to the study of population genetics--that of selection, dispersal, drift and speciation--or in the case of microbes, diversification. Included in this Figure is a role for priority effects (Zhou and Ning, 2017) that would have niche-based aspects and thus could be a secondary process under the main processes of selection and dispersal as defined by Vellend.

This ecological framework also falls within the meta-community framework. A meta-community is defined by a set of interacting communities connected by dispersal and composed of multiple, potentially interacting taxa. A meta-community therefore defines a regional pool of taxa representing multiple local communities sampled at a place and time. A goal of the framework, therefore, is the simultaneous consideration of multiple processes occurring at multiple scales and on how they influence observed community patterns. By including patterns of phylogenetic relatedness (Horner-Devine and Bohannan, 2006; Emerson and Gillespie, 2008) it may be possible to detect the drivers of selection within a subsurface extent.

Table 1-1
Terminology used in this thesis related to community assembly (modified from Zhou and Ning (2017))

Term	Definition
Meta-community	A set of interacting communities connected by dispersal and composed of multiple, potentially interacting taxa.
Taxon Plural: taxa	A grouping of closely related individuals; in this thesis, a taxon is represented as sequence similarity of variable regions within the 16S rRNA gene. Taxa were identified to the level genus.
Community assembly	Processes by which taxa migrate within a regional pool of potentially interacting taxa to establish and maintain local communities.
Selection	Niche-based process that shape community structure due to fitness differences among different taxa, including effects of abiotic conditions, biotic interactions and niche based priority effects.
Homogeneous selection	Selection under homogeneous abiotic and biotic environmental conditions and that lead to more similarity among communities.
Variable selection	Selection under heterogeneous abiotic and biotic environmental conditions leading to more dissimilarity among communities.
Dispersal	Movement of taxa from one location to another and connecting regional and local spatial scales.
Homogenizing dispersal	Dispersal of taxa among communities having the effect of homogenizing the communities such that their structures appear to be similar.
Dispersal limitation	Isolation of taxa among communities, limiting interactions where their structures appear to be dissimilar.
Diversification	Evolutionary process of generating new genetic variants; balancing between speciation and extinction
Drift	Variation in community structure with respect to taxa identities and/or functional traits due to stochastic processes of birth, death, immigration and emigration, spatiotemporal variation, and/or historical contingency

Within an environment, selection of bacterial taxa can occur under unvarying environmental conditions; the homogeneous biotic and abiotic conditions of this environment (Figure 1-1, selection under a common environment “I”) results in low rates of community turnover,

consequently, local communities observed within each sampling location would appear to be highly similar. Selection of bacterial taxa can also occur under variable environmental conditions. The heterogeneous biotic and abiotic conditions of variable environments (Figure 1-1, selection under different environments "I" and "II") result in high rates of community turnover; consequently, local communities observed at each sampling location would appear to be dissimilar. If these differences in environment occurred along an environmental or geographic gradient, community turnover may display decay of similarity with distance along the gradient.

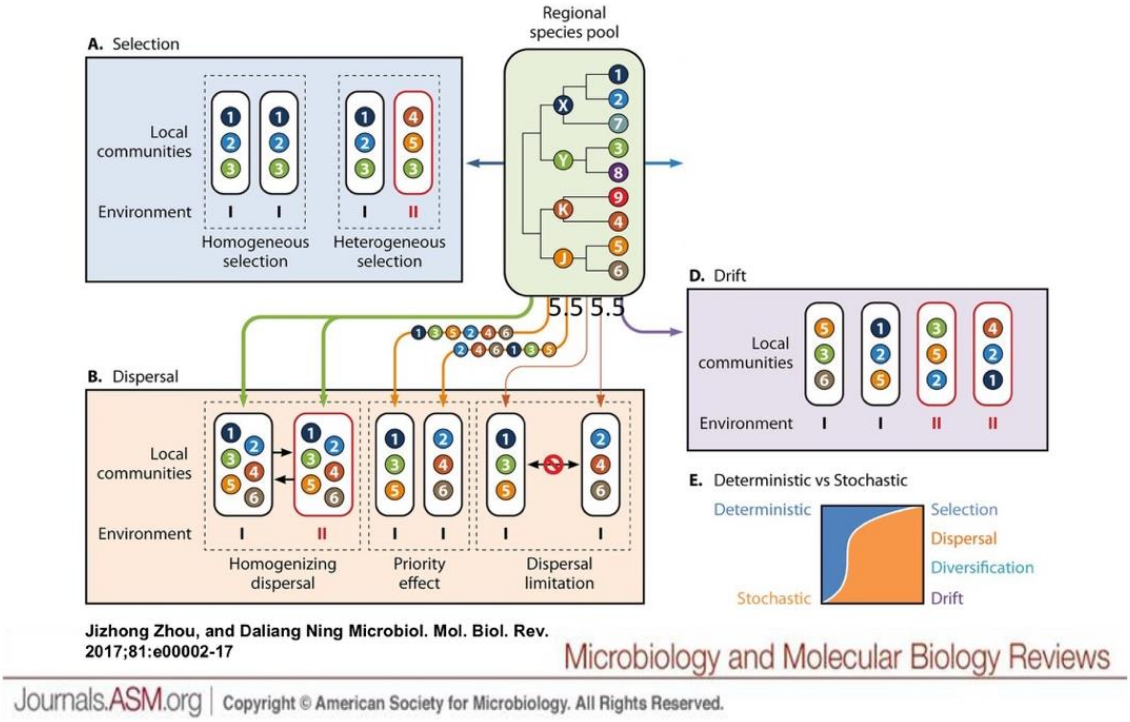


Figure 1-1 Schematic representation of microbial community assembly by the ecological processes proposed by Vellend (Vellend, 2010; Vellend, 2016) and their influences on observed community composition and abundances across spatial scales. This Figure was taken from Zhou and Ning (2017).

Depending on the concomitant dispersal between a pair of communities (sampling locations), the influences of a variable environment can be obscured (Figure 1-1, homogenizing dispersal)

by the homogenizing effect of a high dispersal rate, even when selection pressure is strong. Conditions of high dispersal has the effect of lowering compositional differences between communities (Vellend, 2010) and thus communities at each sampling location would appear to be similar. On the other extreme, conditions when dispersal of taxa between communities is low or even limiting (Figure 1-1, limiting dispersal), observed patterns may be dissimilar, even if the environment is similar. Lastly, when selection pressure is weak, and the dispersal rate between communities is limiting, the dynamics of turnover within a community are subject to drift and random variations in community composition and abundances (Figure 1-1, drift). Depending on the influences of dispersal and drift, it may be difficult to identify the drivers of selection.

By accounting for multiple spatial scales, Vellend's four main ecological processes provide a rational basis for evaluating microbial processes within an extent of the Canadian Shield subsurface—an environment that is difficult to access and therefore microbial community assembly and drivers of selection are largely unknown. Being able to model the natural processes within such an extent—and over the timeframe required for radioactivity to decay to background--can form part of the models for repository performance (the so-called near field) and long-term safety (the so-called far field).

1.4 Overview of the Sample Processing and Main Findings

The work described in this thesis was part of a larger project for characterizing the geology of an approximately 25 km^2 area at Chalk River Laboratories (Ontario, Canada); as such, this thesis offers an exploratory analysis of multiple data sets collected between 2007 and 2014. I applied multivariate analysis, null models and generalized linear models to assess correlations between the fracture water microbiology and its geochemistry. I also assessed spatial correlations between sampling locations.

Analyses for fracture water microbiology included:

- Total and viable microscope counts using fluorescent dyes
- Viable counts by most probable number and heterotrophic plate counts
- 16S rRNA gene:

    - Performed on samples collected in 2008 and 2011: variable regions targeted: V1-V2 using Roche 454 pyrosequencing Genome Sequencer FLX with GS FLX Titanium series reagents. Spatial extent: ~25 km^2 area and ~1 km depth.
    - Samples collected in 2013 and 2014: variable region targeted: V4 using Illumina paired end on the MiSeq technology, PE250 V2 chemistry. Spatial extent: ~1 km^2 area and ~0.9 km depth.
- Shotgun metagenome:
    - Performed on samples collected in 2013 and 2014: random primers using Illumina paired end on the MiSeq technology. Spatial extent: ~1 km^2 area and ~0.9 km depth.

The area characterized is located within the boundary of the Chalk River Laboratories site, situated within the Central Gneissic Belt of the Grenville province (formed 1.5 and 1.0 ba years before present); the site is also situated within the Ottawa-Bonnechere graben (formed 0.5 ba before present). Fracture water was accessed from one open drill hole and six cased and sealed drill holes via the installed Westbay™ Multilevel Groundwater Monitoring system; his system was installed to preserved the natural fracture flow and to enable sampling fracture water at depths. Table 1-2 lists the boreholes by name and by the interval numbers. This table also lists the sampling dates, the DNA sequencing technology employed, the interval midpoint elevation relative to sea level and the distances between interval midpoints within the respective borehole.

Borehole intervals were sampled for microbiological analysis only after confirming that the fluorescein dye content—added to the drilling fluid at a nominal concentration of 1000 μg/L—was less than 1% of the fracture water. The stainless steel Westbay™ sampling tubes were sterilized by autoclave; validation of the sampling process involved filling sterilized tubes with sterile deionized water and performing a mock sampling and tube handling; afterwards the water from the validation was processed for DNA extraction and tested for contaminating DNA by polymerase chain reaction using a primer pair targeting bacterial 16S rRNA gene and agarose gel electrophoresis. Each sampling campaign occurred over two days—two samples collected on day-one and another two samples collected on day-two. Filtering the fracture water for

DNA extraction commenced immediately upon receipt within the laboratory. Aliquots were also tested for total and viable microscope counts and for aerobic and anaerobic heterotrophic plate counts. Most probable number was performed by a separate laboratory within Canadian Nuclear Laboratories. The filters for DNA extraction from day-one were stored at -20°C overnight; DNA from of all samples from the campaign were then extracted at the same time on day-two. All DNA extracts from the first day were stored at -20°C; DNA yield and quantitative PCR—if performed—was performed by the laboratory that prepared the libraries for sequencing. DNA extracts from 2008 and 2011 (Table 1-2) were sequenced in 2011 by 454 pyrosequencing; DNA extracts from 2013 and 2014 (Table 1-2) were sequenced in 2014 by Illumina paired end sequencing. All libraries were prepared by the same individual and sequencing was performed within the same facility.

The reads for the targeted 16S rRNA gene variable regions were filtered for quality (phred scores), read length and contaminating sequences from the library preparation. The conservative clustering algorithms, USEARCH (454 pyrosequencing) and UPARSE (Illumina paired end sequencing), were used for read binning into organizational taxonomical units (OTUs). These algorithms minimize the risk of classifying spurious OTUs compared to the default clustering algorithm used within QIIME 1.9.1, UCLUST. Even so, the single end reads from 454 pyrosequencing would be more prone to spurious OTUs identification than the Illumina paired end reads. The higher number of OTUs identified by 454 pyrosequencing (Chapter 3) than by Illumina paired end sequencing (Chapter 4 and Chapter 5) could reflect either the sequencing technology employed or the larger spatial extent covered in the earlier analysis—25 km^3 compared to 1 km^3. The following sections provide an overview of the main findings documented in Chapter 2 through to Chapter 5.

Table 1-2
Details of sampling locations used for DNA sequence analysis

Borehole	Interval:	Sampling Date	Easting *	Northing *	Elevation **	Distance between sampling locations		Sequencing technology ***
			m	m	m	sequential m	cumulative m	
CRG-1	3	05/30/2011	313743.6	5102838	31	--	0	454
	6	05/31/2011	313760.8	5102857	-68	99	99	
	8	06/01/2011	313771.9	5102870	-131	63	162	
	14	06/02/2011	313815.2	5102917	-356	225	387	
CRG-2	2	09/20/2011	314757.5	5104142	92	--	0	454
	4	09/21/2011	314777	5104160	-38	130	130	
CRG-4A	9	08/23/2011	312203.4	5104452	-440	--	--	454
CRG-3	6	23/09/2013	313344.3	5102404	-122	--	0	Illumina
	8	25/09/2013	313335.7	5102336	-247	125	125	
	11	26/09/2013	313318.3	5102241	-417	170	295	
	14	27/09/2013	313293.2	5102122	-617	200	495	
CRG-6	3	06/05/2013	313391.2	5102135	48	--	0	Illumina
	4	12/05/2013	313373.8	5102116	-4	52	52	
	5	14/05/2013	313351.6	5102092	-68	64	116	
	11	13/05/2013	313234.8	5101980	-390	322	438	
CR-9	2	11/04/2008	312721.5	5102262	29	--	0	454
	3	11/04/2008	312691.7	5102245	-26	55	55	454
	5	11/06/2008 05/05/2014	312449.75	5102130	-154	128	183	454 Illumina
	8	11/06/2008 06/05/2014	312466.6	5102145	-364	210	393	454 Illumina
	11	11/13/2008 12/05/2014	312449.75	5102130	-384	20	413	454 Illumina
	12	11/13/2008 13/05/2014	312416.14	5102120	-426	42	455	454 Illumina
CR-18 (open)	--	09/22/2011	311262.7	5103333	111	--	--	454

* elevation relative to sea level

** Zone 18 of the Universal Transverse Mercator coordinate system

***Roche 454 pyrosequencing, Genome Sequencer FLX, GS FLX Titanium series reagents; Illumina PE, MiSeq, PE250 V2 chemistry

1.4.1 Spatial Correlations in the Abundance and Geochemical Data

Chapter 2 explored correlations between the fracture water geochemistry and the total and viable microscope counts using generalized linear models and multivariate analyses. It was found that, at the spatial scale of the samplings, abundances linked with geochemistry were not unambiguously discerned because of spatial autocorrelation in the geochemical compositions. When Moran's eigenvector map (MEM) coefficients were included in a generalized linear model (GLM) for cell abundances, the distribution of model residuals was improved. Although the GLM identified spatial correlations in both the geochemistry and the abundances, the distributions of cell abundance were normal to the distributions of fracture water sulfate (total cell count) and fracture manganese (viable cell count), a finding that suggests they may have a common cause.

A second goal of the analysis was to assess recharge on the assembly of subsurface communities. By multivariate mixing and mass balance (M3) modeling, all fracture water samples were identified as being of meteoric origin. The geochemical composition of the fracture water was bounded by three endmembers: modern recharge, glacial melt and a saline source—a finding that is consistent with the post glacial history of the site. From these data, the bicarbonate content of the fracture water was used as proxy for modern recharge, stable oxygen isotopic content was used as proxy for glacial melt water and the chloride content was used as proxy for Champlain Sea. A generalized linear model of total and viable abundances of the component microbial cells correlated with the spatially correlated bicarbonate; therefore, it was not possible to unambiguously identify recharge as a driver of microbial total and viable abundances.

1.4.2 Diversity of the 16S rRNA Gene across a 25 km^3 Subsurface Spatial Extent

Chapter 3 explored the diversity of the 16S rRNA gene and possible correlations with the subsurface geochemistry. Taxa identities, abundances and phylogenetic relationships from the 16S rRNA gene V1-V2 region within each sampling location (assemblages of local communities) were combined into a regional pool of taxa (assemblages forming a meta-community). This is the first time I applied Vellend's ecological framework. Null models and multivariate analyses

were applied to compositional and phylogenetic diversity metrics for alpha (local) and beta (among communities). The operational categories defined by Stegen (2015) were used to identify the relative contributions of selection, dispersal and drift on the Chalk River subsurface communities.

A main finding from this analysis was that the component taxa within each community were more closely related than expected by chance—a sign of selective forces acting on community assembly. A second main finding was that compositional and phylogenetic differences between sampling locations were classified as dispersal limited and drift—a sign of randomness on community assembly. The distance decay of similarity revealed apparent biogeochemistry at distances up to 1.5 km. It is thought that this decay of similarity reflects fracture water mixing. It was concluded that the spatial extent of the sampling locations precluded being able to relate selective process with environment. Decreasing the spatial extent of sampling locations might improve resolution to reveal environmental influences that have shaped the subsurface communities.

It was also found that bacterial taxa were numerically dominant in the groundwater. Although these were mainly uncultured, the closest cultivated representatives were from the phenotypically diverse Betaproteobacteria, Deltaproteobacteria, Bacteroidetes, Actinobacteria, Nitrospirae, and Firmicutes. Hundreds of taxa were identified but only a few were found in abundance (>1%) across all assemblages. The remainder of the taxa were low abundance.

1.4.3 Diversity of the 16S rRNA Gene across a 1 km^3 Subsurface Spatial Extent

Chapter 4 explored the diversity of the 16S rRNA gene at a smaller spatial extent—1 km^3 compared to 25 km^3. Taxa identities, abundances and phylogenetic relationships from the 16S rRNA gene V4 region within each sampling location (assemblages of local communities) were combined into a regional pool of taxa forming a meta-community. Focusing on phylogenetic diversity metrics, it was found that the component taxa within each community were more closely related than expected by chance. The random processes of dispersal limitation and drift accounted for up to 50% of the variance in beta diversity; with an associated distance decay of similarity of up to 0.5 km. Procrustes analyses for distance decay of similarity and three

phylogenic beta diversity metrics were significant for the spatial distances between sampling locations but not for the environmental distances between sampling locations. These metrics were sensitive to the tips of the 16S rRNA gene V4 region phylogenetic tree; a phylogenetic beta diversity metric more sensitive to the basal part of the tree was not significant with either the spatial or the environmental distances between sampling locations.

Data from the phylogenetic tree edge lengths revealed that 57% of the identified taxa were phylogenetic sister pairs—pairs of taxa whose edge lengths connected a common node within the V4 region phylogenetic tree were the same length. Within the ecological framework, this pattern may indicate diversification and thus co-existence of related taxa within the subsurface, however, given the variable copy number of the 16S rRNA gene, even at a 97% threshold for sequence similarity, intra-genomic heterogeneity in the V4 region of the gene may be higher than 3%; if this is the case, this finding may reflect the choice of using the 16S rRNA gene rather than an ecological process within the subsurface.

Another main finding from this analysis was that compositional and phylogenetic differences between sampling locations were classified as dispersal limited and drift—a sign of randomness on community assembly even at the smaller spatial extent of 1 km^3. The distance decay of similarity revealed apparent biogeochemistry at distances up to 0.5 km, suggesting that to be able to identify selective processes within this subsurface requires an even smaller spatial extent. For this reason, the meta-community was split into borehole specific biomes having a maximum distance between sampling locations of close to 500 m. These biomes were assessed (Chapter 5) for differential abundances linked to spatial and environmental factors.

1.4.4 Using the 16S rRNA Gene V4 region, Metagenome Data and Demonstrated Metabolism to Assess the Role of Selection in the Subsurface Community Assembly

Chapter 5 expanded on the analysis presented in Chapter 4 by focusing on selection as a driver of subsurface community assembly. Since randomness accounted for 50% of the variance in the phylogenetic beta diversity, sampling locations that clustered close together in the principle co-ordinate analysis of UniFrac distances (Chapter 4) were combined into separate biomes: these were the sampling locations accessed by borehole CR9 and by borehole CRG6. A BioEnv

analysis identified environmental and spatial data that co-varied with the taxa compositional data within each biome; these variables were then used as input variables to a generalized linear model designed to evaluate differential gene expression – except in this case, it was used to evaluate differential 16S rRNA gene abundances that. This analysis identified ~8% of the component taxa displayed differential abundances: within the CR9-biome, ~4% of the component taxa were identified as differentially abundant with respect to fracture water manganese and a spatial gradient; within the CRG6-biome, ~4% of the component taxa were identified as differentially abundant with respect to fracture water organic carbon and sulfate content. The outcome for the CRG6-biome analysis was consistent with demonstrated sulfate reduction by most probable number—an metabolism observed only during this sampling campaign.

A common metabolism across all sampling locations and for all sampling campaigns was nitrate reduction, albeit at 1% or less of the total cell count. Nitrogen compounds, however, where not measured within the fracture water. Data from rock core porewater did identify ammonia, nitrite and nitrate, suggesting the rock as a possible source of nitrogen compounds. Rock weathering as a driver of community assembly warrants further analysis.

1.5 References

Adler, P.B., Hillerislambers, J. and Levine, J.M., 2007, “A niche for neutrality”, Ecology Letters, 10: 95–104, doi:10.1111/j.1461-0248.2006.00996.x.

Cadotte, M.W., Tucker, C.M. 2017, Should Environmental Filtering be Abandoned?, Trends in Ecology and Evolution , 32 , 429 - 437.DOI: http://dx.doi.org/10.1016/j.tree.2017.03.004.

Chesson, P., 2000, “Mechanisms of maintenance of species diversity”, Annual Review of Ecology and Systematics, 31:1, 343-366.

Becking, L. G. M. Baas, I. R. Kaplan, and D. Moore, 1960, "Limits of the natural environment in terms of pH and oxidation-reduction potentials." The Journal of Geology 68, 243-284. http://www.jstor.org/stable/30059218.

Brock, T.D. , 'Milestones in Microbiology'. Translated and edited by Thomas D. Brock. Prentice-Hall, Englewood Cliffs, N.J., 1961.

Horner-Devine, M.C. and Bohannan, B.J.M., 2006, "Phylogenetic clustering and over dispersion in bacterial communities", Ecology, 87 S100–S108, 10.1890/0012-9658.

Hubbell, S. P. (2001). The Unified Neutral Theory of Biodiversity and Biogeography. Princeton, NJ: Princeton University Press.

Pedersen, K., "Microbial life in Deep Granitic Rock", 1997, FEMS Microbiology Reviews, 20, 3-4, 399–414, https://doi.org/10.1016/S0168-6445(97)00022-3.

Keddy, P.A., 1992, "Assembly and response rules: two goals for predictive community ecology", Journal of Vegetation Science, 3: 157–164, doi:10.2307/3235676.

King-Sharp, K.J., Frape, S.K., Peterman, Z., Gwynne, R., Tian, L., Gurban, I., 2016, "Synthesis of Geochemical and Fracture Mineral Studies Relevant to a Deep Geological Repository for Nonfuel Wastes at Chalk River", Canadian Nuclear Review, 10.12943/CNR.2016.00015.

Kraft, N. J. B., Adler, P. B., Godoy, O., James, E. C., Fuller, S. and Levine, J. M. (2015), Community assembly, coexistence and the environmental filtering metaphor. Functional Ecology, 29: 592–599. doi:10.1111/1365-2435.12345.

Lovley, D. R., and F. H. Chapelle, 1995, "Deep Subsurface Microbial Processes", Reviews in Geophysics., 33(3), 365–381, doi:10.1029/95RG01305.

Mayfield, M.M. and Levine, J.M., 2010, "Opposing effects of competitive exclusion on the phylogenetic structure of communities", Ecology Letters, 13, 1085-1093.

Nekola, J. C. and White, P. S. (1999), The distance decay of similarity in biogeography and ecology. Journal of Biogeography, 26: 867–878. doi:10.1046/j.1365-2699.1999.00305.x.

Neymark, L.A., Peterman, Z.E., Moscati, R.J., Thivierge, R.H., 2013, "U–Pb, Rb–Sr, and U-Series Isotope Geochemistry of Rocks and Fracture Minerals from the Chalk River Laboratories Site, Grenville Province, Ontario, Canada", Applied Geochemistry, 36, 10-33.

Nuclear Waste Management Office Backgrounder 2016: What Is Used Nuclear Fuel?

https://www.nwmo.ca/~/media/Site/Files/PDFs/2016/11/10/12/38/EN_Backgrounder_UsedNuclearFuel_LowRes.ashx?la=en.

Mullis, K.F., Faloona, F., Scharf, S., Saiki, R., Horn, G. Erlich, H. 1986, 'Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction'. Cold Spring Harbor Symposium in Quantitative Biology, 51:263–273.

O'Malley, M.A., 2007, "The nineteenth century roots of 'everything is everywhere'", Nature Reviews Microbiology, 5, 647-651,

O'Malley, M.A., 2008, "'Everything is everywhere: but the environment selects': ubiquitous distribution and ecological determinism in microbial biogeography", Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 39, 314-325.

Overballe-Petersen, S, Harms, K. Orlando, L.A.A, Moreno Mayar J.V.,Rasmussen, S, Dahl, T.W., Rosing, M.T. Poole, A.M., Sicheritz-Ponten, T., Brunak, S., Inselmann, S.de Vries, J., Wackernagel, W., Pybus, O.G., Nielsen R., Johnsen, P.J., Magne Nielsen K., Willerslev, E. 'Bacterial natural transformation by highly fragmented and damaged DNA', Proceedings of the National Academy of Sciences Dec 2013, 110 (49) 19860-19865; DOI: 10.1073/pnas.1315278110.

Ricklefs, R.E., 1987, "Community diversity: relative roles of local and regional processes", Science, 235(4785):167-71, DOI: 10.1126/science.235.4785.167.

Seaborn Panel report documenting the public review of Atomic Energy of Canada Limited Concept for Deep Geological Disposal of Used Nuclear Fuel, 1998, Canadian Environmental Assessment Agency, Catalogue No.: EN-106-30/1-1998E, ISBN: 0-662-26470-3, https://www.ceaa.gc.ca/default.asp?lang=En&n=0B83BD43-1

Stegen, J. C., Lin, X., Fredrickson, J. K., and Konopka, A. E. (2015). Estimating and mapping ecological processes influencing microbial community assembly. Front. Microbiol. 6:370. doi: 10.3389/fmicb.2015.00370

Vellend, M. (2010). Conceptual synthesis in community ecology, Quarterly Review of Biology, 85, 183–206. doi: 10.1086/652373.

Vellend, M., 2016, “The Theory of Ecological Communities”, Monograph in Population Biology, 57, Princeton University Press.

Zhou, J., Ning, D., 2017, ‘Stochastic Community Assembly: Does It Matter in Microbial Ecology?, Microbiology and Molecular Biology Reviews,81:e00002-17

2. SPATIAL AUTOCORRELATION, SOURCE WATER AND THE DISTRIBUTION OF TOTAL AND VIABLE MICROBIAL ABUNDANCES WITHIN A CRYSTALLINE FORMATION TO A DEPTH OF 800 M[3]

2.1 Abstract

Proposed radioactive waste repositories require long residence times within deep geological settings for which we have little knowledge of local or regional subsurface dynamics that could affect the transport of hazardous species over the period of radioactive decay. Given the role of microbial processes on element speciation and transport, knowledge and understanding of local microbial ecology within geological formations being considered as host formations can aid predictions for long term safety. In this relatively unexplored environment, sampling opportunities are few and opportunistic. We combined the data collected for geochemistry and microbial abundances from multiple sampling opportunities from within a proposed host formation and performed multivariate mixing and mass balance (M3) modeling, spatial analysis and generalized linear modeling to address whether recharge can explain how subsurface communities assemble within fracture water obtained from multiple saturated fractures accessed by boreholes drilled into the crystalline formation underlying the Chalk River Laboratories site (Deep River, ON, Canada). We found that three possible source waters, each of meteoric origin, explained 97% of the samples, these are: modern recharge, recharge from the period of the Laurentide ice sheet retreat (ca. ~12000 years before present) and a putative saline source assigned as Champlain Sea (also ca. 12000 years before present). The distributed microbial abundances and geochemistry provide a conceptual model of two distinct regions within the subsurface associated with bicarbonate – used as a proxy for modern recharge – and manganese; these regions occur at depths relevant to a proposed repository within the formation. At the scale of sampling, the associated spatial autocorrelation means that abundances linked with geochemistry were not unambiguously discerned, although fine scale Moran's eigenvector map (MEM) coefficients were correlated with the abundance data and suggest the action of localized processes possibly associated with the manganese and sulfate content of the fracture water.

2.2 Introduction

A goal of ecology is to relate population densities from within a region of interest to local or regional environmental conditions, however, analyses of spatially distributed sampling

[3] Adapted from:

E. D. Beaton, Marilyne Stuart, Sim Stroes-Gascoyne, Karen J. King-Sharp, Ioana Gurban, Amy Festarini, Hui Q. Chen, 2017, Spatial Autocorrelation, Source Water and the Distribution of Total and Viable Microbial Abundances within a Crystalline Formation to a Depth of 800 m, Frontiers in Microbiology, 8, 1731, DOI=10.3389/fmicb.2017.01731.

locations can be complicated by autocorrelation (Dormann et al., 2007; Gilbert and Bennet, 2010) or a lack of independence between nearby sampling locations. This characteristic, if not recognized, can lead to incorrect conclusions for population and environment interrelationships. When modeling population densities within a region of interest, autocorrelation can be caused by, for example, distance relationships in biological processes such as dispersal, by assuming an incorrect relationship between abundances and environment within a model, or by not accounting for an important environmental determinant that is, itself, spatially structured and thus causes spatial structuring in the response (Dormann et al., 2007). Discovery of distance-relationships associated with biological processes provides an important and interesting insight on community patterns while the assumptions made when modeling population abundances can lead to incorrect conclusions by having model residuals that are not randomly distributed, and so are themselves autocorrelated (Dormann et al., 2007).

Within the volume of proposed geologic repositories for hosting waste with inventories of long-lived radionuclides, information on the microbial abundances within an undisturbed setting at depth can help formulate conceptual models for long-term subsurface dynamics over the expected inventory decay period. A microbial community is defined as an assemblage of potentially interacting taxa that co-occur over space and time (Nemergut et al., 2013). Differences in abundances over space and time can occur through a combination of processes such as by abiotic selection and biotic competition or by speciation and drift between unconnected communities (Hubbell, 2001; Vellend, 2010). Microbial distributions in natural water systems also tend to be dispersed (Bliss and Fisher, 1953; El-Shaarawi et al., 1981; Haas and Heller, 1986; Hilbe, 2011; Harrison, 2014); occurring as clusters of cells or associated with suspended particles.

In this study, distributions of the total and viable count data and the geochemistry data were derived from sampling multiple saturated fractures that were accessed from boreholes drilled into overlapping bedrock assemblages underlying the Chalk River Laboratories (Deep River, ON, Canada) site. Data collection was part of a siting assessment for a potential future geologic waste management facility at the CRL site (Thompson et al., 2011). The locations of these boreholes are shown in Figure 2-1. Previous studies performed within this formation (Stroes-

Gascoyne et al., 2011; Beaton et al., 2016) showed that bacterial taxa were numerically dominant in the fracture water and that these bacteria displayed nitrogen metabolism with episodes of sulfur metabolism. This finding is akin to other crystalline subsurface environments hosting microbial communities that display metabolic activity such as nitrate, iron and sulfate reduction (Kieft, 1990; Jain et al., 1997; Haveman et al., 1999; Sahl et al., 2008; Nyyssönen et al., 2012). Although the bacteria were mainly uncultured, the closest cultivated representatives were from the phenotypically diverse *Betaproteobacteria*, *Deltaproteobacteria*, *Bacteroidetes*, *Actinobacteria*, *Nitrospirae*, and *Firmicutes*. Hundreds of taxa were identified but only a few were found in abundance (>1%) across all 16S rRNA assemblages. A decay of phylogenetic similarity with distance up 1.5 km was evident within sampling locations separated by up to 5 km of rock (Beaton et al., 2016). We propose that this decay distance is related to dispersal within vertical oriented fractures. To test for the possible influence of recharge and metabolism on total and viable abundances we extend our findings for nitrogen metabolism and sulfate reduction (Stroes-Gascoyne et al., 2011) and for the distance decay of similarity (Beaton et al., 2016) by analyzing the relative influences of the fracture water on microbial abundances and viability; an aspect of this subsurface habitat that had not been evaluated previously. Isotopic analysis of the dilute fracture water indicates it is of meteoric origin – with no significant rock-water interactions (King-Sharp et al., 2016); Supplementary Figure S1 shows the stable isotope composition for hydrogen and oxygen in the fracture water relative the Vienna Standard Mean Ocean Water (VSMOW). This recharge provides a possible source of soluble species for microbial processes and is a medium for dispersal. Porewater analysis from rock cores (Peterman et al., 2016) identified nitrogen compounds within the porewater composition that were not detected within the fracture water, so despite the stable isotope compositions relative to VSMOW (Supplementary Figure S1), rock-water interactions relevant to microbial abundances may still be occurring.

To gauge interrelationships between subsurface microbial abundances with the geochemistry we combined the abundance and geochemical data from multiple sampling opportunities and performed modeling to address whether recharge can explain how subsurface communities assemble within these fractures. The chemical species within the fracture water were evaluated

for their significance as explanatory variables by a multivariate approach in which the fracture water compositions were compared with the compositions of known and derived compositional end-members. The explanatory power of the end-member compositions provide insight into probable source waters and, therefore, insight into the history of recharge, mixing and other geological processes (Laaksoharju et al., 1999; Laaksoharju et al., 2008) that may have shaped the current fracture water compositions. Moran's I was used to determine spatial autocorrelation between sampling locations and the fracture water components were evaluated by a generalized linear model (GLM) (Venables and Ripley, 2002) for their significance as possible metabolic substrates associated with microbial abundances. Positive Moran's eigenvector map (MEM) coefficients were included as independent variables in the GLM to gauge for spatial autocorrelation within the model residuals.

2.3 Materials and Methods

2.3.1 Fracture Water Sampling and Analysis

Fracture water was collected using a Westbay Multilevel Groundwater Monitoring System (Nova Metrix Groundwater Monitoring (Canada) Ltd). Supplementary Figure S2 shows a schematic of a borehole with an installed Westbay System®. This Figure illustrates how the Westbay tubing and packers isolate multiple zones within the borehole thus preventing unnatural fracture water flow within the borehole itself. The tubing fluid is isolated from the formation fluid. In this arrangement, ambient formation fluid flow can pass through the annulus. From inside the tubing, formation fluid can be accessed by lowering a Westbay sampler and container assembly (also shown) to normally closed valved ports positioned between the packers. A larger schematic illustrates a deployed Westbay sampler assembly that is engaged at a selected port. Once the sampler is positioned and engaged, the remotely operated control valve in the sampler is opened to allow formation fluid from the zone to flow into the empty container. The process is monitored by observing changes in fluid pressure during the sequence of operations (see a typical trace of pressure vs. time in Supplementary Figure S2). Once the container is filled, the sampler valve is closed to seal the formation fluid inside the container at in situ pressure. The assembly is disengaged from the port (the port

valve automatically closes) and the fluid in the sealed container is retrieved to the surface for further handling.

The fracture water sampler consists of four 250 mL stainless steel tubes connected in series by tubing and Swagelok fittings. Prior to each sampling, the tubes were sterilized by autoclave and the fittings were sterilized by washing them with 70% ethanol. Validation of the sterilization and transport procedures was performed using sterilized water and PCR with bacterial rRNA 16S primers (Muyzer et al., 1993). Since the tube assemblies contacted only the interior of the casing surface, the probability of introducing surface microbes into the sampled volumes was minimal.

The borehole locations within the study site region of interest, and their names, are shown in Figure 2-1. These sampling locations are situated between the geological boundaries created by the Maskinonge Lake fault, the Mattawa fault (Ottawa River) and by East-West trending diabase dykes that traverse the study site along the boreholes CR-9, CRG-3 and CRG-6. Fracture water was collected from sealed borehole CRG-1, CRG-2, CRG-3, CRG-4A, CRG-6 and CR-9. Fracture water from an open unsealed borehole, CR-18, was also sampled. Depths of the sampled fracture water ranged from 35 to 780 m (137 to -800 m elevation, relative to sea level).

The fracture water pH [Beckman PHI 265 pH/Temp/mV meter (Beckman Coulter, Inc.)] and conductivity [YSI Model 30 Conductivity Meter (YSI Inc., Yellow Springs, OH, United States) were measured and 100 mL aliquots were filtered through a 0.45 µm filter (isopore polycarbonate, Millipore, Billerica, MA, United States) then immediately preserved by adding 0.5 mL nitric acid (ultra-trace grade, SeastarTM, Baseline®, Fisher Scientific, Ottawa, ON, Canada) for elemental analysis. Elemental composition of the fracture water was determined by inductively coupled plasma-mass spectrometry (ICP-MS, using either a Varian 820-MS (Agilent Technologies, Inc.) or an Element XR (Thermo Scientific)) and by inductively coupled plasma atomic emission spectroscopy (ICP-AES, Optima 3300, Perkin Elmer). Anion concentrations were determined using a Dionex 3000 ICS ion chromatograph (Dionex, Sunnyvale, CA, United States). Dissolved organic (DOC) and inorganic carbon (DIC) were determined using a Dohrmann, model Phoenix 8000-UV Persulfate TOC Analyzer (Teledyne Teckmar, Mason, OH, United States).

Total and viable microbial densities were determined by fluorescence microscopy with a Nikon E600 microscope and a Zeiss Axiophot microscope after filtering the separate stained samples onto black polycarbonate filters (Fisher Scientific, 25 mm, 0.22 µm pore size); at least fifteen fields of view and at least 300 cells were counted per filter for a coefficient of variation of 5.8% per filter. Direct counts for total cell densities were determined in triplicate 1 mL volumes – within 4 h of sampling at the formation pressure and within 1 h of opening the sample tubes including a 30 min incubation time. Total cell densities were determined using the DNA intercalating dye, Sybr Green II (Life Technologies); because separate aliquots were shipped to another laboratory, total cell counts were also determined within 24 h of opening the sample tubes, in this case, using Acridine Orange (Sigma–Aldrich) to emulate the procedure employed at the receiving laboratory; the two dyes and two time points gave similar results. Direct counts for viable cell densities were determined in triplicate 1 mL volumes, also within 1 h of opening the sampling tubes, using dyes that are sensitive to different characteristics of viable microbial cells: the soluble 5-cyano-2,3-ditolyl tetrazolium chloride (CTC, Sigma–Aldrich) was used to evaluate respiratory activity within the microbial population as detected by the reduction of CTC to the insoluble fluorescent CTC-formazan (Schaule et al., 1993); the lipophilic cation, rhodamine-123 (R123 Sigma–Aldrich) (McFeters et al., 1998; Fuller et al., 2000) was used to evaluate cells within the microbial population that display a membrane potential difference; and carboxyfluorescein diacetate (CFDA, Sigma–Aldrich) was used to evaluate enzymatic activity (Schaule et al., 1993).

2.3.2 Multivariate Mixing and Mass Balance Analysis

The fracture water components sodium (Na^+), calcium (Ca^{2+}), magnesium (Mg^{2+}), bicarbonate (HCO_3^-), chloride (Cl^-), sulfate (SO_4^{2-}) and the isotopes tritium, deuterium and stable oxygen, $\delta^{18}O$, were used as input data for the multivariate, mixing and mass balance analysis model (M3, performed by 3D Terra, Montreal, Quebec). The M3 model consists of four steps: a principal component analysis (PCA); selection of reference waters (end-members) followed by

calculations of mixing proportions; and finally mass balance calculations (Laaksoharju et al., 1999; Laaksoharju et al., 2008).

Three end-members were found to describe the fracture water; these are referred to as: (1) 'recharge,' (2) 'Champlain Sea' (or 'saline'), and (3) 'glacial melt' (not shown). The stable isotope values for the melt water end member were taken from the literature: the $\delta^{18}O$ value from Frape and Fritz (1987) and Remenda et al. (1994). The deuterium value was determined by Rozanski et al. (1993). The tritium value, which governs the proportion of recharge, was considered decayed to zero. The end-member referred to as 'Champlain Sea' was obtained from nearby sediment pore water from this period in the site history that had a salinity of 6.1% (Torrance, 1988). The end-member referred to as modern recharge was calculated as an average of the chemistry of the upper section of boreholes CRG-2, CRG-3, CRG-4A and CRG-6-1 and CR-9-1. The software, Surfer (Golden Software), was used to create 2D cross section maps.

2.3.3 Generalized Linear Model with a Negative Binomial Distribution

The replicate values for microbial cell densities and geochemistry were averaged for each borehole interval sampling location. Supplementary Figure S3 shows a comparison of quantile-by-quantile plots for the total cell count distribution against theoretical normal and negative binomial distributions. The environmental and spatial data were evaluated as explanatory variables using the glm.nb() function from the R package 'MASS' (Venables and Ripley, 2002). Significant variables were determined by stepwise modeling. Model selection was based on minimizing the Akaike information criterion (AIC). Analysis of variance was applied to the reduced model to determine the significance of the retained variables. Only those values with $p < 0.05$ were considered significant. All the model results are provided in an Excel file in the Supplemental Information.

2.3.4 Spatial Autocorrelation and Moran's Eigenvector Map Coefficients

Moran's eigenvector map were created by principle coordinates of neighbor matrices (Borcard and Legendre, 2002; Dray et al., 2006) from within the R packages 'sdep' and 'adespatial'. A matrix of spatial eigenvectors was built from a distance matrix of Easting and Northing, zone 18, Universal Transverse Mercator coordinates for each borehole interval. The functions used to create the spatial weightings matrix were nbtri(), that converts the spatial coordinates of the

sampling locations into a distance neighbors map, and the function nb2listw() that creates the weightings matrix from the neighbors map. The eigenvectors for positive values for Moran's I reveal different spatial structures over the entire range of scales encompassed by the geographical sampling area. The first MEM values generated in the analyses represent broader spatial structures, and the last MEM values represent finer spatial structures. Values for Moran's I at each sampling location were compared to a null distribution of the global Moran's I using the function localmoran(). The resulting z-values were plotted to display locations with spatial correlations that were more than two standard deviations from the null mean.

2.4 Results

2.4.1 Multivariate Mixing and Mass Balance (M3) Modeling

The results from the PCA are shown in Figure 2-2. The PCA results are displayed three times to illustrate modeling results for mixing of the three fracture water compositional end-members; these were the percent mixing proportion for glacial melt water (Figure 2-2, upper left panel), the percent mixing proportion for Champlain Sea (Figure 2-2, upper right panel), and the percent mixing proportion for modern recharge (Figure 2-2, bottom panel). The first and second principal components accounted for 71% of the variance in the geochemistry. The area encompassed by the three end-members (Figure 2-2, triangle joining the compositional end-members) explains over 97% of the fracture water samples; most of the individual fracture water compositions plot between the three reference waters. The fracture water samples that plot outside the region of the three end members are listed as open circles (Figure 2-2, all three panels). The explanatory power of glacial melt, Champlain Sea and modern recharge may indicate that these waters have affected the present fracture water composition and thus represent historical events that could have influenced the fracture water microbial populations. The modeled mixing proportions of the three source waters suggests that fracture water sampled from boreholes CRG-1, CRG-2, CRG-3, CRG-4A, CRG-6, and CR-18 contain mainly modern recharge with a small glacial melt mixing proportion of up to ~40%. Fracture water sampled from borehole CR-9 includes proportions from these source waters and an additional mixing proportion from a saline water source, referred to here as Champlain Sea. Fracture water accessed from intervals 11 and 12 from borehole CR-9 have mainly a saline water type

signature of ~70%. By this model, the fracture water from CR-9-3, CR-9-8 and CR-18 is a mixture of Champlain Sea, melt water and modern water. Distributions of the three possible water sources is represented in cross section in Figure 2-3 (left panel) for modern recharge (top left panel), glacial melt (middle left panel) and Champlain Sea (bottom left panel). The depth and orientation profiles for boreholes CR-9, CRG-1 and CRG-3 are also shown. These boreholes form an East-West transect across the study site (depicted by the shaded area in Figure 2-1); the sampling locations within these boreholes that were used for microbial abundance determinations are shown in Figure 2-3 as white dots. The visualizations were created by 2-D kriging interpolating between the sampling locations within each of these boreholes and do not account for the fractures that would provide the water flow paths throughout the rock matrix. The left-hand side of Figure 2-3 shows the mixing proportion by prospective source water and the right-hand side of the Figure 2-3 shows the distributions of geochemical signatures that correspond to these water sources: bicarbonate (Figure 2-3, top right panel) for modern recharge; measured $\delta^{18}O$ values (Figure 2-3, middle right panel) for glacial melt water and chloride (Figure 2-3, bottom right panel) for a saline source water. These components of the fracture water, therefore, may represent a signature for source water in a GLM.

2.4.2 Spatial Autocorrelation

Spatial autocorrelation refers to similarities in attributes between adjacent locations compared to the attributes between more distant locations (Miller, 2004). Spatial autocorrelation in abundance data can be informative of processes that drive community patterns. Spatial autocorrelation in model residuals, however, can lead to incorrect interpretation of the processes that drive community patterns. To test for spatial autocorrelation within the sampled fracture water, MEM coefficients were calculated and those coefficients associated with positive Moran's I were added to the GLM as independent variables. These coefficients may represent unknown processes occurring locally within the projected area. Local values for Moran's I were also compared with a null distribution of the global Moran's I to identify attributes at sampling locations (for example cell counts) with Moran's I values that were more than two standard deviations from the null mean. A local Moran's I for an attribute that is more than two standard deviations from the null mean in the positive direction indicates that the

spatial distribution of that attribute is more clustered than would be expected if underlying spatial processes were random; in this case, the null hypothesis of random distribution of a given 'attribute' would be rejected. A local Moran's I for an attribute that is more than two standard deviations from the null mean in the negative direction indicates the spatial distribution of high and low values for that attribute were more spatially dispersed than would be expected if underlying spatial processes were random; in this case, the null hypothesis would also be rejected.

The z-values calculated for the distribution of the various 'attributes' – namely, the cell count densities, concentrations of soluble compounds, pH and the positive MEM coefficients – are shown in Figure 2-4; the sampling locations are listed by borehole and interval following a West-to-East direction from borehole CR-18 to borehole CRG-2 (see Figure 2-1). The dashed gray lines and the solid gray lines mark where the first and second standard deviations from the null mean lie. The bars for attributes that extend beyond the mark for the second standard deviation, in the positive or negative direction, identify the sampling locations with spatially non-random attributes. From the plots in Figure 2-4, the deeper sampling locations within borehole CR-9 at intervals 8, 11, and 12, display non-random attributes relative to the global distribution of cell counts, or clustering as lower total cell counts; lower bicarbonate concentrations and higher sulfate and manganese concentrations; and by the MEM coefficients labeled MEM5, MEM7 and MEM10. These intervals are also the sampling locations with a saline signature (Figures 2, 3, bottom left panel); even so, the chloride was not identified as being spatially autocorrelated.

The total cell counts from the shallow sampling location within the same borehole, located at interval 2, is dispersed compared to the null mean of spatially distributed counts values; and the MEM coefficient, MEM1. Only two other sampling locations had cell count densities that were outside the bounds of the null spatial distribution: interval 7 of borehole CRG-4A and interval 14 of borehole CRG-6. For these locations, interval 7 of borehole CRG-4A is elevated with respect to: the CTC positive counts; the concentration of bicarbonate; and the MEM coefficients, MEM2 and MEM5. Interval 14 of borehole CRG-6 is reduced with respect to: R123 positive counts; and the MEM coefficient MEM4. These deviations from the global Moran's I for

all the sampling locations suggest there are local processes that influencing the microbial abundances.

2.4.3 Distribution of the Count Data

Values for total and viable cell densities in fracture water sampled from each borehole are provided in Supplementary Table S1 and Table S2. How the cell densities distribute across the sampling locations is shown in Figure 2-5 as histograms and as boxplots by borehole arranged in a West to East direction (from borehole CR-18 to borehole CRG-2 as shown in Figure 2-1). The cell densities within boreholes CR-18, CR-9, CRG-1 and CRG-2 form the lower density part of the histograms and the cell densities within boreholes CRG-3, CRG-6 and CRG-4A form the higher density part of the histograms. The same data is plotted as scatter plots by sampling location elevation relative to sea level (Supplementary Figure S4, left panel) and by the Longitude value for the borehole collar, were the borehole enters the subsurface (Supplementary Figure S4, right panel; these two figures show that borehole location show a wider range of cell densities than does elevation.

To help identify possible drivers for the microbial abundances, the distribution patterns for the total and viable cell densities were also compared with the distributions of the fracture water geochemistry (using data taken from Supplementary Table S1 and Table S2) and to the rock porewater components: sulfate, bicarbonate, ammonia, nitrate and nitrite (from Peterman et al., 2016). The resulting quantile-by-quantile plots are shown in Supplementary Figures S5–S8 and in Figure 2-5 beside the histograms for the cell densities. Quantile-by-quantile plots allow for the distribution patterns between two datasets to be compared; if the datasets follow a similar distribution the data points plot along a straight line; if the datasets do not follow a similar distribution pattern, the points diverge from a straight line. We find from these comparisons that the microbial cell densities distribute within the subsurface like that for the fracture water and porewater sulfate, the porewater ammonia and for the fracture water manganese, but not for the fracture water bicarbonate or for the porewater bicarbonate. Supplementary Figure S5 show the cell density distributions with the fracture water and porewater sulfate. The total cell count data appear to have a distribution like that for the fracture water sulfate while the CTC and R123 cell density distributions deviate from the

straight line (Supplementary Figure S5, top panel). The opposite patterns are seen for porewater sulfate; the distributions for total cell count and the porewater sulfate deviate from a straight line while the CTC and R123 cell densities appears to have distributions like that for the porewater sulfate (Supplementary Figure S5, bottom panel).

Comparisons with the distributions of fracture water and porewater bicarbonate (Supplementary Figure S6) and of fracture water manganese (Supplementary Figure S7) show that the count data distributions are not like that for bicarbonate in the lower quantiles (Supplementary Figure S6, top panel for fracture water, bottom panel for porewater) but they are distributed that for like manganese (Supplementary Figure S7).

Supplementary Figure S8 show the quantile-by-quantile plots for the distributions of cell density the porewater nitrogen compounds: ammonia, nitrite and nitrate. These components of the porewater were not detected within the fracture water. The plot shows that the distributions for total and CTC cell densities are roughly linear with the distribution for ammonia (Supplementary Figure S8, top panel). The total, CTC and R123 cell densities are also roughly linear with the distribution of nitrate concentrations but there is flattening in the middle quantiles for nitrate. The plot comparing the cell densities with porewater nitrite suggest these datasets follow different distributions across the formation (Supplementary Figure S8, bottom panel).

2.4.4 Generalized Linear Modeling of the Count Data

The geochemical and descriptive data used for the GLM are given in Supplementary Table S2. The negative binomial GLM function within the R package 'MASS' (Venables and Ripley, 2002) provides a model to assign linear predictors (β) and a description of the random error distribution of the count data. The total and viable count data across all the sampling locations were each used as response variables. Data for the geochemical and positive MEM coefficients across all sampling locations were used as the dependent variables. Unmeasured environmental variables associated with the microbial cell density distribution would form part of the random component of the resulting linear model.

The independent variables were evaluated first for model selection then stepwise model fitting was performed. Metabolically relevant components of the geochemistry were: pH; dissolved

organic carbon (DOC); bicarbonate; sulfate; iron; manganese; and phosphate. Bicarbonate ion is also a possible signature for modern recharge (Figure 2-3). Data for chloride ion were included as an explanatory variable for a saline source water component, and the stable oxygen isotope ($\delta^{18}O$) data were included as an explanatory variable for a glacial melt source water component. The spatial weightings matrix identified 11 positive MEM coefficients; these were also included in a model. The resulting coefficients (β) and the 5% confidence interval values for the significant explanatory variables are listed in Supplementary Table S2 (total counts), Supplementary Table S3 (CTC counts), Supplementary Table S4 (R123 counts) and Supplementary Table S5 (CFDA counts). Only those explanatory variables with a significance of $p < 0.05$ or lower are shown.

When the models were run with the geochemistry – without the positive MEM coefficients – bicarbonate and manganese were identified as the predictors of total (Supplementary Table S2), CTC (Supplementary Table S3) and R123 (Supplementary Table S4) cell counts. When the model was run with the positive MEM coefficients – without the geochemistry – between two and four coefficients were significant: MEM2 and MEM4 were identified for both total and viable cell counts; and either MEM1, MEM5, MEM7 or MEM10 were identified depending on the count data (Supplementary Tables S2–S5). An analysis of the model residuals for the total count data are provided in Figure 2-6: environmental variables (Figure 2-6A), the positive MEM coefficients (Figure 2-6B) and the measured variables and spatial coefficients (Figure 2-6C). These plots show that the distribution of the residuals and fitted total count are more randomly distributed when the positive MEM coefficients are included in the model (Figure 2-6B) than when the model included only environment (Figure 2-6A). Combining spatial and environmental inputs did not improve the distribution of the model residuals (Figure 2-6C).

2.5 Discussion

The concept of geological radioactive waste repositories is to provide secure locations over long residence times within deep geological settings allowing for the decay of long-lived radionuclides to background levels. The feasibility of emplacing a repository relies on having knowledge of local and regional subsurface dynamics that would form the basis of predicting the transport of hazardous species from a repository over the period of radioactive decay.

Given the role of microbes on element speciation and transport and their effect on element retention by microbial-derived metal oxides (Kennedy et al., 2011), knowledge and understanding of local microbial ecology within prospective host formations can aid predictions used for making long term safety cases. To gauge interrelationships between subsurface microbial abundances with the geochemical data we combined the abundance data obtained from multiple sampling opportunities from within the crystalline formation underlying the Chalk River Laboratories site (Deep River, ON, Canada) and performed multivariate mixing and mass balance (M3) modeling, spatial analysis and GLM. We considered the dual role of fracture water – as the medium for transport of soluble species suitable for microbial metabolism and its role as the medium for dispersal.

In our analyses, we identified possible sources of the fracture water and evaluated the fracture water composition as predictors of total and viable microbial cell densities within these fractures including their distribution patterns. The dilute character of the fracture water (King-Sharp et al., 2016) compared to other sites on the Canadian Shield is thought to reflect recharge that occurred at the end of the last glaciation followed by a gradual recharge with meteoric water. The fracture water ages date from 5000 to 10,000 years (King-Sharp et al., 2016). Modern recharge does not appear to extend deeper than approximately 100 m (King-Sharp et al., 2016). We therefore performed modeling to address whether recharge can explain how subsurface communities assemble within these fractures. The main findings from M3 modeling were that three possible meteoric source waters account for 97% of the samples: glacial melt water, a saline source and modern recharge. The mixing proportion for modern recharge and glacial melt water describe most of the samples; the mixing proportion of a saline source water is localized to deeper fractures transected by boreholes CR-9 (Figure 2-3). Although the stable isotope data for oxygen and hydrogen align with the VSMOW (Supplementary Figure S1), supporting the notion of recharge as a main driver of microbial assembly, rock water interactions may still be important in explaining the fracture water microbiology. A porewater analysis of the drilled rock cores identified nitrogen compounds (Peterman et al., 2016) that were not detected within the fracture water; a finding that corroborates both the measured nitrogen metabolism within the fracture water (Stroes-

Gascoyne et al., 2011) and the identified taxa within the fracture water (Beaton et al., 2016) whose cultured relatives encompass the complete nitrogen cycle, including nitrogen fixation. In a study of the component taxa (manuscript in preparation), nitrogen metabolism was detected within all sampling locations; sulfate reduction was detected only within borehole CRG-6.

Recharge into fractures is topology driven. An analysis of the cell density distribution patterns identified location specific patterns and patterns that were generalized across the sampling locations. The total and viable cell densities fell into two categories: those locations with lower cell densities (location within boreholes CR-9, CR-18, CRG-1, CRG-2) (Figure 2-5 and Supplementary Figure S4) and those locations with higher cell densities (locations within boreholes CRG-3, CRG-4A and CRG-6). The sampling locations with the lowest total cell counts were those locations that had the highest mixing proportions of a saline source water (Figure 2-2) and the sampling locations with the highest total cell densities were those locations with higher mixing proportions of modern recharge (Figure 2-2). Variation in the count data appears to be localized to the region around each of the boreholes and not to the elevation of the sampling locations (Supplementary Figure S4). If the abundance data can be linked to recharge, the influences of local conditions on recharge may need to be accounted for by, for example, overburden thickness and hydraulic conductivity. In a study of modern recharge into another fractured crystalline aquifer that is overlain by variable thicknesses of overburden (Gleeson et al., 2009), the authors conclude that overburden thickness and hydraulic conductivity were major parameters that controlled modern recharge into the underlying bedrock aquifer and that a thicker overburden meant modern recharge was slower and more widespread (Gleeson et al., 2009). A slower recharge rate and higher surface area of unconsolidated overburden would favor higher cell densities.

The distributions of the abundances and the geochemistry were more generalized across the site. The quantile-by-quantile plots show that the total cell count distributions aligned with the distributions of fracture water sulfate, fracture water manganese, porewater sulfate, porewater ammonia and porewater nitrate (Figure 2-5 and Supplementary Figures S5, S7, S8). The viable cell count distributions also aligned with the distributions of sulfate, manganese, ammonia and nitrate while their distributions compared with bicarbonate were heavy tailed in the lower

quantiles – namely, those sampling locations within borehole CR-9; a region of the subsurface found to be distinct in the M3 modeling (Figure 2-3).

Analysis of spatially distributed sampling locations can reveal distance-relationships in abundance data (Dormann et al., 2007). The assumptions made when modeling population abundances can lead to incorrect conclusions if the model residuals are not randomly distributed (Dormann et al., 2007). We therefore performed a spatial analysis to test for a role for meteoric water recharge on total and viable abundances by comparing a null distribution of the global Moran's I value (Figure 2-4) and by adding the resulting eigenvector map coefficients into a GLM (Figure 2-6). The GLM identified bicarbonate and manganese as significant predictors of microbial abundances (Supplementary Tables S2–S5). Both bicarbonate and manganese also show spatial autocorrelation at sampling locations within borehole CR-9; as does sulfate (Figure 2-4). In our analysis, bicarbonate was considered as a proxy for modern recharge; the proxy for a saline source recharge, chloride, was not identified as significant (Supplementary Tables S2–S5) and, despite the M3 modeling showing the localization of this saline signature (Figures 2, 3), chloride was not spatially clustered with the bicarbonate, manganese and sulfate (Figure 2-4).

The GLM also identified positive MEM coefficients of which four coefficients clustered within borehole CR-9 and two coefficients, MEM2 and MEM4, were randomly distributed. The improved GLM residuals with these coefficients suggest that the significance of the bicarbonate and manganese was due to the localized and distinct fracture water conditions that exist within borehole CR-9 and further suggests that their significance in the GLM reflects this spatial correlation. Inclusion of the MEM coefficients within the GLM improved the distribution of the GLM residuals. The finer scale influences represented by these coefficients may indicate unmeasured/unknown processes; the distribution pattern similarities observed with the quantile-by-quantile plots may help reconcile these processes.

2.6 Conclusion

The main findings of this work are that M3 modeling identified three possible meteoric sources water for recharge; of these three, modern recharge appears to be the most likely source water to explain, in part, microbial abundances within the projected area of the sampling locations.

Chloride, as a proxy for a saline source water, was not a significant explanatory variable for the total or viable count data. Stable oxygen isotope ($\delta^{18}O$), as a gauge of glacial melt water, was also not a significant explanatory variable of microbial abundance distributions.

Spatial autocorrelation analysis show that low total cell counts co-localize with lower bicarbonate, higher manganese and higher sulfate. These locations are associated with the saline source water signatures. The spatial correlation of both the bicarbonate and the manganese suggest that their significance in the GLM reflects this spatial correlation and not a direct effect on microbial abundances, per se. Inclusion of positive MEM coefficients into the GLM improved the distribution of the model residuals. The finer scale influences represented by the significant MEM coefficients suggest there are unmeasured/unknown processes occurring within these sampling locations.

While the fracture water is dilute, and of mainly meteoric origin (King-Sharp et al., 2016), the prospect of porewater sulfur and porewater nitrogen (Peterman et al., 2016) potentially leaching from the host rock suggest there may be localized processes that are separate from a role of source water recharge in explaining microbial abundance distributions within the projected area of the sampling locations.

2.7 Author Contributions

All authors collected and contributed data sets for analysis as well as participated in the conceptual drafting and revision of this manuscript. KK-S oversaw the groundwater sampling. IG performed modelling of the geochemistry. EB conducted subsequent analyses and was the primary author in writing and revising the manuscript. SS-G, KK-S, and MS contributed significantly in manuscript development and revision.

2.8 Funding

This work was supported by Natural Resources Canada through the Canadian Nuclear Laboratories Nuclear Legacy Liability Program under which the Geological Waste Management Facility (GWMF) project was funded.

2.9 Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

2.10 Acknowledgments

We thank Vinicus Anghel, Canadian Nuclear Laboratories, for his helpful review of this manuscript and David Larssen, Westbay Instruments, Vancouver, British Columbia, for providing the image and description of the fluid sampling system. Funding of this work was provided by Natural Resources Canada Ltd.

2.11 References

Beaton, E. D., Stevenson, B. S., King-Sharp, K. J., Stamps, B. W., Nunn, H. S., and Stuart, M. (2016). Local and regional diversity reveals dispersal limitation and drift as drivers for groundwater bacterial communities from a fractured granite formation. Front. Microbiol. 7:1933. doi: 10.3389/fmicb.2016.01933

Bliss, C. I., and Fisher, R. A. (1953). Fitting the negative binomial distribution to biological data. Biometrics 9, 176–200. doi: 10.2307/3001850

Borcard, D., and Legendre, P. (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecol. Model. 153, 51–68. doi: 10.1016/S0304-3800(01)00501-4

Dormann, C. F., McPherson, J., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., et al. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30, 609–628. doi: 10.1111/j.2007.0906-7590.05171.x

Dray, S., Legendre, P., and Peres-Neto, P. R. (2006). Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecol. Model. 196, 483–493. doi: 10.1016/j.ecolmodel.2006.02.015

El-Shaarawi, A. H., Esterby, S. R., and Dutka, B. J. (1981). Bacterial density in water determined by Poisson or negative binomial distributions. Appl. Environ. Microbiol. 41, 107–116.

Frape, S. K., and Fritz, P. (eds) (1987). “Geochemical trends from groundwaters from the Canadian Shield,” in saline Water and Gases in Crystalline Rocks, Special Paper 33 (St. John’s, NL: Geological Association of Canada), 19–38.

Fuller, M. E., Streger, S. H., Rothmel, R. K., Mailloux, B. J., Hall, J. A., Onstott, T. C., et al. (2000). Development of a vital fluorescent staining method for monitoring bacterial transport in subsurface environments. Appl. Environ. Microbiol. 66, 4486–4496. doi: 10.1128/AEM.66.10.4486-4496.2000

Gilbert, B., and Bennet, J. R. (2010). Partitioning variation in ecological communities: do the numbers add up? J. Appl. Ecol. 47, 1071–1082. doi: 10.1111/j.1365-2664.2010.01861.x

Gleeson, T., Novakowski, K., and Kurt Kyser, T. (2009). Extremely rapid and localized recharge to a fractured rock aquifer. J. Hydrol. 376, 496–509. doi: 10.1016/j.jhydrol.2009.07.056

Haas, C. N., and Heller, B. (1986). Statistics of enumerating total coliforms in water samples by membrane filter procedures. Water Res. 20, 525–530. doi: 10.1016/0043-1354(86)90203-4

Harrison, X. A. (2014). Using observation-level random effects to model overdispersion in count data in ecology and evolution. PeerJ 2:e616. doi: 10.7717/peerj.616

Haveman, S. A., Pedersen, K., and Ruotsalainen, P. (1999). Distribution and metabolic diversity of microorganisms in deep igneous rock aquifers of Finland. Geomicrobiol. J. 16, 277–294. doi: 10.1080/014904599270541

Hilbe, J. M. (2011). Negative Binomial Regression, Second Edn. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511973420

Hubbell, S. P. (2001). The Unified Neutral Theory of Biodiversity and Biogeography. Princeton, NJ: Princeton University Press.

Jain, D. K., Providenti, M., Tanner, C., Cord, I., and Stroes-Gascoyne, S. (1997). Characterization of microbial communities in deep groundwater from granitic rock. Can. J. Microbiol 43, 272–283. doi: 10.1139/m97-038

Kennedy, C. B., Gault, A. G., Fortin, D., Clark, I. D. F. G., and Ferris, F. G. (2011). Retention of iodide by bacteriogenic iron oxides. Geomicrobiol. J. 28, 387–395. doi: 10.1080/01490451003653110

Full Text | Google Scholar

Kieft, T. L. (1990). Environmental Parameters Controlling Microbial Activities in Terrestrial Subsurface Environments. Washington, DC: US Department of Energy.

King-Sharp, K., Frape, S. K., Peterman, Z. E., Gwynne, R., Tian, L., and Gurban, I. (2016). Synthesis of geochemical and fracture mineral studies relevant to a deep geological repository for non-fuel wastes at Chalk River. CNL Nuclear Rev. 6, 117–130. doi: 10.12943/CNR.2016.00015

Laaksoharju, M., Gascoyne, M., and Gurban, I. (2008). Understanding groundwater chemistry using mixing models. Appl. Geochem. 23, 1921–1940. doi: 10.1016/j.apgeochem.2008.02.018

Laaksoharju, M., Skårman, C., and Skårman, E. (1999). Multivariate mixing and mass balance (M3) calculations, a new tool for decoding hydrogeochemical information. Appl. Geochem. 14, 861–871. doi: 10.1016/S0883-2927(99)00024-4

McFeters, G. A., Pyle, B. H., Lisle, J. T., and Broadaway, S. C. (1998). Rapid direct methods for enumeration of specific, active bacteria in water and biofilms. J. Appl. Microbiol. 85, 193S–200S. doi: 10.1111/j.1365-2672.1998.tb05299.x

Miller, H. J. (2004). Tober's first law and spatial analysis. Ann. Assoc. Am. Geogr. 94, 284–289. doi: 10.1111/j.1467-8306.2004.09402005.x

Muyzer, G., de Waal, E. C., and Uitterlinden, A. G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl. Environm. Microbiol. 59, 695–700.

Nemergut, D. R., Schmidt, S. K., Fukami, T., O'Neill, S. P., Bilinski, T. M., Stanish, L. F., et al. (2013). Patterns and processes of microbial community assembly. Microbiol. Mol. Biol. Rev. 7, 342–356. doi: 10.1128/MMBR.00051-12

Nyyssönen, M., Bomberg, M., Kapanen, A., Nousiainen, A., Pitkänen, P., and Itävaara, M. (2012). Methanogenic and sulphate-reducing microbial communities in deep groundwater of crystalline rock fractures in Olkiluoto, Finland. Geomicrobiol. J. 29, 863–878. doi: 10.1080/01490451.2011.635759

Peterman, Z. E., Neymark, L. A., King-Sharp, K. J., and Gascoyne, M. (2016). Isotope hydrology of the Chalk River Laboratories site, Ontario, Canada. Appl. Geochem. 66, 149–161. doi: 10.1016/j.jenvrad.2014.07.007

Remenda, V. H., Cherry, J. A., and Edwards, T. W. (1994). Isotopic composition of old ground water from Lake Agassiz: implications for late Pleistocene climate. Science 266, 1975–1978. doi: 10.1126/science.266.5193.1975

Rozanski, K., Araguás-Araguás, L., and Gonfiantini, R. (1993). "Isotopic patterns in modern global precipitation," in Climate Change in Continental Isotopic Records, eds P. K. Swart, K. C. Lohmann, J. McKenzie, and S. Savin (Washington, DC: American Geophysical Union).

Sahl, J. W., Schmidt, R., Swanner, E. D., Mandernack, K. W., Templeton, A. S., Kieft, T. L., et al. (2008). Subsurface microbial diversity in deep-granitic-fracture water in Colorado. Appl. Environ. Microbiol. 74, 143–152. doi: 10.1128/AEM.01133-07

Schaule, G., Flemming, H. C., and Ridgway, H. F. (1993). Use of 5-cyano-2,3-ditolyl tetrazolium chloride for quantifying planktonic and sessile respiring bacteria in drinking water. Appl. Environ. Microbiol. 59, 3850–3857.

Stroes-Gascoyne, S., Hamon, C. J., Audette-Stuart, M., Beaton, E. D., King-Sharp, K., Festarini, A., et al. (2011). Microbial Characterization of Groundwater from Boreholes CR9 and CR18 at CRL (2007–2009) – Implications for a Possible Future Repository for Radioactive Non-Fuel Waste. Toronto, ON: Canadian Nuclear Society.

Thompson, P., Baumgartner, P., Beaton, E. D., Chan, T., Kitson, C., Kozak, E., et al. (2011). An Investigation of the Suitability of the Chalk River Site to Host a Geologic Waste Management Facility for AECL's Low and Intermediate Level Wastes. Toronto, ON: Canadian Nuclear Society.

Torrance, J. K. (1988). "Mineralogy, pore-water chemistry and geotechnical behaviour of Champlain Sea and related sediments," in The Late Quaternary Development of the Champlain Sea Basin, ed. N. R. Gadd (St. John's, NL: Geological association of Canada).

Vellend, M. (2010). Conceptual synthesis in community ecology. Q. Rev. Biol. 85, 183–206. doi: 10.1086/652373

Venables, W. N., and Ripley, B. D. (2002). odern Applied Statistics with S. New York, NY: Springer Science Business Media.

2.12 List of Figures

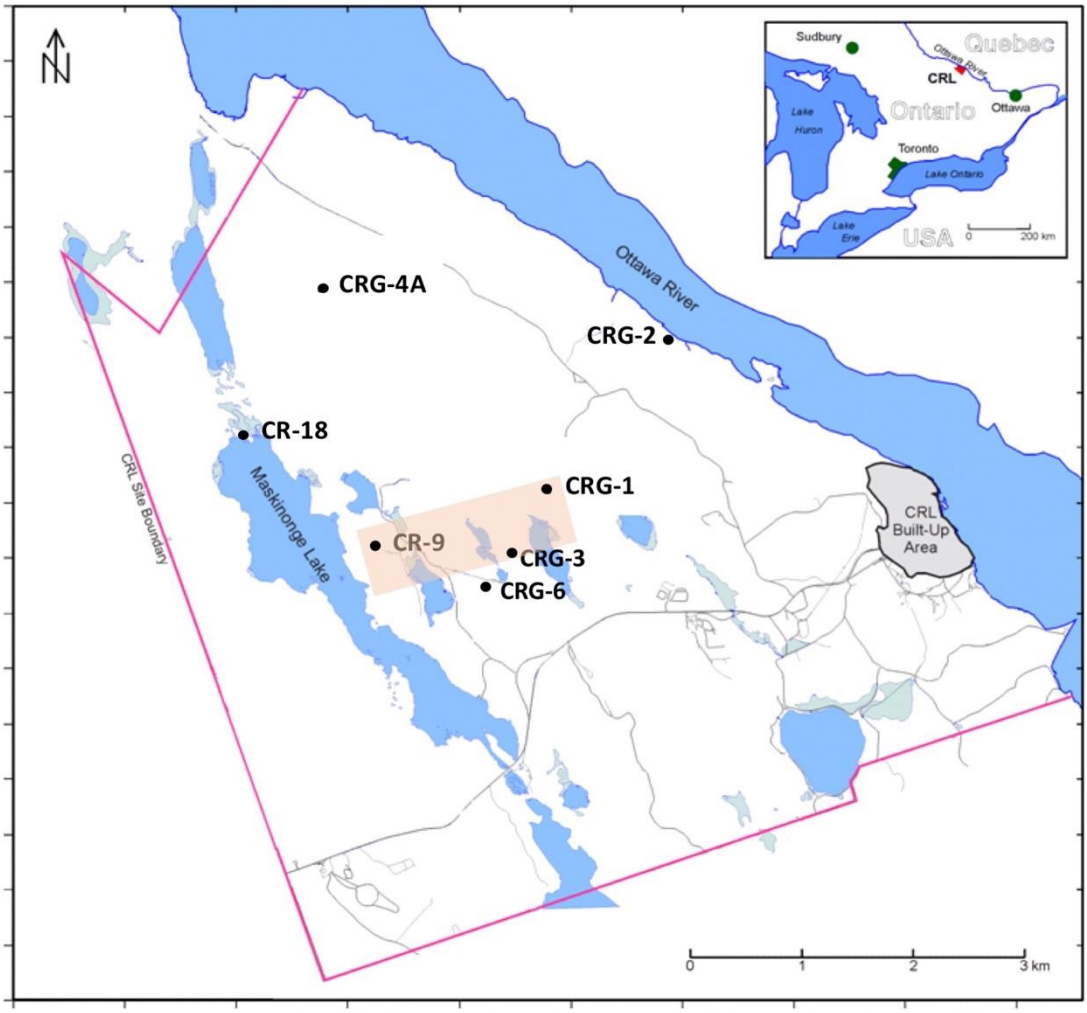


Figure 2-1 Fracture water sampling locations, Chalk River Laboratories, Deep River, ON, Canada. Boreholes CR-9 and CR-18 were drilled into the bedrock ca. 1980, boreholes CRG-1 through CRG-7 were drilled into the bedrock since 2005. All boreholes except CR-18 were sealed and isolated into multiple intervals using Westbay Systems. Borehole CR-18 is an open borehole. The shaded area transects the boreholes used for visualizing the M3 results shown in Figure 2-3.

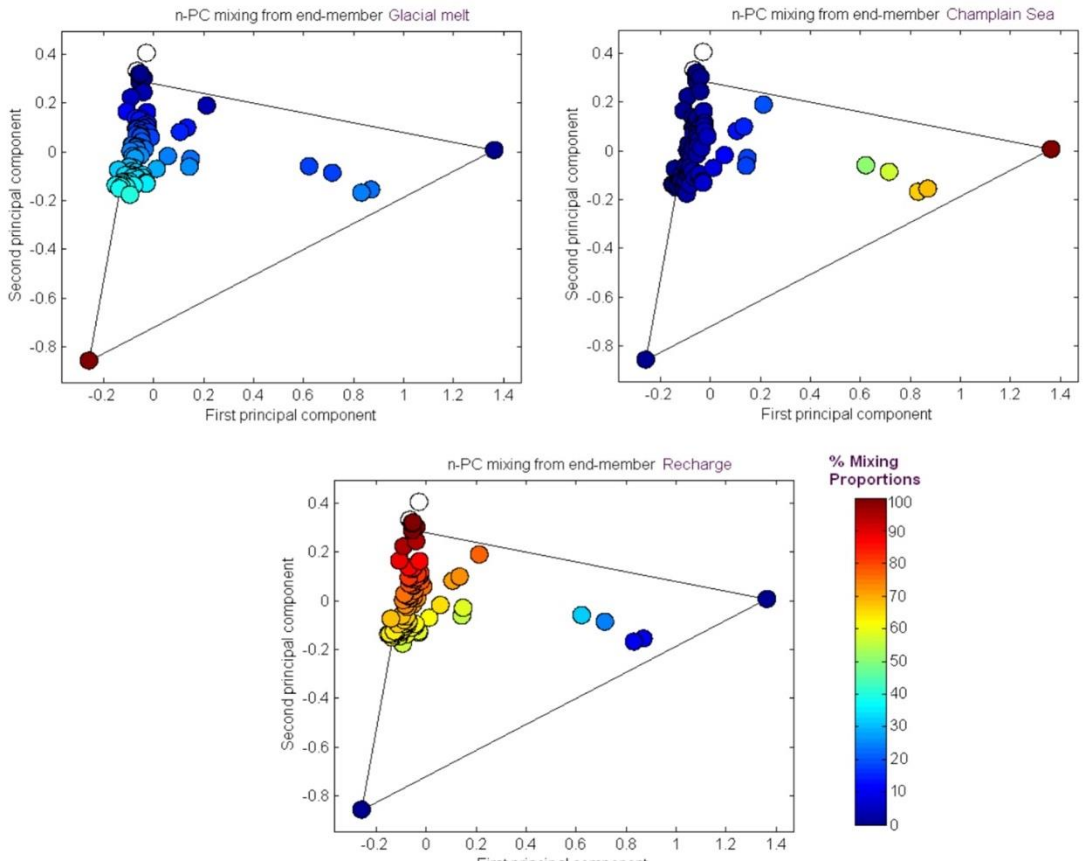


Figure 2-2 M3 Principal Component analysis of the major fracture water ions and stable oxygen $\delta^{18}O$. The results are shown in relation to the M3 modeling results for source water mixing proportions: glacial melt, saline and recharge waters.

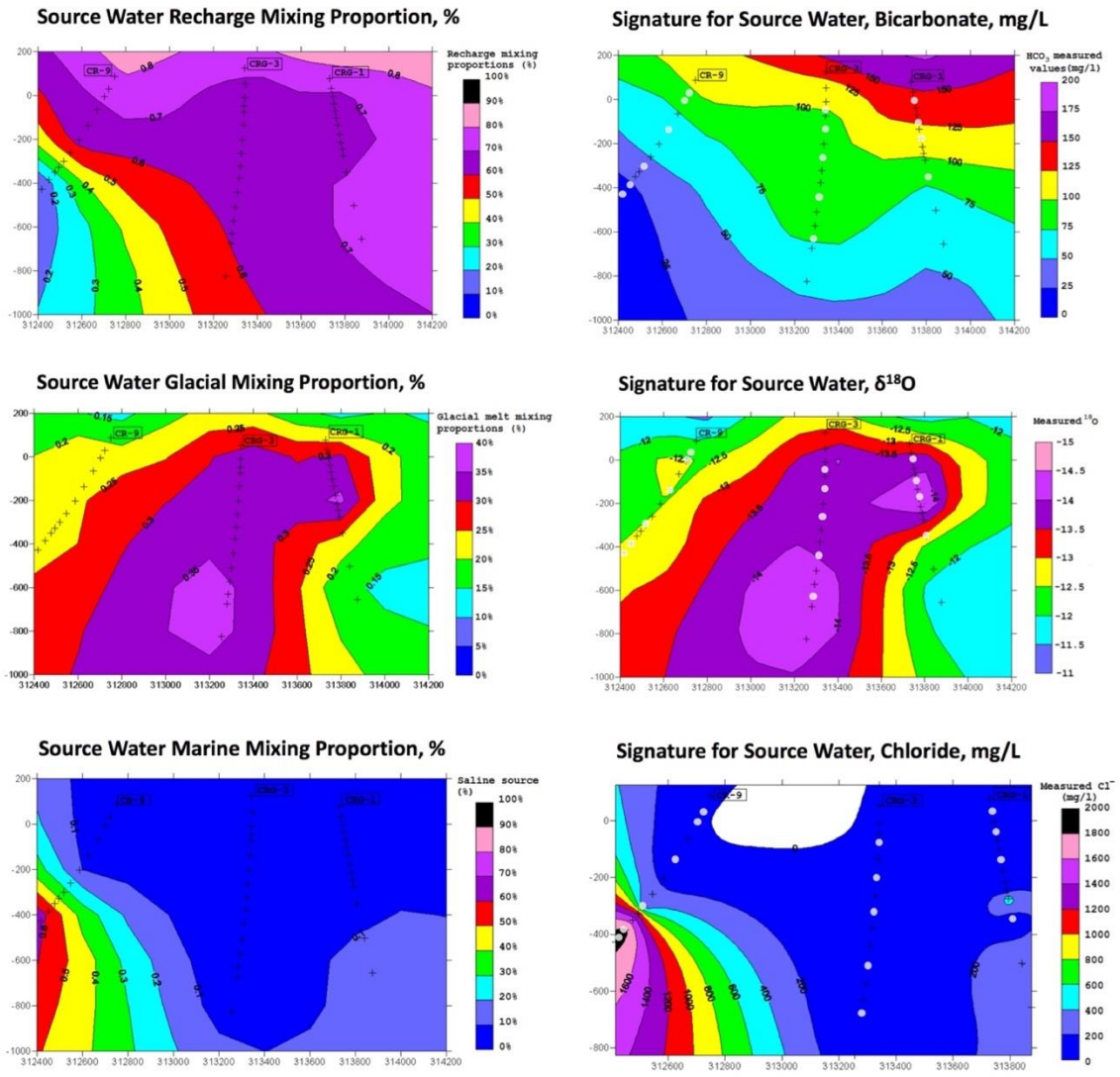


Figure 2-3 M3 results for source water end members (left) and a candidate signature associated with the source water (right). The results are shown in cross section referenced to the boreholes CR-9, CRG-1, and CRG-3.

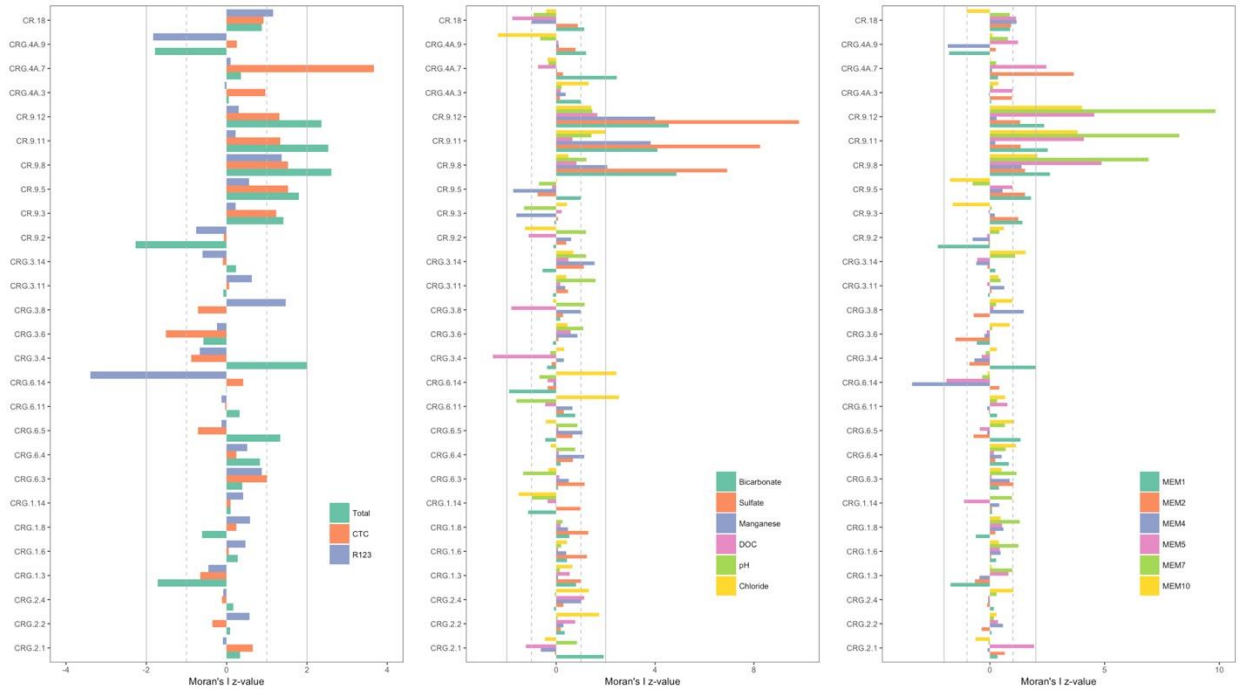


Figure 2-4 Spatial autocorrelation by sampling location (borehole and interval listed in a West-to-East direction starting with borehole CR-18 (see Figure 2-1)). (Left): total and viable cell densities. (Middle): soluble components of the fracture water. (Right): Positive Moran's eigenvector map coefficients. The bars represent the z-values calculated by subtracting the local Moran's I value from the null mean value and normalizing by the null standard deviation. Those locations with z-values more than two standard deviations from the null mean are significant. The gray lines mark first (dashed line) and second (solid line) standard deviations from the null mean.

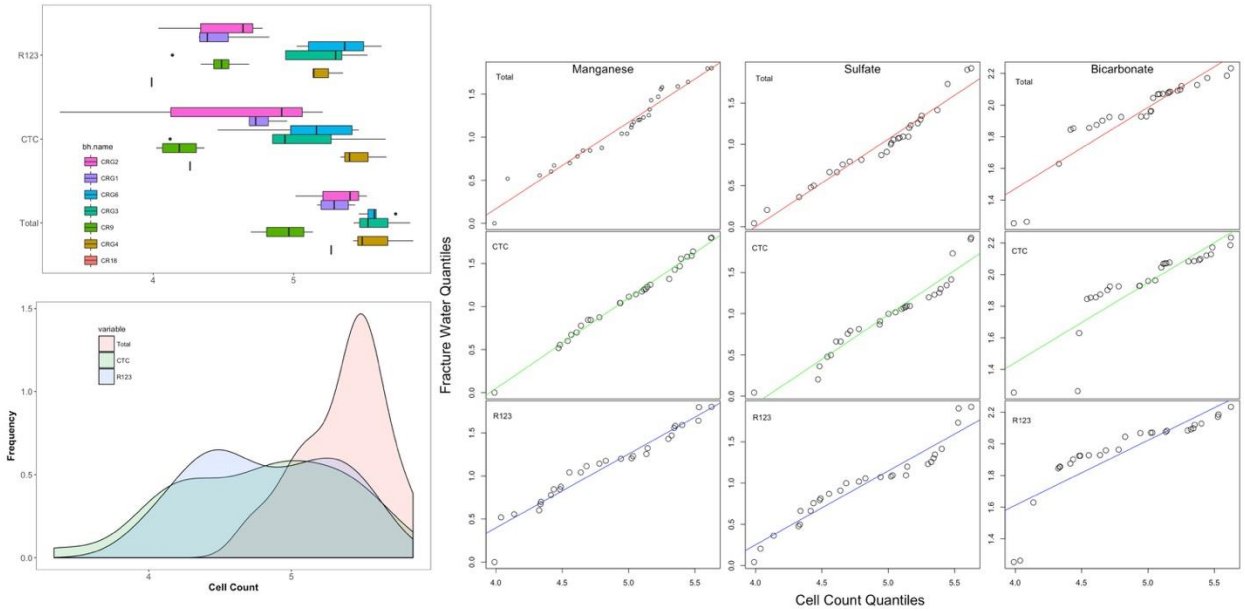


Figure 2-5 Distribution of total and viable cell counts. The boxplots are listed by borehole in a West-to-East direction starting with borehole CR-18 (see Figure 2-1). Also shown are quantile-by-quantile plots for cell density distributions compared to the distributions for fracture water manganese, sulfate and bicarbonate.

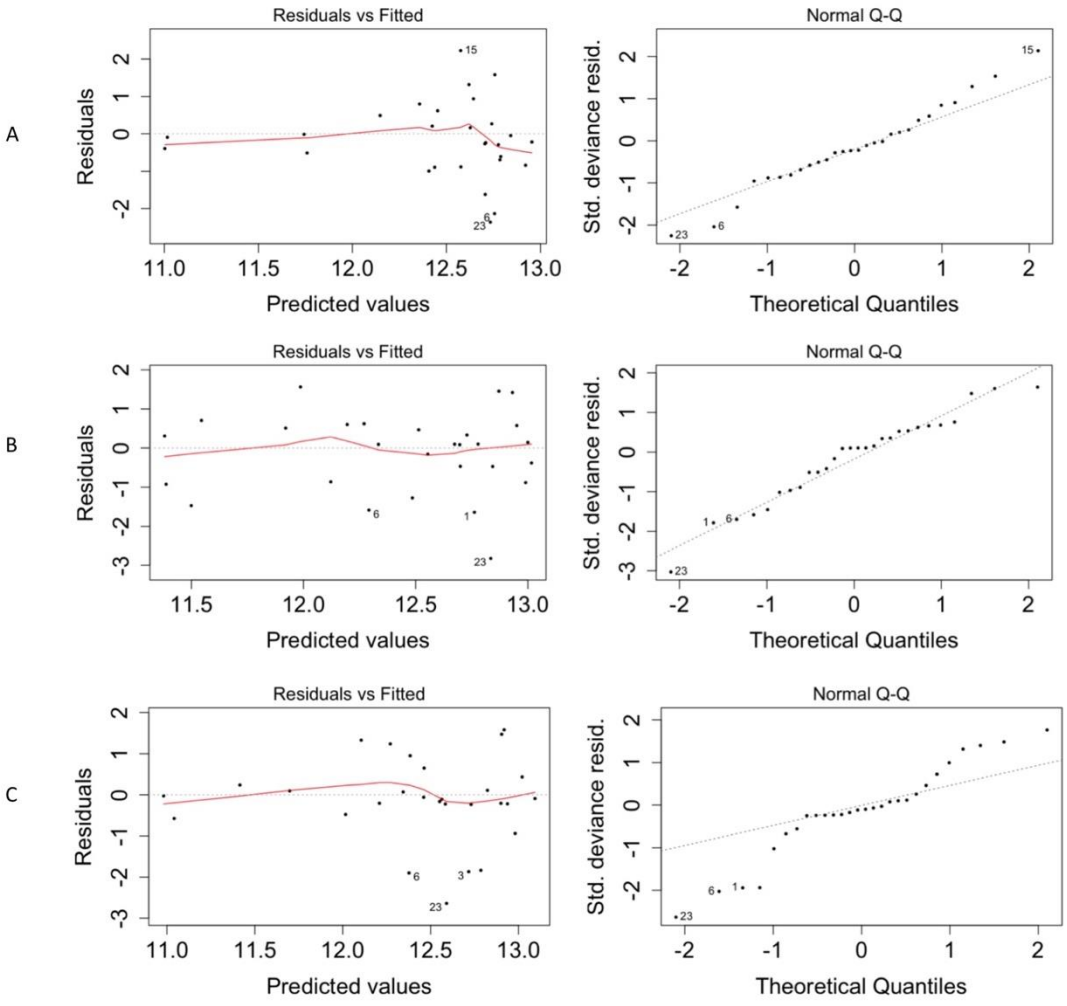


Figure 2-6 Spatial autocorrelation in the model residuals for the total cell count. (A) Significant environmental explanatory variables from Supplementary Table S1 (bicarbonate and manganese). (B) Significant spatial explanatory variables from Supplementary Table S1 (MEM1, MEM2, MEM4 and MEM5). (C) Significant environmental and spatial explanatory variables from Supplementary Table S1 (manganese, MEM1, MEM2 and MEM5).

2.13 Supplemental Information

Table S1A. The Cell Count Data Used for Modelling Total and Viable Microbial Cell Densities

Borehole and Interval	Total Cell Count	CTC Count	R123 Count	CFDA Count
	cells/mL	cells/mL	cells/mL	cells/mL
CRG.1.3	1.60×10^5	4.80×10^4	2.73×10^4	1.56×10^4
CRG.1.6	2.36×10^5	4.84×10^4	2.16×10^4	4.41×10^3
CRG.1.8	1.47×10^5	5.96×10^4	2.11×10^4	7.11×10^3
CRG.1.14	2.74×10^5	9.02×10^4	6.73×10^4	3.43×10^4
CRG.2.1	3.32×10^5	1.61×10^5	1.09×10^4	2.68×10^3
CRG.2.2	1.03×10^5	2.17×10^3	6.02×10^4	2.01×10^3
CRG.2.4	2.52×10^5	8.22×10^4	4.36×10^4	3.68×10^3
CRG.3.4	4.69×10^5	1.32×10^4	1.37×10^4	5.18×10^3
CRG.3.6	6.76×10^5	4.52×10^5	3.36×10^5	2.04×10^5
CRG.3.8	2.67×10^5	7.09×10^4	1.99×10^5	3.08×10^4
CRG.3.11	3.38×10^5	1.85×10^5	2.20×10^5	2.40×10^4
CRG.3.14	2.95×10^5	8.66×10^4	8.76×10^4	1.17×10^4
CRG.4A.3	2.65×10^5	2.50×10^5	1.36×10^5	6.23×10^4
CRG.4A.7	3.08×10^5	4.58×10^5	1.39×10^5	2.66×10^4
CRG.4A.9	7.10×10^5	2.15×10^5	2.25×10^5	2.85×10^4
CRG.4A.13	9.74×10^5	2.90×10^5	4.28×10^5	1.31×10^5
CRG.6.3	3.29×10^5	2.90×10^5	2.53×10^5	1.66×10^4
CRG.6.4	3.87×10^5	2.09×10^5	2.11×10^5	2.74×10^4
CRG.6.5	5.32×10^5	2.88×10^4	1.05×10^5	6.97×10^3
CRG.6.9	3.81×10^5	2.80×10^5	4.22×10^5	5.77×10^4
CRG.6.11	3.69×10^5	9.36×10^4	1.08×10^5	1.79×10^4
CRG.6.14	2.90×10^5	1.01×10^5	3.39×10^5	3.60×10^4
CR.9.2	8.62×10^4	2.32×10^4	4.82×10^4	1.91×10^4
CR.9.3	1.25×10^5	1.64×10^4	3.09×10^4	4.18×10^3
CR.9.5	1.37×10^5	1.44×10^4	3.04×10^4	5.09×10^3
CR.9.8	9.94×10^4	2.18×10^4	3.58×10^4	8.49×10^3
CR.9.11	5.82×10^4	1.05×10^4	2.60×10^4	1.03×10^4
CR.9.12	4.96×10^4	1.09×10^4	2.18×10^4	9.24×10^3
CR.18	1.85×10^5	1.83×10^4	9.75×10^3	2.34×10^3

* On each sampling day, three samplings were performed. Total direct counts for each sampling were performed in triplicate. Viable counts were performed on one of the samplings, also in triplicate.

Table S2. The Geochemical and Cell Count Data Used for Modelling Total and Viable Microbial Cell Densities

Borehole and Interval	Number of Sampling Days*	Easting	Northing	Elevation	pH	Dissolved Organic Carbon	Bicarbonate	Sulfate	Chloride	Iron
				m		mg/L	mg/L	mg/L	mg/L	µg/L
CRG.1.3	2	313743.6	5102838	31.3	8.5	13.4	133.3	1.3	8.2	42
CRG.1.6	2	313760.8	5102857	-67.7	8.4	12.5	120.1	0.6	34.5	48.5
CRG.1.8	2	313771.9	5102870	-130.7	8.6	12.1	118.1	0.1	9.0	84.5
CRG.1.14	2	313815.2	5102917	-355.7	9.1	6.5	70.8	3.6	58.0	22
CRG.2.1	1	314741	5104127.2	137.9	7.8	1.2	152.4	19.0	1.8	29
CRG.2.2	2	314757.5	5104142	92.2	8.6	23.0	116.8	11.4	0.5	20
CRG.2.4	1	314777	5104160	38.2	9.4	17.4	83.1	10.8	1.0	100
CRG.3.4	1	313346.9	5102436.2	-60.3	9.5	3.0	84.0	21.2	7.5	20
CRG.3.6	1	313344.3	5102404	-122.0	9.2	33.0	90.0	16.0	12.5	49
CRG.3.8	2	313335.7	5102336	-247.0	9.2	1.5	116.6	11.0	56.0	14
CRG.3.11	2	313318.3	5102241	-417.0	9.3	1.6	83.3	9.0	21.5	145.5
CRG.3.14	2	313293.2	5102122	-617.0	9.5	1.2	78.9	2.2	21.9	70.5
CRG.4A.3	1	312209.8	5104600.7	1.8	8.6	24.7	122.5	14.8	2.0	30
CRG.4A.7	1	312208	5104495.3	-310.9	8.8	2.5	147.4	5.2	4.9	0
CRG.4A.9	1	312203.4	5104452	-440.0	7.9	18.4	124.3	2.0	14.0	50
CRG.4A.13	1	312178.3	5104338	-799.2	7.3	51.0	463.2	0.3	398.3	1900
CRG.6.3	2	313391.2	5102135	48.0	8.4	9.5	110.0	3.6	6.0	47
CRG.6.4	2	313373.8	5102116	-4.0	9.1	9.9	91.0	5.5	15.0	68
CRG.6.5	1	313351.6	5102092	-68.0	9.0	10.4	74.0	4.7	21.0	430
CRG.6.9	1	313294.5	5102036.1	-229.0	8.7	1.9	170.2	7.1	33.0	73
CRG.6.11	1	313234.8	5101980	-390.0	8.0	12.9	116.0	10.4	32.0	178
CRG.6.14	1	313203.3	5101951.7	-469.0	9.0	1.8	131.0	11.4	32.2	20
CR.9.2	2	312721.5	5102262	29.0	9.0	28.5	120.6	6.4	35.0	79
CR.9.3	2	312691.7	5102245.9	-26.2	8.1	19.1	69.0	17.0	137.7	47.7
CR.9.5	4	312617.4	5102207	-154.0	9.0	12.7	83.8	9.4	43.3	19.8
CR.9.8	4	312466.6	5102145	-364.0	8.2	5.0	41.6	53.0	152.6	51.5
CR.9.11	4	312449.7	5102130.9	-385.0	7.7	6.4	17.3	79.3	1422.5	21
CR.9.12	2	312416.1	5102120.0	-426.3	7.9	2.1	16.9	82.8	1665.8	3.3
CR.18	4	311262.7	5103333	111.0	9.0	1.3	70.4	25.0	95.5	39.5

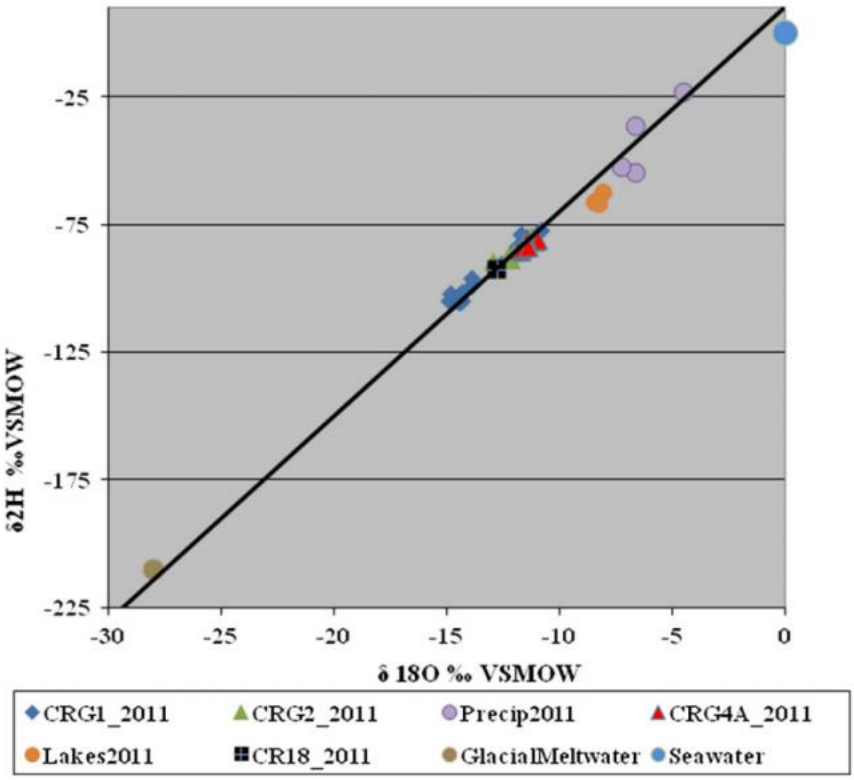


Figure S1. Stable oxygen and stable hydrogen isotopes of dilute fracture groundwater compared

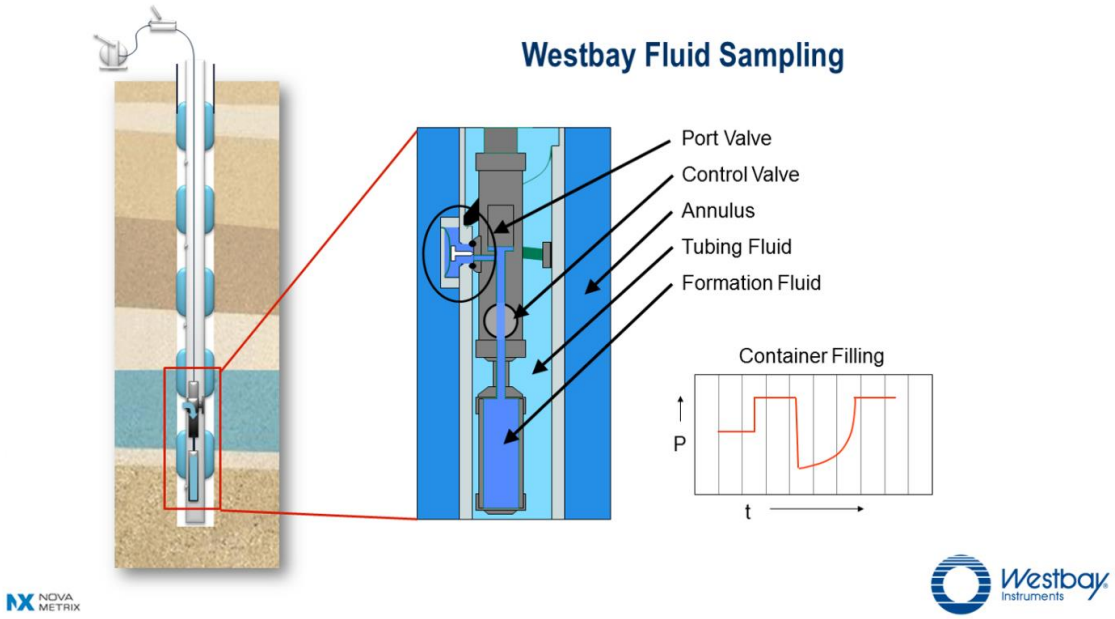


Figure S2. Schematic of the Westbay System.

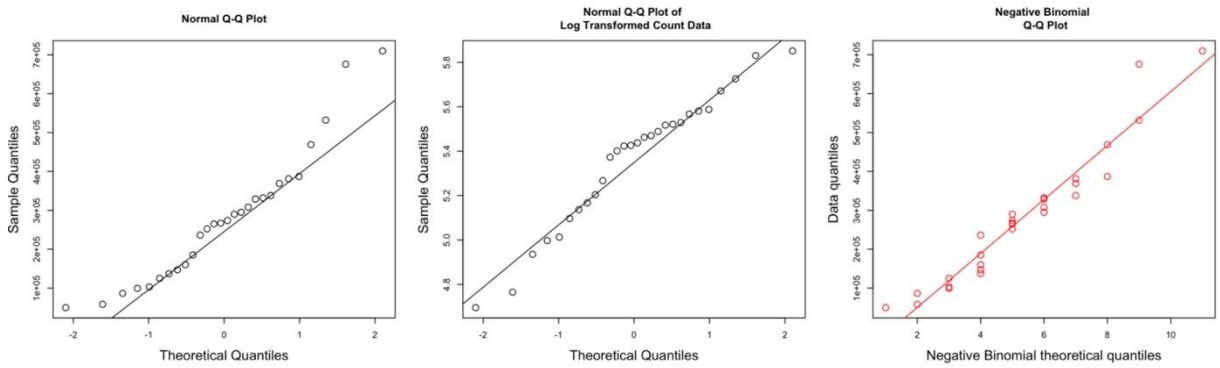


Figure S3. Distribution of the subsurface total cell abundance with a theoretical normal distribution (left and middle panel) and a negative binomial distribution (right panel)

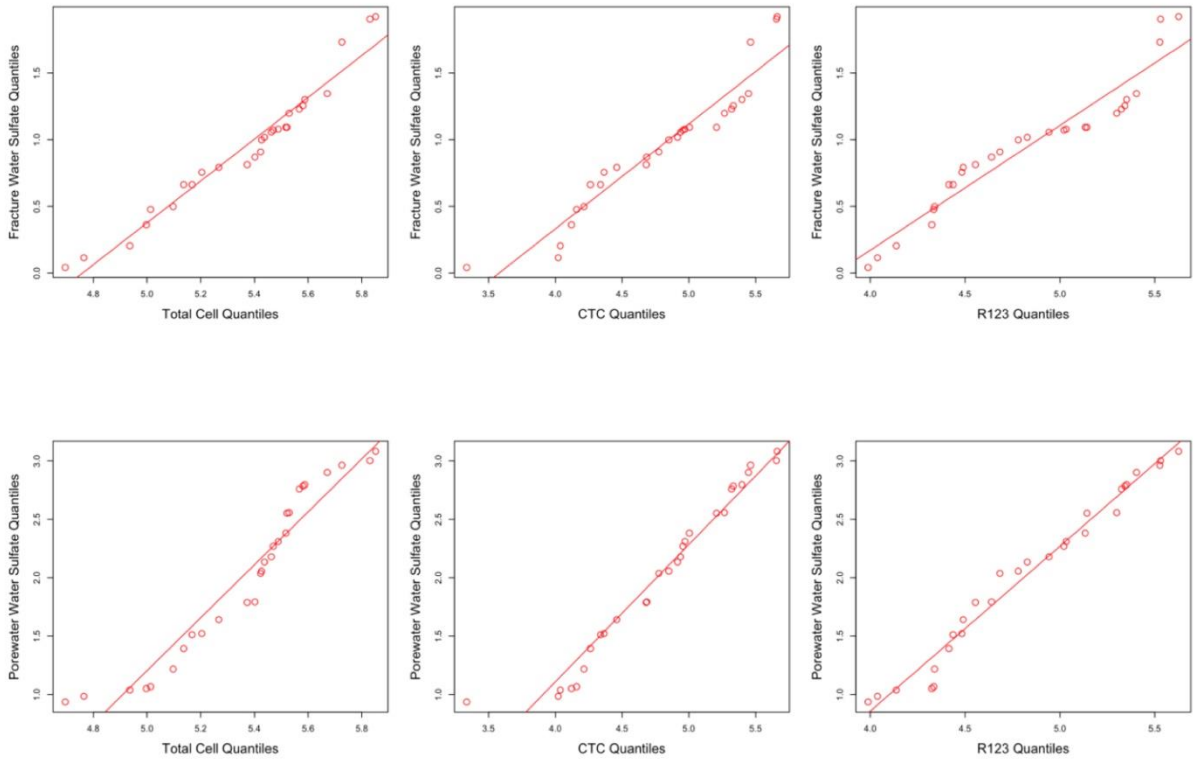


Figure S4 Quantile-by-quantile plot (qqplot) comparing the distribution of sulfate in fracture water (top panel) and porewater water (bottom panel) with the distributions of the total and viable microbial abundances. The data were log transformed for the comparison. The data for porewater was taken from Peterman et.al. (2016), and the data for fracture water was taken from Table S1.

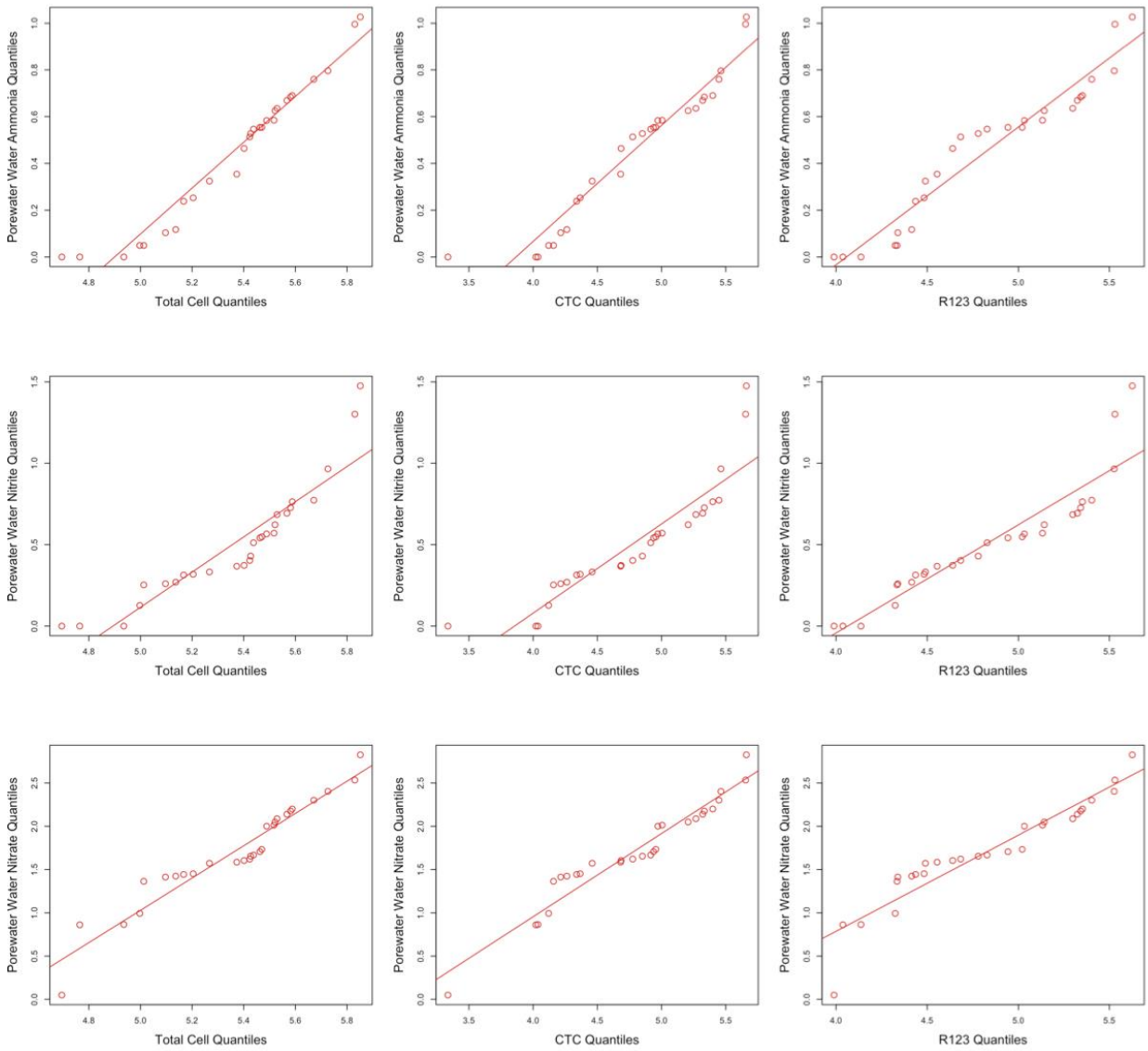


Figure S5 Quantile-by-quantile plot (qqplot) comparing the distribution of ammonia (top panel), nitrite (middle panel) and nitrate (nitrate) in porewater with the distributions of the total and viable microbial abundances. The data were log transformed for the comparison. The data for porewater was taken from Peterman et.al. (2016), and the data for fracture water was taken from Table S1.

3. LOCAL AND REGIONAL DIVERSITY REVEALS DISPERSAL LIMITATION AND DRIFT AS DRIVERS FOR GROUNDWATER BACTERIAL COMMUNITIES FROM A FRACTURED GRANITE FORMATION[4]

3.1 Abstract

Microorganisms found in terrestrial subsurface environments make up a large proportion of the Earth's biomass. Biogeochemical cycles catalyzed by subsurface microbes have the potential to influence the speciation and transport of radionuclides managed in geological repositories. To gain insight on factors that constrain microbial processes within a formation with restricted groundwater flow we performed a meta-community analysis on groundwater collected from multiple discrete fractures underlying the Chalk River Laboratories site (located in Ontario, Canada). Bacterial taxa were numerically dominant in the groundwater. Although these were mainly uncultured, the closest cultivated representatives were from the phenotypically diverse Betaproteobacteria, Deltaproteobacteria, Bacteroidetes, Actinobacteria, Nitrospirae, and Firmicutes. Hundreds of taxa were identified but only a few were found in abundance (>1%) across all assemblages. The remainder of the taxa were low abundance. Within an ecological framework of selection, dispersal and drift, the local and regional diversity revealed fewer taxa within each assemblage relative to the meta-community, but the taxa that were present were more related than predicted by chance. The combination of dispersion at one phylogenetic depth and clustering at another phylogenetic depth suggest both niche (dispersion) and filtering (clustering) as drivers of local assembly. Distance decay of similarity reveals apparent biogeography of 1.5 km. Beta diversity revealed greater influence of selection at shallow sampling locations while the influences of dispersal limitation and randomness were greater at deeper sampling locations. Although selection has shaped each assemblage, the spatial scale of groundwater sampling favored detection of neutral processes over selective processes. Dispersal limitation between assemblages combined with local selection means the meta-community is subject to drift, and therefore, likely reflects the differential historical events that have influenced the current bacterial composition. Categorizing the study site into smaller regions of interest of more closely spaced fractures, or of potentially hydraulically connected fractures, might improve the resolution of an analysis to reveal environmental influences that have shaped these bacterial communities.

[4] Adapted from

Beaton E. D., Stevenson Bradley S., King-Sharp Karen J., Stamps Blake W., Nunn Heather S., Stuart Marilyne, 2016, Local and Regional Diversity Reveals Dispersal Limitation and Drift as Drivers for Groundwater Bacterial Communities from a Fractured Granite Formation, Frontiers in Microbiology, 7, 1933, DOI=10.3389/fmicb.2016.01933

3.2 Introduction

Biogeochemical cycles of the subsurface have the potential to influence the speciation and transport of radionuclides managed in geological repositories. There is limited knowledge on how microbial diversity relates to biogeochemical processes (Griebler and Lueders, 2009) such as the flow of energy through an ecosystem, and how elements like carbon are recycled. These processes involve metabolism and competition, where microorganisms act as catalysts and the available free energy supports a community sensing and responding to environmental changes (Lever, 2011; Jorgensen et al., 2012; Flynn et al., 2013; Algora et al., 2015). While these interactions can connect the supply of electron donor and acceptor compounds to abundant taxa, when viewed at multiple spatial scales, random factors like dispersal, speciation and extinction also influence diversity (Hubbell, 2001). Random processes, therefore, may also explain community dynamics irrespective of the available free energy. Taxa distributions between sampling locations, then, can stem from a combination of competition, environmental constraints, differences in dispersal among a regional pool of taxa and drift due to dispersal limitation.

In the study of subsurface environments, fractured granite formations represent an ecosystem with circuitous hydraulic flow paths that restrict both the flow direction and flow rate of the groundwater; these flow paths reflect complex local and regional recharge. Dispersal limitation, therefore, may obscure the detection of in situ biogeochemical processes except at sufficiently closely spaced sampling locations relative to recharge and discharge. Diverse microbial communities within these formations (Jain et al., 1997; Haveman et al., 1999; Sahl et al., 2008; Itävaara et al., 2011; Thompson et al., 2011; Hallbeck and Pedersen, 2012; Nyyssönen et al., 2012, 2014) display activities for nitrate, iron and sulfate reduction (Pedersen, 1996; Jain et al., 1997; Haveman et al., 1999; Hallbeck and Pedersen, 2012). Our aim is to understand how communities in fractured granite form and what these findings mean in a broader context of a study site.

Patterns of phylogenetic relatedness within a community enable detection of selection as a processes governing community assembly (Horner-Devine and Bohannan, 2006; Emerson and Gillespie, 2008). Beta diversity is a measure of differences in taxa identities, abundances and

phylogenies among locations within a region of interest (Graham and Fine, 2008; Cavender-Bares et al., 2009; Anderson et al., 2011). This measure represents the variation among communities, linking the local community (alpha diversity) to other communities within the region (gamma diversity). Because both selection and dispersal (amongst other neutral processes) contribute to beta diversity (Vellend, 2010), these measures can help explain the organization and functioning of microbial communities within a region of fractured granite. Understanding how a pool of taxa assemble and are maintained over a region will inform how active in situ metabolic processes evolve in space and time, how biogeochemical processes differ across a formation and ultimately how functional and phylogenetic diversity (PD) can affect the solubility and transport of compounds through the formation.

Recently, a conceptual ecological framework (Vellend, 2010) was transformed into an operational framework and applied to a meta-community to compare the relative influences of deterministic and neutral processes on subsurface microbial communities within and across the Ringold and Hanford geologic formations (Stegen et al., 2015). Under this framework, meta-community dynamics differentiate into a combination of selection, dispersion and random processes; differentiation is through the combined results of the beta diversity metrics for the β-Nearest Taxon Index (βNTI) and the Raup Crick index (Chase et al., 2011). βNTI is a measure of a community's phylogenetic composition in terms of relatedness of co-occurring taxa relative to a meta-community (Webb et al., 2008). These calculations, therefore, provide a measure of the phylogenetic relatedness within and between each sampling location. Raup-Crick beta diversity is a measure of dissimilarity between communities compared to a null expectation. The resulting Raup-Crick beta diversity, relative to the corresponding alpha and gamma diversity, provides an indication of whether deterministic or neutral processes have influenced overall community dynamics. It may be that no process dominates and the communities are randomly assembled (Stegen et al., 2015).

In this study, we evaluated a bacteria meta-community of 16S rRNA gene libraries from multiple groundwater-filled granite fractures located within the boundary of Chalk River Laboratories, Ontario, Canada. We accessed groundwater to a depth of 670 m from drill holes transecting low permeability stacked gneiss assemblages underlying the study site. We

compared the alpha and beta diversity to the measured groundwater variables and to the spatial locations of the sampled fractures. By comparing different sampling locations, we were able to relate ecological assembly processes to possible environmental and spatial drivers that govern the structure of the microbial communities.

3.3 Materials and Methods

3.3.1 Sampling Groundwater from Fracture Zones

The locations of boreholes within the study site are shown in Figure 3-1. The Mattawa fault (Ottawa River) and the Maskinonge Lake fault, also shown in Figure 3-1, bound the study site. Diabase dykes traversing the site form another boundary. These features may isolate groundwater into different zones across the area of study. Underlying the site are gneiss forming stacked assemblages consisting of an overlying and underlying garnet-poor assemblage, and a central garnet-rich assemblage. A more detailed description of the site is provided in the Supplementary Information. Samples were collected for geochemical and microbiological analysis from four sealed boreholes (Westbay Multilevel Groundwater Monitoring System (Nova Metrix Groundwater Monitoring (Canada) Ltd) and one open borehole.

Figure 3-2 shows a schematic of a borehole with an installed Westbay System® for multi-level groundwater monitoring. This Figure illustrates how the Westbay tubing and packers isolate multiple zones within the borehole thus preventing unnatural vertical groundwater flow within the borehole itself. The tubing fluid is isolated from the formation fluid. In this arrangement, ambient formation fluid flow can pass through the annulus. From inside the tubing, formation fluid can be accessed by lowering a Westbay sampler and container assembly (also shown) to normally closed valved ports positioned between the packers. A larger schematic illustrates a deployed Westbay sampler assembly that is engaged at a selected port. Once the sampler is positioned and engaged, the remotely operated control valve in the sampler is opened to allow formation fluid from the zone to flow into the empty container. The process is monitored by observing changes in fluid pressure during the sequence of operations (see a typical trace of pressure vs. time in Figure 3-2). Once the container is filled, the sampler valve is closed to seal the formation fluid inside the container at in situ pressure. The assembly is disengaged from the

port (the port valve automatically closes at this time) and the fluid in the sealed container is retrieved to the surface for further handling.

We sampled three fracture locations isolated by borehole CR-9 and seven fracture locations isolated by boreholes CRG-1, CRG-2, and CRG-4A (Table 3-1). The open borehole, CR-18, was accessed by lowering a pump line into the borehole. The groundwater sampling tube assembly consists of four 250 mL stainless steel tubes connected in series by tubing and swagelok fittings. Prior to each sampling, the tubes were sterilized by autoclave and the fittings were sterilized by washing them with 70% ethanol. Validation of the sterilization and transport procedures was performed using sterilized water and PCR with bacterial rRNA 16S primers (Muyzer et al., 1993). Since the tube assemblies contacted only the interior of the casing surface, the probability of introducing surface microbes into the sampled volumes was minimal.

After sampling and transport to the laboratory, opening the filled tubes and dispensing the sampled groundwater, took place inside of a glove box under an atmosphere of filtered nitrogen gas. Inside the glovebox, the groundwater pH (Beckman PHI 265 pH/Temp/mV meter (Beckman Coulter, Inc.)) and conductivity (YSI Model 30 Conductivity Meter (YSI Inc., Yellow Springs, OH, USA) were measured and 100 mL aliquots were filtered through a 0.45 μm filter (isopore polycarbonate, Millipore, Billerica, MA, USA) and immediately acidified by adding 0.5 mL nitric acid (ultra-trace grade, SeastarTM, Baseline®, Fisher Scientific, Ottawa, ON, Canada) for elemental analysis. Elemental composition of the groundwater was determined by inductively coupled plasma-mass spectrometry (ICP-MS, using either a Varian 820-MS (Agilent Technologies, Inc.) or an Element XR (Thermo Scientific)) and by inductively coupled plasma atomic emission spectroscopy (ICP-AES, Optima 3300, Perkin Elmer). Anion concentrations were determined using a Dionex 3000 ICS ion chromatograph (Dionex, Sunnyvale, CA, USA). Dissolved organic (DOC) and inorganic carbon (DIC) were determined using a Dohrmann, model Phoenix 8000-UV Persulfate TOC Analyzer (Teledyne Teckmar, Mason, OH, USA). The concentration of the tracer dye, fluorescein, that was added to the drilling fluid during drilling provided a measure of residual drill water after the drill holes were flushed (nominal fluorescein concentration in the drilling fluid: 1000 μg/kg, fluorescein concentration in sampled fracture water <10 μg/kg.

3.4 Characterization of the Bacterial Assemblages

3.4.1 Direct Cell Counts

Epifluorescent direct cell counting was used to enumerate total microbes in the groundwater. Triplicate 1 mL volumes of the groundwater was incubated with a DNA intercalating dye then filtered onto black polycarbonate filters (0.22 µm, Millipore) for viewing and counting. The dyes were Sybr Green-I and acridine orange (Sigma). The microscope was a Nikon E600 epifluorescent microscope equipped with a Nikon 100x Plan Fluor 100x oil immersion objective and filters for blue excitation/green emission (B-3A, ex: 420-490 nm, longpass emission filter, 520 nm) and green excitation/red emission (G-2E/C (ex: 530-560 nm, 565 nm dichromatc mirror, em: 590-650 nm).

3.4.2 Nucleic Acid Extraction, Quantitative PCR, and Creation of 16S rRNA Gene Libraries

Biomass from each groundwater sample (0.8–1.0 L) was collected by filtration onto sterile 47 mm diameter a 0.22 µm polyethersulfone filters within a sterile 100 mL filter housing (Millipore Corp., Billerica, MA, USA) within 2 h of sample collection. The filters were processed for DNA extraction either immediately or after storage overnight at -20℃. Nucleic acids were extracted from the filters using the UltraClean® microbial DNA isolation kit (Mo Bio Laboratories, Inc) (CRG-1, CRG-2, CRG-4A, CR-18) and the Rapid Water® DNA Isolation Kit (Mo Bio Laboratories, Inc) (CR-9). The two extractions kits have the same chemistry but differed by the volume of the final extract; therefore, the extracts from borehole CR-9 were concentrated by adding 3M sodium acetate and precipitation with ethanol.

Quantitative PCR (qPCR) was used to enumerate the number of copies of the bacterial ssu rRNA genes as a proxy for population density (copies/mL). The group-specific primers used for qPCR were 27F and 338R for bacteria (250 and 125 nM, respectively) and A8F and A44R (500 nM each) for archaea (Nakatsu and Marsh, 2007). Each qPCR contained 2 µL of sample DNA, primers, and 15 µL of 2x SYBR® Green PCR Master Mix (Applied Biosystems), with a total volume of 30 µL. Each experimental DNA was amplified in triplicate. The nearly full-length ssu

rRNA gene from the bacterium Thermacetogenium phaeum or archaeon Methanospirillum hungatei JF-1, cloned into the pCR4 vector (TOPO TA Cloning, Invitrogen), were used for triplicate control reactions. Ten-fold serial dilutions of the control DNA were used for the standard curve, ranging from 3.42×10^3 to 3.42×10^9 copies per reaction. Thermal cycling and detection was performed with a 7300 Real Time PCR System (Applied Biosystems). The cycling conditions for bacterial qPCR included 10 min at 95°C followed by 30 cycles of 30 s at 96°C, 45 s at 55°C, and 45 s at 72°C. The cycling conditions for archaeal qPCR were identical, except for 40 cycles of amplification. The amplified product from each reaction was visualized on an agarose gel (1% w/v in 0.5x TAE, 80 v, 1 h) stained with ethidium bromide to confirm the product size and lack of primer dimer.

The 16S rRNA gene libraries were prepared directly from bacterial qPCR products by pooling and purifying triplicate reactions using the Wizard PCR Preps DNA Purification System (Promega, Madison, WI, USA). The negative qPCR control reactions (no DNA added) were also included. Triplicate PCRs for each library were pooled and purified as described above, and a unique barcode (forward primer only) was attached with the primers TiA-8nt-M13F and TiB-M13R, along with the necessary 454 Fusion A and B tags (454 Life Sciences, Branford, CT, USA). A customized 2-step barcoding approach was used similar to that used by Herbold et al. (2015). Purified libraries were then barcoded with a unique 8 bp barcode, which consisted of a PCR with 5 µL of template, 1X PCR buffer, 0.625 U Taq DNA polymerase (Fermentas, Thermo Fisher, Waltham, MA, USA), and 0.05 µM of the primers (TiA-8nt-27F and TiB-338R) in a total volume of 25 µL. Six cycles of PCR were carried out; otherwise, reaction conditions were identical for bacterial qPCR. The tagged PCR products were purified again using the Wizard PCR Preps DNA Purification System and quantified using the Qubit HS assay (Life Technologies, Carlsbad, CA, USA). Equimolar amounts of each library were pooled and then sequenced using a Genome Sequencer FLX instrument with the GS FLX Titanium series reagents (454 Life Sciences).

3.4.3 Analysis of 16S rRNA Gene Libraries

Raw sequence data was quality filtered and demultiplexed using QIIME (Caporaso et al., 2010b). Demultiplexed libraries were denoised to reduce the error profile inherent within 454 pyrosequencing (Reeder and Knight, 2010).

Sequences were clustered into operational taxonomic units (OTUs) at 97% similarity with USEARCH (Edgar, 2010), and chimeric sequences were removed using de novo and reference-based searches with UCHIME (Edgar et al., 2011). Representative sequences for each OTU were aligned using PyNAST against the SILVA reference database [release 123, (Caporaso et al., 2010a; Quast et al., 2013)] and classified using the RDP naïve Bayesian classifier with the SILVA database (release 123).

3.5 Material Availability

The 16S rRNA sequence reads were submitted to the short read archive under accession number SRR1261803.

3.6 Selection, Dispersal, and Random Processes

3.6.1 Phylogenetic Measures of Assemblage Diversity

The R package (R Core Team, 2015), 'picante' (Kembel et al., 2010) was used to calculate taxa richness, PD, mean pairwise distance (MPD), and the mean nearest taxon distance (MNTD). These calculations compare taxa within assemblages from each sampling location with the taxa from the meta-community tree. PD is the summed phylogenetic branch length connecting all taxa in the meta-community (Faith, 1992). Taxa richness was also derived from the PD calculation by scaling the maximum edge length to one, the resulting PD is the taxa richness. To account for differences in taxa richness on PD, we evaluated PD by using the standardized effect size for PD (by the function ses.pd, abundance.weighted = TRUE). The null model 'taxa.labels' was set at 999 randomizations.

The MPD and MNTD calculations provide measures of co-existence and phylogenetic turnover between the assemblages that make up the meta-community. MPD compares taxa relatedness to the average tree edge length while MNTD compares taxa relatedness at closer phylogenetic depth. To calculate MPD and MNTD, a pair-wise distance matrix of the meta-community tree was first calculated using the function cophenetic(); the assemblages from each sampling location were compared with the mean of the non-diagonal elements (MPD) and with the smallest non-diagonal matrix values for each taxon (MNTD). Patterns of co-existence – of closely related (clustered) or more distantly related (overdispersed) taxa – were evaluated using the standardized effect size for MPD (by the function ses.mpd, abundance.weighted =

TRUE) and for MNTD (by the function ses.mntd, abundance.weighted = TRUE). For each of these calculations the model 'taxa.labels' was set at 999 randomizations.

3.6.2 Beta Diversity

To evaluate the strength of selection, phylogenetic turnover between communities was calculated as the β-nearest taxon index (βNTI) (Stegen et al., 2013). These calculations were performed within the R environment using the codes provided by Stegen et al. (2013) and that are available from Github at https://github.com/stegen/Stegen_etal_ISME_2013/blob/master/bNTI_Local_Machine.r. Within this code, the standardized effect size for MNTD between sampling locations, βMNTD (abundance.weighted = TRUE) is calculated and then multiplied by minus one to give values for the β-nearest taxon index, βNTI.

For those pairwise comparisons that were not significant for selection by βNTI, the strength of dispersal on assembly was evaluated by calculating Raup-Crick (Chase and Myers, 2011; Stegen et al., 2013) that was extended to include comparisons between relative abundances. These calculations helped to discern between the contributions of limiting and homogenizing dispersal. These calculations were performed within the R environment using the codes provided by Stegen et al. (2013); these are available from Github at https://github.com/stegen/Stegen_etal_ISME_2013/blob/master/Raup_Crick_Abundance.r. The result of this analysis is referred to as RCbray. The code was run at 999 repetitions to generate a distribution of null values.

3.6.3 Distance Decay of Similarity

The distance decay of similarity was determined using Bray–Curtis similarity (from the R package 'picante') and weighted UniFrac similarity (from the R package 'GUniFrac'). The geographic distance between sampling locations was calculated using the function earth.dist from the R package 'fossil' and the data fit using the function scatter.smooth().

3.7 Spatial Descriptors and Selection of Explanatory Variables

A Moran's eigenvector map (MEM), that was created by principle coordinates of neighbor matrices (Borcard and Legendre, 2002; Legendre et al., 2009) from within the R packages 'sdep' and 'adespatial,' was used to build a matrix of spatial eigenvectors from a distance matrix of

Easting and Northing, zone 18, Universal Transverse Mercator coordinates for each borehole interval (Table 3-1). The functions used to create the spatial weightings matrix were nbtri(), that converts the spatial coordinates of the sampling locations into a distance neighbors map, and the function nb2listw() that creates the weightings matrix from the neighbors map. The eigenvectors with positive Moran's I values reveal different spatial structures over the entire range of scales encompassed by the geographical sampling area. The first MEM values generated in the analyses represent broader spatial structures, and the last MEM values represent finer spatial structures. The MEM values were combined and standardized with the geochemistry for dissolved organic carbon, bicarbonate, sulfate, iron, manganese, and chloride (Table 3-1) using the function decostand() (within the package 'vegan'). The standardized values were then used to perform a constrained redundancy analysis (RDA). Standardized taxa abundances (by the function decostand() within the package 'vegan', specifying the Hellinger transformation) was the dependent variable. The functions ordistep() and ordiR2step() were then used in forward model selection to define the overall adjusted R2 value.

3.8 Results

3.8.1 Groundwater Geochemistry

Table 3-1 shows the groundwater concentrations of dissolved organic carbon, bicarbonate, sulfate, iron, manganese and major ions from the isolated fractures. The groundwater from fractures isolated by boreholes CR-9 and CR-18 displayed higher chloride and sulfate concentrations than the groundwater from boreholes CRG-1, CRG-2, and CRG-4A. The bicarbonate content of the groundwater decreased with borehole interval depth and displayed a high of 146.5 mg/L (borehole CRG-1) and a low of 19.8 mg/L (borehole CR-9). The soluble iron content of the groundwater tended to increase with depth for groundwater taken from the CRG-boreholes, ranging from a low of 20 µg/L to a high of 100 µg/L, and tended to increase with depth for groundwater taken from borehole CR-9, ranging from a high of 80 µg/L to being below detection. The soluble manganese content of the groundwater tended to decrease with depth in groundwater taken from CRG-boreholes, ranging from a high of 30 µg/L to a being below detection, and to tended to increase with depth in groundwater taken from borehole CR-9, ranging from below detection to 80 µg/L. The nitrate and phosphate ions were below

detection across all samples. A principle component analysis of the major ions and stable oxygen isotope (Supplementary Figure S1) suggested that the groundwater is of meteoric origin. The groundwater pH was slightly alkaline ranging from pH 7.9 to 9.4 and the in situ temperature ranged from 8.8 to 11.7°C.

3.8.2 Bacterial Composition of the Groundwater

Bacteria were dominant in the groundwater, relative to the Archaea. Estimates of the 16S rRNA gene copy numbers for Archaea were at or below the values of the negative control (0 to <100 cells/mL; data not shown), whereas the estimates of bacterial 16S rRNA genes were between 7.8×10^4 and 1.3×10^7 copies per mL of groundwater filtered (Supplementary Table S1). Direct counts of cells by DNA staining estimated the cell densities to be between 5.2×10^4 and 1.1×10^6 cells/mL (Supplementary Table S1).

A total of 114,787 sequences were grouped into 780 OTUs; these varied in their distribution across all libraries. Most of the OTUs were low abundance across the meta-community as only 19 of all the OTUs were represented >1% of the average relative abundance. The distribution of these abundant OTUs is presented in a heatmap (Figure 3-3). The most abundant OTUs identified were a Comamonadaceae (OTU 0), a Desulfovibrio (OTU 1) and a Bacteroidetes WCHB1-32 (OTU 2). The other abundant OTUs were uncultured representatives of Betaproteobacteria [including an Azospira (OTU 17), Sulfuritalea (OTU 5) and a Ferribacterium (OTU 4) from the Rhodocyclales and Simplicispira (OTU 16) and Polarimonas (OTU 8) from the Burkholderiales; Deltaproteobacteria, Bacteroidetes, Firmicutes, Alphaproteobacteria, and Nitrospirae. The heatmap in Figure 3-3 also shows the proportion of OTUs making up <1% of the average relative abundance that were combined as “other OTUs.” The proportion of low abundance OTUs were highest from assemblages within borehole CR-9 (52–63%) and lowest from assemblages within borehole CRG-1 (22–43%).

When viewed at a higher taxonomic level, the most abundant OTUs (Supplementary Table S2) were members of the class Betaproteobacteria (20–40%). The other abundant OTUs (Supplementary Table S2) were members of the Deltaproteobacteria (1–49%), Bacteroidetes (3–26%), Alphaproteobacteria (1–19.8%) and Firmicutes (<1–16%). The assemblages from within the fracture zone accessed by borehole CR-18 had the highest abundances of

Acidobacteria (12.6%) and Alphaproteobacteria (19.8%). Abundant Nitrospiracaea (up to 21.5%) were found in groundwater from boreholes CR-9 and CR-18. Notably, 49% of the bacteria community within the fracture zone accessed by borehole CRG-4A was composed of a single taxon, OTU 1 (shown in the heatmap in Figure 3-3 and identified as Desulfovibrio; this OTU was also found in the assemblages from boreholes CRG-1 (intervals 3, 6, and 14), CRG-2 (interval 2) and in CR-9 (interval 2).

3.8.3 Influences of Selection and Dispersal

3.8.3.1 Diversity and Co-existence

By normalizing the difference between observed values and the null mean as multiples of the standard deviation significant differences from the null expectation are those values (z-values) that are two or more standard deviations from the mean (Vellend et al., 2010; Cadotte and Davies, 2016). The observed values for PD, MPD, and MNTD are plotted in Figure 3-4 (black filled triangles) as these measures relate to the observed taxa richness (the calculated values are listed in Supplementary Tables S3–S5, for PD, MPD and MNTD, respectively.) The regression lines for the corresponding null expectations are shown in red. When the values for an assemblage (sampling locations) displayed observed PD, MPD, or MNTD values that were within the respective null expectations for the meta-community, these locations are marked in Figure 3-4 with both a filled and an open triangle. Otherwise, the values were outside the null expectation and were depicted with a single black triangle; these assemblages display either clustering (located below the regression line for the null expectation) or over dispersion (located above the regression line for the null expectation). We found that at the phylogenetic depth for MPD (Figure 3-4, middle panel), the observed MPD values for all the assemblages were within null distribution, which indicates the assemblages were dispersed evenly relative to the meta-community. Only one observed PD value (Figure 3-4, top panel) was within two standard deviations of the null mean values (marked with both open and closed triangles); this assemblage corresponds to borehole CR-9, interval 2 that had the highest taxa richness (Supplementary Table S3). The remainder of the observed PD values, and all of the observed MNTD values (Figure 3-4, bottom panel), were more than two standard deviations from the

corresponding null expectations. Since these values were below the regression line for the null expectation, the assemblages from these locations displayed lower taxa richness (PD) and co-occurrence of phylogenetically related taxa (MNTD). These measures indicate selection as a driver shaping each of these communities.

3.8.4 Beta Diversity

To help identify the basis for the lower assemblage level diversity and the co-occurrence of related taxa, beta diversity was also calculated and interpreted according to the ecological modeling framework of Stegen (Stegen et al., 2015) for selection, dispersal and drift. A listing of the null comparisons for βNTI and Raup-Crick (as RCbray) are provided in Table 3-2. Briefly, phylogenetic beta diversity βNTI values that are more than two standard deviations on either side of the mean (values of +2 or higher, and of -2 or lower) indicate selection as a driver for community assembly. If a βNTI value falls within the null expectation, namely between -2 and +2, observed differences between the paired sampling locations are random. The resulting distance matrix for βNTI is listed in Supplementary Table S6. Values outside the null expectation for βNTI occur between sampling locations within boreholes CR-9 and CR-18 when paired with each other and the other sampling locations. The phylogenetic beta diversity, therefore, indicates selection as one of the drivers of the community assembly.

To discern the contribution of dispersal on assembly, the sampling locations with βNTI values that were within the null expectation were therefore also used to calculate beta diversity based on taxa identities, Raup-Crick. The Raup-Crick values were further processed using Bray–Curtis dissimilarity, therefore, for RCbray values that were more than +0.95 or less than -0.95 corresponded to values that were more than two standard deviations on either side of the mean.

When values for βNTI are also within the null range and RCbray values are outside of the null range, then either limiting dispersal or homogenizing dispersal, as listed in Table 3-2, are drivers of community assembly, respectively. The resulting distance matrix for RCbray is listed in Supplementary Table S7. When the observed values for both βNTI and RCbray fall within their null distributions, the bacterial dynamics between paired sampling locations are random. Based on the combinations of βNTI and RCbray values from Supplementary Tables S6 and S7, the

assemblages from boreholes CRG-1, CRG-2, and CRG-4A indicate limiting dispersal and random processes. The phylogenetic beta diversity, therefore, identifies dispersal and random processes as drivers of the community assembly.

3.8.5 Distance Decay of Similarity

Distance decay refers to the decrease of similarity as the distance between observations increases. A negative relationship between distance and similarity is implicit in taxa turnover occurring along an environmental gradient but can also occur by isolation created by geography (Nekola and White, 1999). When the decay in similarity occurs along an environmental gradient, the cause of the decay is attributable to competition and fitness, or environmental filtering. When the decay in similarity occurs by spatial isolation, the combined influences of space and time limit movement across landscapes (Nekola and White, 1999). The decay of Bray–Curtis and weighted UniFrac similarities of the meta-community is shown in Figure 3-5. Both similarity measures display a rapid decay distance between 0 and 1.5 km and a shallower decay distance between 1.5 and 5 km.

3.9 Identification of Significant Environmental and Spatial Variables

Next, we compared taxa abundances with the geochemical and spatial variables by performing a constrained RDA followed by forward selection to define a reduced model. The RDA indicated whether there were significant relationships but did not identify explanatory variables. The resulting global adjusted R2 was used as an upper limit for subsequent comparisons to identify the explanatory variables. The variables included were the prospective electron donor and electron acceptor compounds from the groundwater: dissolved organic carbon, bicarbonate, sulfate, iron, and manganese (Table 3-1), and the signatures for a proposed seawater origin (Table 3-1, Supplementary Figure S1). The spatial variables were from the Moran's eigenvector map (MEM) analysis. The details of the modeling are provided in the Supplementary Information.

The resulting model for taxa abundances accounted for 29.4% of the global adjusted R2 variance, $p < 0.01$. The contributions of the explanatory variables for each sampling location are shown in Figure 3-6. There is clear separation between boreholes and the accessed intervals. Also shown in Figure 3-6 are the vectors for environmental variables: organic and inorganic

carbon, iron, manganese, and sulfate, and the first four spatial variables: MEM1, MEM2, MEM3, and MEM4. Initial model selection identified the two largest spatial scales, MEM1 and MEM2 ($p < 0.05$) as the variables associated with taxa abundances. On model reduction, MEM1 was identified as being the most significant variable accounting for 20.4% of the adjusted R2 variance.

None of the geochemistry in the model was found to be significant. Spatial correlograms support these model outcomes – bicarbonate, iron, manganese, and sulfate each display spatial correlation (Figure 3-7). Other components of the groundwater, the dissolved organic carbon, chloride and pH were not spatially correlated (Figure 3-7). The groundwater meta-community appears to be subject to dispersal limitation. Relative to the meta-community, therefore, individual assemblages appear to be subject to random variation in population and size caused by drift.

3.10 Discussion

The microorganisms found in terrestrial subsurface environments are abundant, accounting for 2–19% of Earth's biomass (Kallmeyer et al., 2012) and contribute to important biogeochemical cycles, but their relative inaccessibility means that they are largely unknown. The terrestrial subsurface environment potentially represents a unique script for ecological function, as physical separation likely limits dispersion, which is highly dependent upon the flow of groundwater through a formation. Rock–water interactions mean that groundwater within fracture zones that traverse rock layers can transition to different geochemistry (Swanner and Templeton, 2011). This change, in turn, can affect the composition and diversity of local microbial assemblages, which can vary considerably across short distances (Pedersen, 1997; Gihring et al., 2007; Sahl et al., 2008). Furthermore, taxa richness in the subsurface tends to decrease with increasing depth, and those few taxa present at greater depths often represent novel phylogenetic lineages (Gihring et al., 2007; Sahl et al., 2008). The physical separation of these low abundance microbial assemblages may explain their apparent evolutionary divergence and diversification (Papke and Ward, 2004). We posit that the microbial assemblages within the terrestrial subsurface provide additional evidence that microbial diversity follows an apparent pattern of biogeography on scales of a few millimeters to

thousands of kilometers (Martiny et al., 2006), challenging the once long-standing assumption of unlimited dispersal of microorganisms that are selected by the environment alone. These patterns in microbial diversity are likely driven by the combined influences of selection, dispersion, and drift (Hanson et al., 2012). Whereas, the dynamics of the uncultivated microbial diversity over space and time can reveal patterns of energy flow and biogeochemical processes within the terrestrial subsurface.

The region under study at Chalk River Laboratories (located in Ontario, Canada and shown in Figure 3-1) is underlain by granite that is fractured to depths of several hundreds of meters. In this Shield environment, these fractures provide the pathways for local and regional groundwater flow. Dispersal limitation of groundwater microbes within these fractures would be expected to represent a virtual time capsule of extant microbes, relative to a historical regional pool of taxa continuously being transported to these fractures via recharge. This phenomenon would date back to the period of glacial retreat approximately 11,000 years before present; with an exception that, once isolated within these fractures, ongoing dynamic processes of selection, dispersal and drift continued to influence the community assembly. Bacterial communities were characterized from groundwater obtained from multiple discrete fractures isolated by five boreholes drilled into the rock mass underlying the study site (Figure 3-1). Determinations of biomass and taxa distributions were characterized to gain insight on likely metabolic reactions prevalent at each sampling location. We also analyzed the meta-community within an ecological framework by calculating bacterial diversity at the local and regional scale. The resulting diversity, within and between assemblages, provided insight on the relative contributions of selection, dispersal and drift. The contributions of possible geochemical gradients and distance on the meta-community was also evaluated by three approaches: distance decay of similarity; RDA; and by calculating Moran's I.

The results represent an improved understanding of saturated fractured granite systems. The microbial community density in our samples was between 104 and 105 cells/mL, which is in line with densities between 103 and 107 cells/mL reported for other terrestrial granitic subsurface environments (Pedersen, 1997). Unlike other granitic groundwater systems; however,

methanogenic archaea were not a significant proportion of these communities since Archaea were below detection based on qPCR (Supplementary Table S1).

The bacterial assemblages across boreholes and depths were composed of many, mainly low abundance (<1% relative abundance) taxa or OTUs. This abundance distribution pattern is commonly seen in microbial ecology (Nemergut et al., 2013) in which only a few taxa are present in abundance; most taxa present are rare. Higher phylogenetic/taxonomic groups (phyla, classes) were shared across sites, consisting mainly of members of the phenotypically diverse Betaproteobacterial families *Comamonadaceae* (Willems, 2014) and *Rhodocyclaceae* (Oren, 2014). The OTUs detected across the boreholes were most closely related to uncultivated taxa; however, the closest cultivated representatives are capable of metabolism as diverse as oxic, microoxic, and anoxic growth through chemoorganoheterotrophy, oxidation of iron, sulfur, and hydrogen, and the reduction of nitrate, iron and manganese, and even N-fixation. Members of the *Deltaproteobacteria* family -- *Desulfovibrionaceae* and *Synthrophaceae* -- were also present in relative abundance, suggesting sulfate reduction and syntrophy in anoxia as probable metabolic niches across sampling locations. Finally, two OTUs related to members of the Nitrospiraceae were abundant in CR-18 and CR-9 with increasing depth. These chemolithoautotrophs are capable of nitrite oxidation (Nitrospira) and iron oxidation (Leptospirillum; Lücker et al., 2010), or anaerobic hydrogenotrophic sulfate reduction (Thermodesulfovibrio) (Daims, 2014). Nitrospira were shown to play a key role in the cycling of nitrogen in a deep terrestrial granitic system in Henderson, CO, USA (Swanner and Templeton, 2011).

By evaluating the study site as individual sampling locations and as a meta-community across all sampling locations, we were able to detect probable drivers of the meta-community assembly. Within each sampling location we detected assemblages with lower PD compared to the meta-community (Figure 3-4, top panel). Values for PD change in proportion to the taxa richness (Cadotte and Davies, 2016). Fewer taxa within an assemblage (sampling location) than what is represented by the meta-community, therefore, would be reflected as lower PD. Lower than predicted PD was also found in assemblages of suspended marine bacteria (Horner-Devine and Bohannan, 2006) and of soil bacteria (Goberna et al., 2014). Within each community we also

found that co-occurrence of these fewer taxa spanned the meta-community phylogeny at the phylogenetic depth of the mean pairwise distances (MPD, Figure 3-4, middle panel), signifying even dispersion. However, their co-occurrence at the phylogenetic depth of the MNTD (Figure 3-4, bottom panel), reveals they are more closely related than predicted by the meta-community, signifying clustering. These outcomes support selection as a driver of the local subsurface bacterial assembly.

One possible selective force is environmental filtering. This filtering refers to the incompatibility of some taxa to environmental factors, and so is thought to explain the co-occurrence of phylogenetically related taxa that share evolutionarily conserved traits, while more distant taxa, lacking similar traits, fail to become established (reviewed in Mayfield and Levine, 2010). Co-occurrence of related taxa within an assemblage is considered unlikely under the concept of competition and exclusion (Webb et al., 2002). We have, in our analysis, fewer taxa based on PD, dispersion at more distant phylogenetic depth (MPD) and clustering at more related phylogenetic depth. Mayfield and Levine (Mayfield and Levine, 2010) recently proposed that competition can explain both dispersion and clustering without invoking environmental filtering as a separate mechanism. Instead, either dispersion (niche) or clustering (competitive ability) may be favored, depending on the relative strengths of phylogenetically correlated niche differences and the competitive ability of shared traits. Similarly, niche differences unrelated to phylogeny, when combined with competitive ability that relates to phylogeny, can also explain clustering. For example, sulfate reduction is not a phylogenetically distinct trait, since this trait is shared between some members of the divergent lineages *Deltaproteobacteria* and *Firmicutes*. Abundant members from these lineages were observed in the study described here (Figure 3-3; Supplementary Table S2). We may need to consider the functional genes represented at each sampling location as well as the phylogenetic relationships. Further testing should focus on the roles of niche and traits by comparing 16S rRNA genes and functional genes within sampling locations.

Dispersal is another mechanism that could be influencing the meta-community. At the regional scale, a meta-community represents a regional pool of taxa, whereby interactions between assemblages rely on dispersal. Under the conceptual framework of Vellend (2010), ecological

selection of bacterial taxa can occur in unvarying environmental conditions (with corresponding low rates of community turnover), and in variable environmental conditions (with corresponding high rates of community turnover). Depending on the concomitant rate of taxa dispersal between a pair of communities (sampling locations), the influences of a variable environment can be obscured, even when selection pressure is strong, by the homogenizing effect of a high dispersal rate that has the effect of lowering compositional differences (Vellend, 2010). Lastly, when selection pressure is weak, and the dispersal rate between communities is limiting, the dynamics of turnover within a community are subject to drift and random variations in population. Under the operational framework devised by Stegen et al. (2015) the contributions of selection, dispersal and drift were evaluated using the phylogenetic and taxonomic compositions at each sampling location relative to the meta-community. Null model expectations based on nearest taxon index (βNTI) and an extended Raup-Crick (Chase et al., 2011), referred to as RCbray, (Stegen et al., 2015), differentiated between these roles. These comparisons are based solely on community compositions without reference to environmental or spatial factors. The resulting matrices, Supplementary Tables S6 and S7, indicate selection as a driver between sampling locations from within boreholes CR-9 and CR-18 relative to the other sampling locations. The combinations of βNTI and RCbray, also indicated limited dispersal and random processes as drivers governing the meta-community dynamics. The signals for selection were greater at shallow sampling locations while the influences of dispersal limitation and randomness were greater at deeper sampling locations. This is a trend also seen by Gihring et al. (2007) in the microbial communities in South African gold mines.

While the beta diversity modeling indicates a role for selection as a driver of community assembly on a regional scale, the lower taxa richness and dispersal limitation also suggest regional scale drift. Isolated communities experience a greater degree of ecological drift, and thus higher beta diversity than more connected communities (reviewed in Chase et al., 2011). A distance decay of similarity based on taxonomic (Bray–Curtis) and phylogenetic (1-UniFrac) comparisons (Figure 3-5) showed that similarity between locations leveled off at inter-location distances of up to 1.5 km. The meta-community may reflect isolation created by the rock mass and limited connectivity between fractures rather than environmental gradients (Nekola and

White, 1999). The constrained RDA with forward selection of geochemical and spatial variables relative to the meta-community taxonomy identified only a spatial component in the model; accounting for 20.4% of the adjusted R2 variance of the pool of taxa. The environmental variables also displayed significant positive Moran's I values (Figure 3-7) in the first spatial lag. Overall, the spatial structure of the sampled fractures described the dynamics of the meta-community. Although individual assemblages suggest selection, the spatial structure of the sampling favored detection of limited dispersal as the main driving force governing the meta-community. The beta diversity of the study site also reflects some selection and some drift. To allow identification of variables associated with in situ processes of energy flow and element cycling, future groundwater studies should consider dividing the study site into smaller regions of interest with more closely spaced sampling locations or to limit sampling to hydraulically connected fractures

3.11 Conclusion

The spatial scale of sampling to create a meta-community of suspended subsurface bacteria favored detection of neutral over selective processes. Most of the taxa identified were low abundance and represented uncultured lineages from the metabolically diverse Betaproteobacteria, Deltaproteobacteria, Bacteroidetes, Actinobacteria, Nitrospirae, and Firmicutes. Each microbial assemblage was composed of fewer taxa relative to the meta-community, but these taxa were more related than would be predicted by chance. The combination of dispersion, at one phylogenetic depth, and clustering, at another phylogenetic depth, suggest both niche (dispersion) and filtering (clustering) as drivers of local assembly. Beta diversity also indicated selection as a driver of the subsurface meta-community; however, when attempting to relate taxa abundance to the environment, the spatial scale of groundwater sampling favored detection of neutral over selective processes. Major geochemical components were also spatially auto-correlated. Selection was detected at the level of the meta-community by βNTI that may be related to the selection that was detected within each assemblage by PD, MPD, and MNTD.

Dispersal limitation between assemblages and local selection means that the meta-community is subject to drift, and therefore, likely reflects differential historical events that have influenced

the current bacterial compositions (Andam et al., 2016). Possible historical events that could shape a subsurface bacterial community include differential periods of surface water recharge, dynamics of the fractured rock mass that could release trapped pore water and alter the rock-water interactions, new fractures formed by rebound, and fractures that restricted hydraulic flow paths. A sampling design that includes more closely spaced fractures, or includes potentially hydraulically connected fractures, might reveal a contribution of environment and selection at the level of the meta-community.

3.12 Author Contributions

All authors collected and contributed data sets for analysis as well as participated in the conceptual drafting and revision of this manuscript. KK-S oversaw the groundwater sampling and geochemical analyses. BWS and HN performed sequencing and sequence data analysis. EB conducted subsequent analyses on the sequence data and was the primary author in writing and revising the manuscript. BSS and MS contributed significantly in manuscript development and revision.

3.13 Funding

This work was supported by Natural Resources Canada through the Canadian Nuclear Laboratories Nuclear Legacy Liability Program under which the Geological Waste Management Facility (GWMF) project was funded.

3.14 Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

3.15 Acknowledgments

We gratefully acknowledge the contributions from S. Stroes-Gascoyne, Canadian Nuclear Laboratories, Pinawa, Manitoba. I. Gurban of 3D Terra, Montreal, Quebec, Canada and David Larssen, Westbay Instruments, Vancouver, British Columbia.

3.16 References

Algora, C., Vasileiadis, S., Wasmund, K., Trevisan, M., Kruger, M., Puglisi, E., et al. (2015). Manganese and iron as structuring parameters of microbial communities in arctic marine sediments from the baffin bay. FEMS Microbiol. Ecol. 91:fiv056. doi: 10.1093/femsec/fiv056

Andam, C. P., Doroghazi, J. R., Campbell, A. N., Kelly, P. J., Choudoir, M. J., and Buckley, D. H. (2016). A latitudinal diversity gradient in terrestrial bacteria of the genus streptomyces. mBio 7:e02200-15. doi: 10.1128/mBio.02200-15

Anderson, M. J., Crist, T. O., Chase, J. M., Vellend, M., Inouye, B. D., Freestone, A. L., et al. (2011). Navigating the multiple meanings of beta diversity: a roadmap for the practicing ecologist. Ecol. Lett. 14, 19–28. doi: 10.1111/j.1461-0248.2010.01552.x

Borcard, D., and Legendre, P. (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecol. Modell. 153, 51–68. doi: 10.1016/S0304-3800(01)00501-4

Cadotte, M. W., and Davies, T. J. (2016). Phylogenies in Ecology: A Guide to Concepts and Methods, Chapter 3. Princeton, NJ: Princeton University Press.

Caporaso, J. G., Bittinger, K., Bushman, F. D., DeSantis, T. Z., Andersen, G. L., and Knight, R. (2010a). PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics 26, 266–267. doi: 10.1093/bioinformatics/btp636

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010b). QIIME aallows analysis of high throughput community sequencing data. Nat. Methods 7, 335–336. doi: 10.1038/nmeth.f.303

Cavender-Bares, J., Kozak, K. H., Fine, P. V., and Kembel, S. W. (2009). The merging of community ecology and phylogenetic biology. Ecol. Lett. 12, 693–715. doi: 10.1111/j.1461-0248.2009.01314.x

Chase, J. M., Kraft, N. J. B., Smith, K. G., Vellend, M., and Inouye, B. D. (2011). Using null models to disentangle variation in community dissimilarity from variation in α-diversity. Ecosphere 2, 1–11. doi: 10.1890/ES10-00117.1

Chase, J. M., and Myers, J. A. (2011). Disentangling the importance of ecological niches from stochastic processes across scales. Philos. Trans. R. Soc. Lond. B Biol. Sci. 366, 2351–2363. doi: 10.1098/rstb.2011.0063

Daims, H. (2014). "The family nitrospiraceae," in The Prokaryotes, eds E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin: Springer), 733–749.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME Improves Sensitivity and speed of chimera detection. Bioinformatics 27, 2194–2200. doi: 10.1093/bioinformatics/btr381

Emerson, B. C., and Gillespie, R. G. (2008). Phylogenetic analysis of community assembly and structure over space and time. Trends Ecol. Evol. 23, 619–630. doi: 10.1016/j.tree.2008.07.005

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. Biol. Conserv. 61, 1–10. doi: 10.1016/0006-3207(92)91201-3

Flynn, T. M., Sanford, R. A., Ryu, H., Bethke, C. M., Levine, A. D., Ashbolt, N. J., et al. (2013). Functional microbial diversity explains groundwater chemistry in a pristine aquifer. BMC Microbiol. 13:146. doi: 10.1186/1471-2180-13-146

Gihring, T. M., Moser, D. P., Lin, L. H., Davidson, M., Onstott, T. C., Morgan, L., et al. (2007). The Distribution of microbial taxa in the subsurface water of the kalahari shield, south africa. Geomicrobiol. J. 23, 415–430. doi: 10.1080/01490450600875696

Goberna, M., Navarro-Cano, J. A., Valiente-Banuet, A., García, C., and Verdú, M. (2014). Abiotic stress tolerance and competition related traits underlie phylogenetic clustering in soil bacterial communities. Ecol. Lett. 17, 1191–1201. doi: 10.1111/ele.12341

Graham, C. H., and Fine, P. V. (2008). Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. Ecol. Lett. 11, 1265–1277. doi: 10.1111/j.1461-0248.2008.01256.x

Griebler, C., and Lueders, T. (2009). Microbial biodiversity in groundwater ecosystems. Freshw. Biol. 54, 649–677. doi: 10.1111/j.1365-2427.2008.02013.x

Hallbeck, L., and Pedersen, K. (2012). Culture-dependent comparison of microbial diversity in deep granitic groundwater from two sites considered for a swedish final repository of spent nuclear fuel. FEMS Microbiol. Ecol. 81, 66–77. doi: 10.1111/j.1574-6941.2011.01281.x

Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C., and Martiny, J. B. (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. Nat. Rev. Microbiol. 14, 497–506. doi: 10.1038/nrmicro2795

Haveman, S. A., Pedersen, K., and Ruotsalainen, P. (1999). Distribution and metabolic diversity of microorganisms in deep igneous rock aquifers of finland. Geomicrobiol. J. 16, 277–294. doi: 10.1080/014904599270541

Herbold, C. W., Pelikan, C., Kuzyk, O., Hausmann, B., Angel, R., Berry, D., et al. (2015). A flexible and economical barcoding approach for highly multiplexed amplicon sequencing of diverse target genes. Front. Microbiol. 6:731. doi: 10.3389/fmicb.2015.00731

Horner-Devine, M. C., and Bohannan, B. J. M. (2006). Phylogenetic clustering and overdispersion in bacterial communities. Ecology 87, S100–S108. doi: 10.1890/0012-9658(2006)87[100:PCAOIB]2.0.CO;2

Hubbell, S. P. (2001). “The unified neutral theory of biodiversity and biogeography,” in Monographs in Population Biology, eds S. A. Levin and H. S. Horn (Princeton, NJ: Princeton University Press), 32.

Itävaara, M., Nyyssönen, M., Kapanen, A., Nousiainen, A., Ahonen, L., and Kukkonen, I. (2011). Characterization of bacterial diversity to a depth of 1500 m in the outokumpu deep

borehole, fennoscandian shield. FEMS Microbiol. Ecol. 77, 295–309. doi: 10.1111/j.1574-6941.2011.01111.x

Jain, D. K., Providenti, M., Tanner, C., Cord, I., and Stroes-Gascoyne, S. (1997). Characterization of microbial communities in deep groundwater from granitic rock. Can. J. Microbiol. 43, 272–283. doi: 10.1139/m97-038

Jorgensen, S. L., Hannisdal, B., Lanzen, A., Baumberger, T., Flesland, K., Fonseca, R., et al. (2012). Correlating microbial community profiles with geochemical data in highly stratified sediments from the arctic mid-ocean ridge. Proc. Natl. Acad. Sci. U.S.A. 109, E2846–E2855. doi: 10.1073/pnas.1207574109

Kallmeyer, J., Pockalny, R., Adhikari, R. R., Smith, D. C., and D'Hondt, S. (2012). Global distribution of microbial abundance and biomass in subseafloor sediment. Proc. Natl. Acad. Sci. U.S.A. 109, 16213–16216. doi: 10.1073/pnas.1203849109

Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., et al. (2010). Picante: r tools for integrating phylogenies and ecology. Bioinformatics 26, 1463–1464. doi: 10.1093/bioinformatics/btq166

Legendre, P., Mi, X. C., Ren, H. B., Ma, K. P., Yu, M. J., Sun, I. F., et al. (2009). Partitioning beta diversity in a subtropical broad-leaved forest of china. Ecology 90, 663–674.

Lever, M. A. (2011). Acetogenesis in the energy-starved deep biosphere - a paradox? Front. Microbiol. 2:284. doi: 10.3389/fmicb.2011.00284

Lücker, S., Wagner, M., Maixner, F., Pelletier, E., Koch, H., Vacherie, B., et al. (2010). A nitrospira metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. Proc. Natl. Acad. Sci. U.S.A. 107, 13479–13484. doi: 10.1073/pnas.1003860107

Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., et al. (2006). Microbial biogeography: putting microorganisms on the map. Nat. Rev. Microbiol. 4, 102–112. doi: 10.1038/nrmicro1341

Mayfield, M. M., and Levine, J. M. (2010). Opposing effects of competitive exclusion on the phylogenetic structure of communities. Ecol. Lett. 13, 1085–1093. doi: 10.1111/j.1461-0248.2010.01509.x

Muyzer, G., de Waal, E. C., and Uitterlinden, A. G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl. Environ. Microbiol. 59, 695–700.

Nakatsu, C. H., and Marsh, T. L. (2007). "Analysis of microbial communities with denaturing gradient gel electrophoresis and terminal restriction fragment length polymorphism," in Methods for General and Molecular Microbiology, Third Edn, eds C. A. Reddy, T. L. Beveridege, J. A. Breznak, G. A. Marzluf, T. M. Schmidt, and L. R. Snyder (Washington, DC: American Society of Microbiology), 909–923.

Nekola, J. C., and White, P. S. (1999). The distance decay of similarity in biogeography and ecology. J. Biogeogr. 26, 867–878. doi: 10.1046/j.1365-2699.1999.00305.x

Nemergut, D. R., Schmidt, S. K., Fukami, T., O'Neill, S. P., Bilinski, T. M., Stanish, L. F., et al. (2013). Patterns and processes of microbial community assembly. Microbiol. Mol. Biol. Rev. 77, 342–356. doi: 10.1128/MMBR.00051-12

Nyyssönen, M., Bomberg, M., Kapanen, A., Nousiainen, A., Pitkänen, P., and Itävaara, M. (2012). Methanogenic and sulphate-reducing microbial communities in deep groundwater of crystalline rock fractures in olkiluoto, finland. Geomicrobiol. J. 29, 863–878. doi: 10.1080/01490451.2011.635759

Nyyssönen, M., Hultman, J., Ahonen, L., Kukkonen, I., Paulin, L., Laine, P., et al. (2014). Taxonomically and functionally diverse microbial communities in deep crystalline rocks of the fennoscandian shield. ISME J. 8, 126–138. doi: 10.1038/ismej.2013.125

Oren, A. (2014). "The family rhodocyclaceae," in The Prokaryotes, eds E. Rosenberg, E. F. DeLong, F. Thompson, S. Lory, and E. Stackebrandt (Berlin: Springer), 975–998.

Papke, R. T., and Ward, D. M. (2004). The importance of physical isolation to microbial diversification. FEMS Microbiol. Ecol. 48, 293–303. doi: 10.1016/j.femsec.2004.03.013

Pedersen, K. (1996). Investigations of subterranean bacteria in deep crystalline bedrock and their importance for the disposal of nuclear waste. Can. J. Microbiol. 42, 382–391. doi: 10.1139/m96-054

Pedersen, K. (1997). Microbial life in deep granitic rock. FEMS Microbiol. Rev. 20, 399–414. doi: 10.1111/j.1574-6976.1997.tb00325.x

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596. doi: 10.1093/nar/gks1219

R Core Team (2015). R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing. Available at: http://www.R-project.org/

Reeder, J., and Knight, R. (2010). Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. Nat. Methods 7, 668–669. doi: 10.1038/nmeth0910-668b

Sahl, J. W., Schmidt, R., Swanner, E. D., Mandernack, K. W., Templeton, A. S., Kieft, T. L., et al. (2008). Subsurface microbial diversity in deep-granitic-fracture water in colorado. Appl. Environ. Microbiol. 74, 143–152. doi: 10.1128/AEM.01133-07

Stegen, J. C., Lin, X., Fredrickson, J. K., Chen, X., Kennedy, D. W., Murray, C. J., et al. (2013). Quantifying community assembly processes and identifying features that impose them. ISME J. 7, 2069–2079. doi: 10.1038/ismej.2013.93

Stegen, J. C., Lin, X., Fredrickson, J. K., and Konopka, A. E. (2015). Estimating and mapping ecological processes influencing microbial community assembly. Front. Microbiol. 6:370. doi: 10.3389/fmicb.2015.00370

Swanner, E. D., and Templeton, A. S. (2011). Potential for nitrogen fixation and nitrification in the granite-hosted subsurface at henderson mine, CO. Front. Microbiol. 2, 254. doi: 10.3389/fmicb.2011.00254

Thompson, P., Baumgartner, P., Beaton, E. D., Chan, T., Kitson, C., Kozak, E., et al. (2011). An Investigation of the Suitability of the Chalk River Site to Host a Geologic Waste

Management Facility for AECL's Low and Intermediate Level Wastes., Waste Management, Decommissioning and Environmental Restoration for Canada's Nuclear Activities, Current Practices and Future Needs. Toronto, ON: Canadian Nuclear Society.

Vellend, M. (2010). Conceptual synthesis in community ecology. Q. Rev. Biol. 85, 183–206. doi: 10.1086/652373

Vellend, M., Cornwell, W. K., Magnuson-Ford, K., and Mooers, A. (2010). "Measuring phylogenetic biodiversity," in Biological Diversity: Frontiers in Measurement and Assessment, 2010, Chap. 14, eds A. E. Magurran and B. J. McGill (Oxford: Oxford University Press).

Webb, C. O., Ackerly, D. D., and Kembel, S. W. (2008). Phylocom: software for the analysis of phylogenetic community structure and trait evolution. Bioinformatics 24, 2098–2100. doi: 10.1093/bioinformatics/btn358

Webb, C. O., Ackerly, D. D., McPeek, M. A., and Donoghue, M. J. (2002). Phylogenies and community ecology. Annu. Rev. Ecol. Syst. 33, 475–505. doi: 10.1146/annurev.ecolsys.33.010802.150448

Willems, A. (2014). "The family comamonadaceae," in The Prokaryotes, eds E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin: Springer), 777–851.

3.17 List of Tables

Table 3-1
Geochemistry and descriptive parameters for groundwater from each borehole sampling location.

		CRG-1				CRG-2		CRG-4A	CR-9			CR-18
Interval:		**3**	**6**	**8**	**14**	**2**	**4**	**9**	**2**	**5**	**12**	**--**
Sampling Date m/d/y		05/30/ 2011	05/31/ 2011	06/01/ 2011	06/02/ 2011	09/20/ 2011	09/21/ 2011	08/23/ 2011	11/04/ 2008	11/06/ 2008	11/13/ 2008	09/22/ 2011
Easting*	**m**	313743.6	313760.8	313771.9	313815.2	314757.5	314777	312203.4	312721.5	312449.75	312416.14	311262.7
Northing*	**m**	5102838	5102857	5102870	5102917	5104142	5104160	5104452	5102262	5102130.9	5102120	5103333
Depth	**m**	145	244	307	532	72	126	604	96	279	672	5
Elevation	**m****	31	-68	-131	-356	92	38	-440	29	-154	-426	111
Conductivity	**µs/cm**	233	292	220	302	223	210	290	258	317	4225	213
Temperature	**ºC**	9.0	8.6	9.1	11.1	9.6	8.4	11.7	8.6	8.4	11.7	-
pH		8.5	8.5	8.6	9.1	8.6	9.4	7.9	9.3	9.1	8.5	8.5
DOC	**mg/L**	25.9	23.5	23.2	7.9	23	17.4	18.4	22	19	1.9	1
HCO_3^-	**mg/L**	146.5	131.1	130.7	76.6	116.8	83.1	124.3	121.9	91.5	19.8	79.5
Cl^-	**mg/L**	5.4	31	6	41	0.5	1	14	58	28	1,572	91
SO_4^{-2}	**mg/L**	1.3	0.5	0.1	0.2	11.4	10.8	2	7.9	7.3	83	25
NO_3^-	**mg/L**	<mdl	<mdl	<mdl	<mdl	<mdl	<mdl	<mdl	<mdl	<mdl	<mdl	<mdl
PO_4^{-3}	**mg/L**	<mdl	<mdl	<mdl	<mdl	<mdl	<mdl	<mdl	<mdl	<mdl	<mdl	<mdl
Ba^{+2}	**mg/L**	0.02	0.04	0.02	0.01	<mdl	<mdl	0.01	0.01	0.01	0.14	0.1
Ca^{+2}	**mg/L**	10.5	22.9	10.8	4.2	11.2	3.1	13	9	7.8	325	18.1
Mg^{+2}	**mg/L**	1.5	4.1	1.3	0.4	1.2	0.2	1.2	0.8	0.3	8.1	1.1
Na^+	**mg/L**	40	31	36	61	39	43	53	47	64	510	80
$Fe_{soluble}$	**µg/L**	40	60	80	20	20	100	50	80	30	<mdl	50
$Mn_{soluble}$	**µg/L**	30	30	20	10	10	<mdl	20	10	<mdl	80	10

Note: na = not available
mdl = method detection limit
* Universal Transverse Mercator coordinates, zone 18.
** elevation relative to sea level

Table 3-2
Beta diversity relationship to ecological process (Stegen et al., 2013).

Ecological process:	βNTI	RC_{bray}
Selection (variable)	> +2	--
Selection (homogeneous)	< -2	--
Dispersal (limiting), drift	null	> +0.95
Dispersal (homogenizing)	null	< -0.95
Random assembly	null	null

3.18 List of Figures

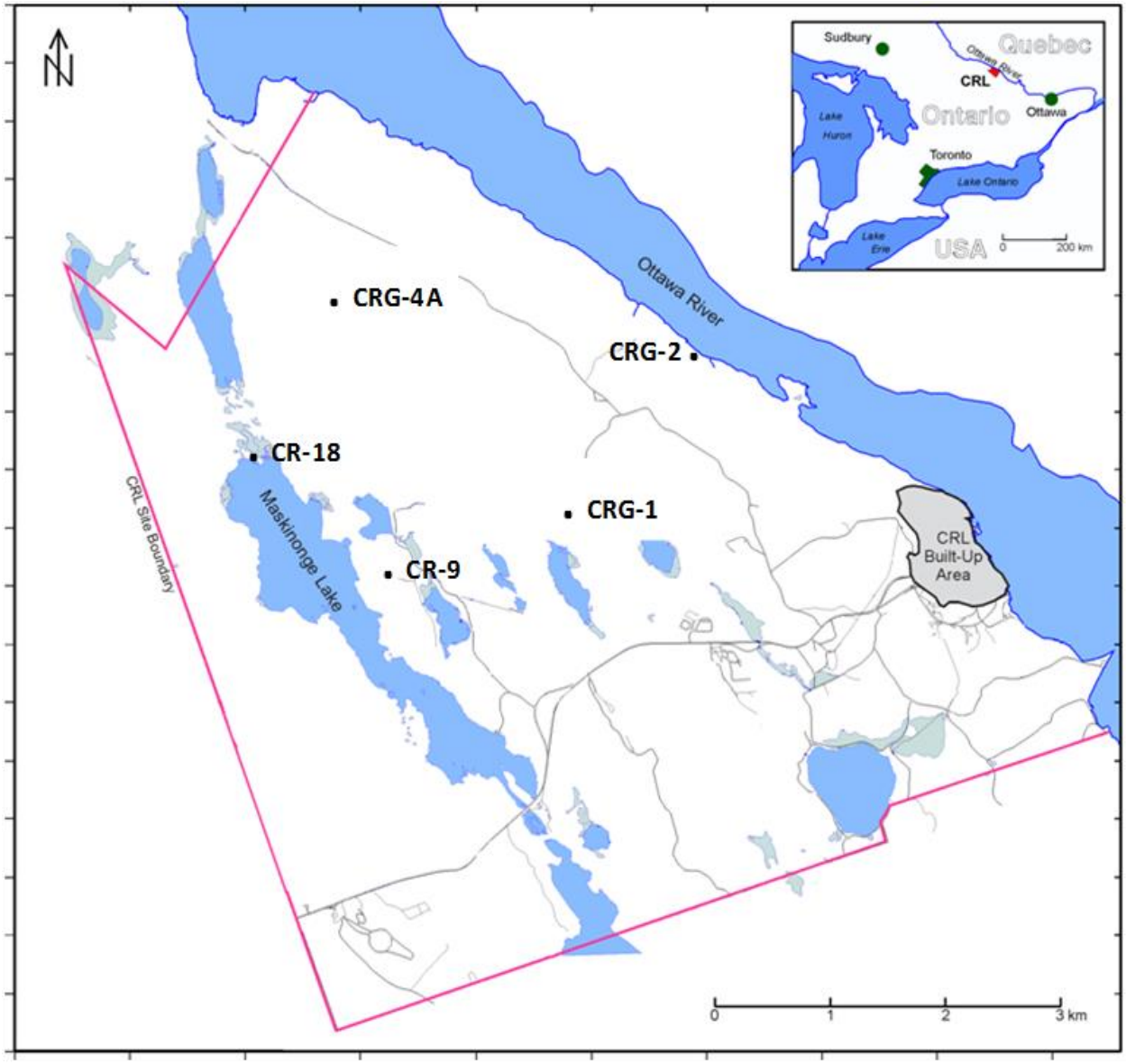


Figure 3-1 Map of the Chalk River Laboratories site located in Eastern Ontario, Canada showing the approximate locations of boreholes included in the study.

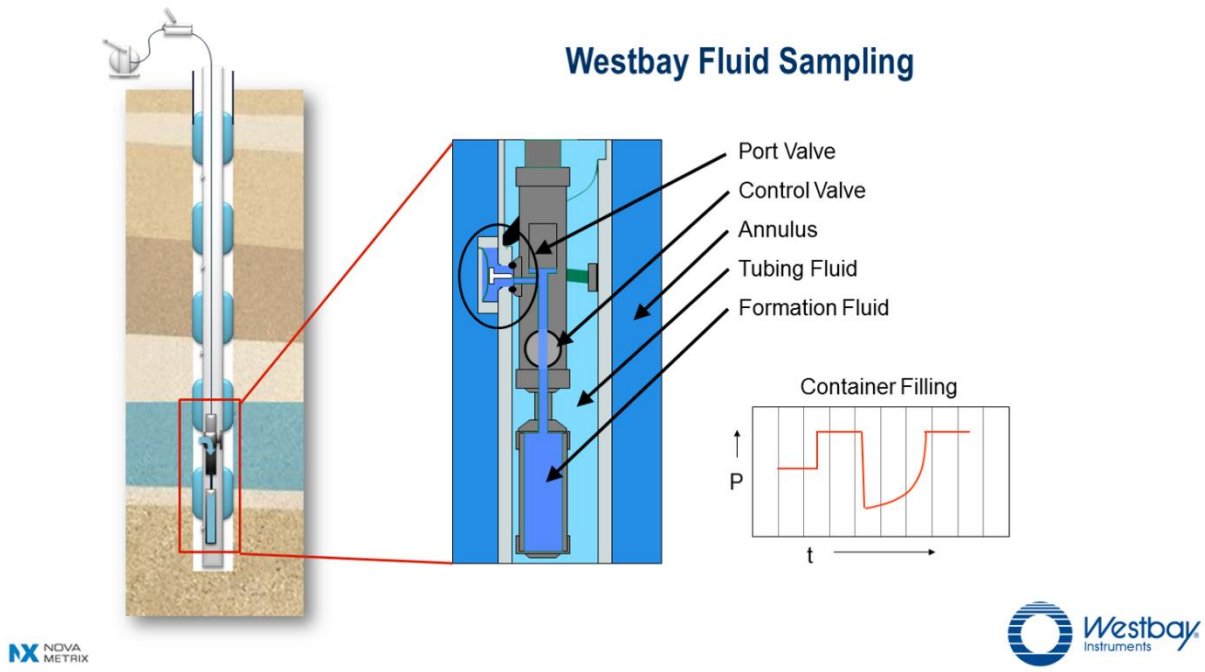


Figure 3-2 Schematic of the Westbay System for multi-level groundwater monitoring (used with permission of Nova Metrix Ground Monitoring Ltd.)

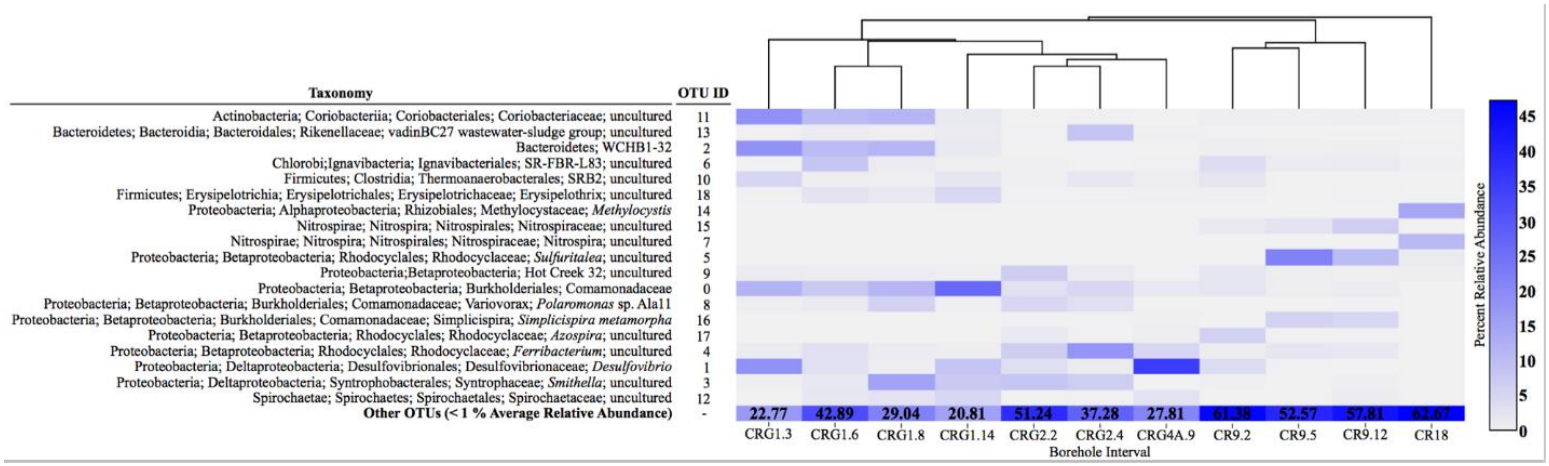


Figure 3-3 Diagram showing the taxonomic identification of the most abundant operational taxonomic units (OTUs, >1% average relative abundance) and their relative abundance in each sample (heatmap). The relative abundance of all other OTUs (< 1% on average) is grouped as "Other OTUs", with their collective relative abundance (%) denoted in each box. The dendrogram above the samples depicts similarity among the sampling locations based on unweighted UniFrac distances.

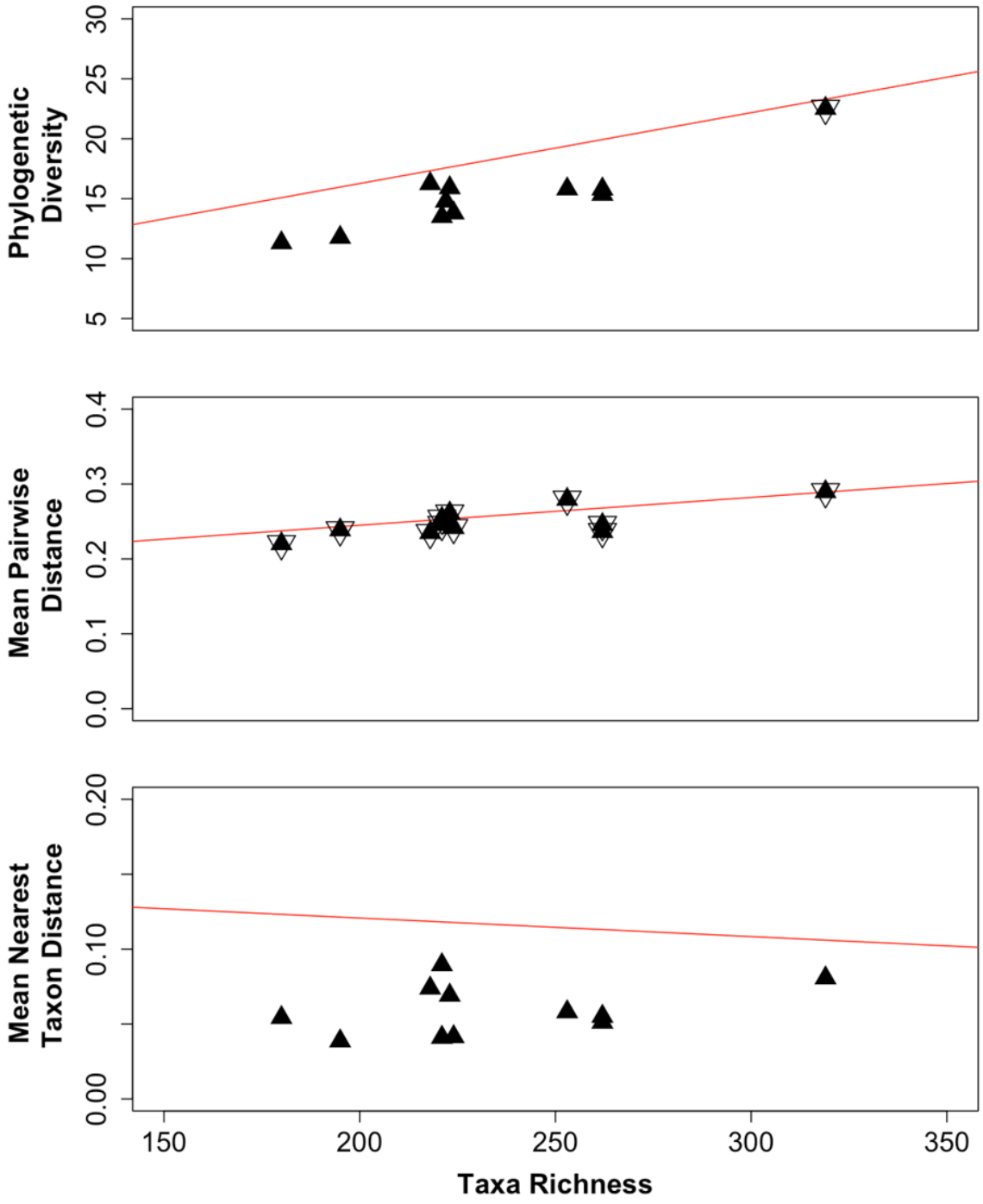


Figure 3-4 Observed values (black triangles) for phylogenetic diversity (PD, top panel), mean pairwise distance (MPD, middle panel) and mean nearest taxon distance (MNTD, bottom panel) of assemblages relative to the taxa richness. The regression line is shown for the corresponding null mean values (red lines). Assemblages marked with an open triangle indicate the observed values were within two standard deviations of the null mean values.

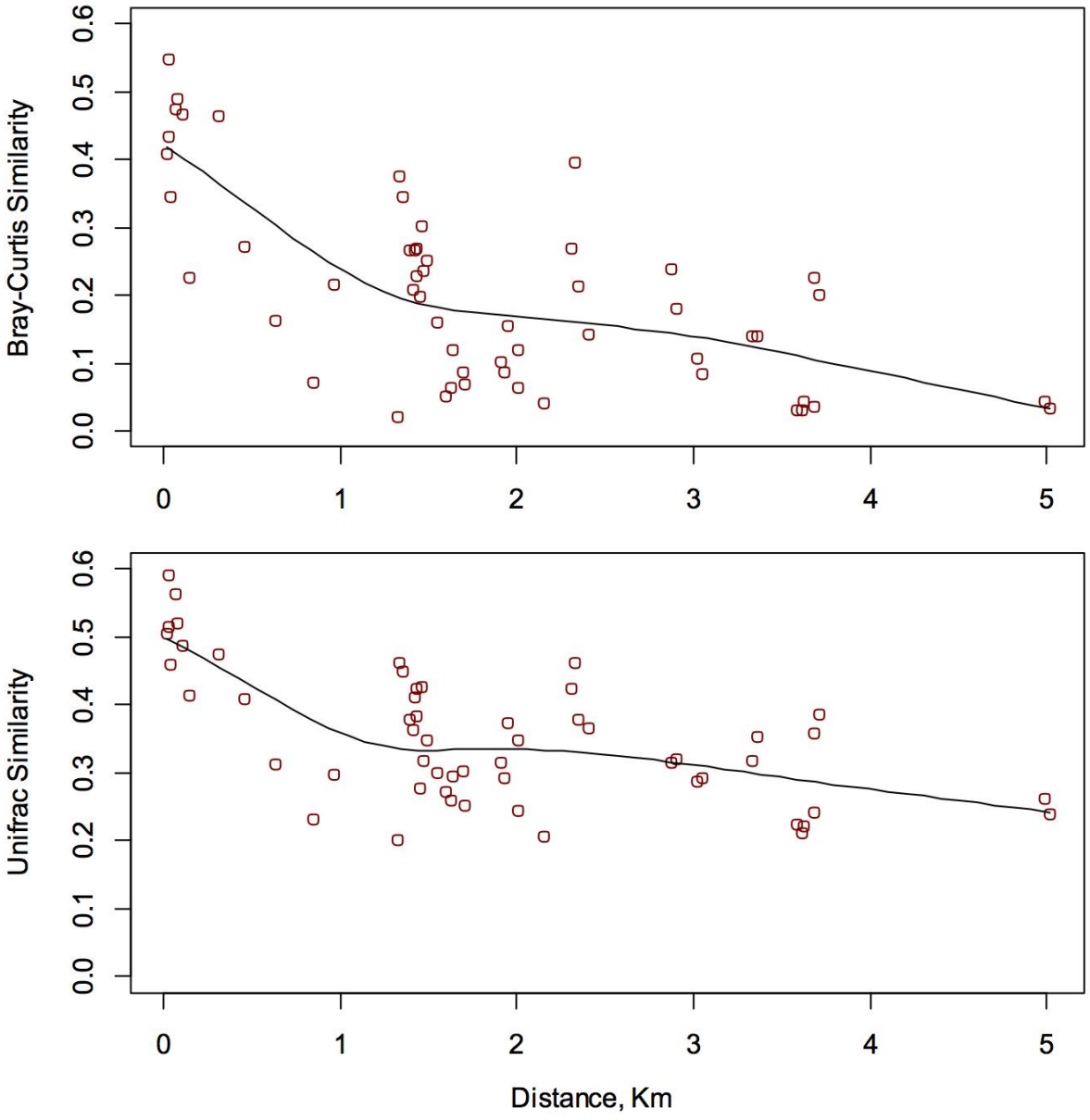


Figure 3-5 Bray-Curtis and UniFrac distance decay relationships of groundwater bacteria across the study site

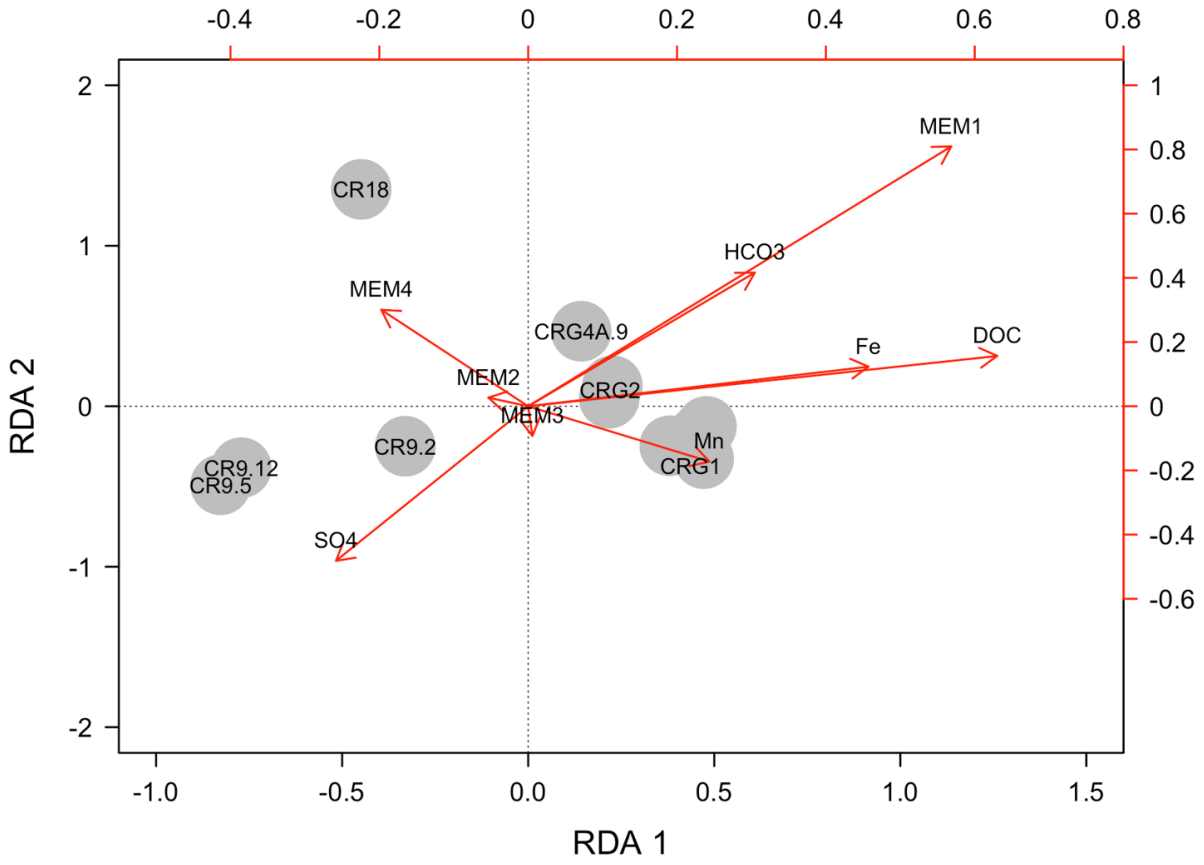


Figure 3-6 Explanatory power of selected environmental and spatial variables on taxa abundances. The axes are the first and second coefficients from the redundancy analysis (RDA). Global adjusted R^2 of 29.4%. The sampling locations are labeled according to the borehole and interval used to access the groundwater.

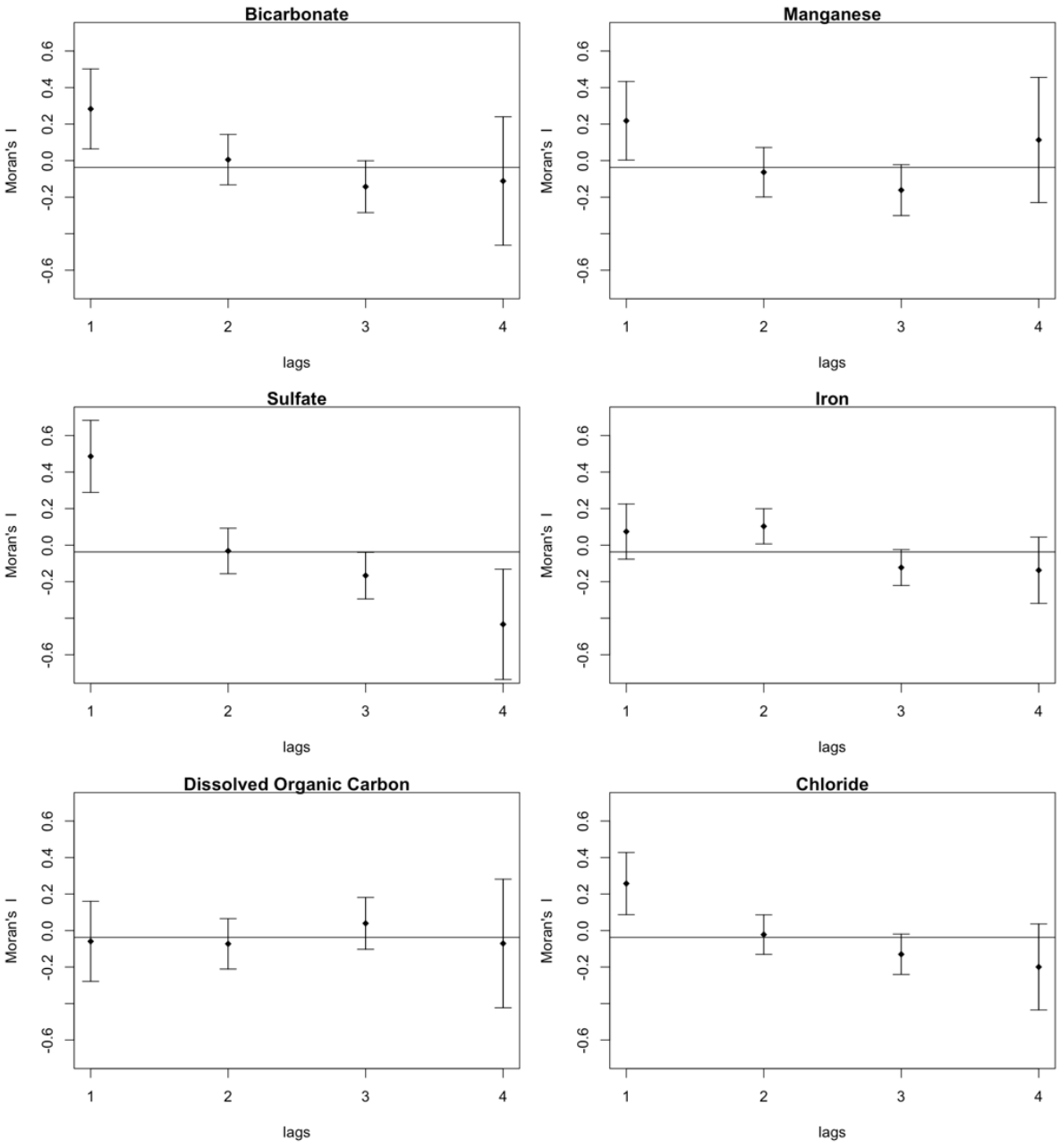


Figure 3-7 Moran's I showing the spatial correlation of some major groundwater components; bicarbonate, manganese, sulfate, chloride and iron, but not for dissolved organic carbon.

3.19 Supplemental Information

The Chalk River Laboratories Study Site

The CRL site is located about 200 km northwest of Ottawa, Ontario. It is situated around the northern margin of the Ottawa-Bonnechere graben (1). Boreholes were drilled into the rock mass to assess the local geological features and then they were cased and sealed to isolate discrete fracture zones and faults to allow an assessment of the hydrogeology of the site (2, 3). The location of the site and the locations of the boreholes included in the study are shown in Figure 1. The site boundaries are formed in part by the Mattawa fault (Ottawa River) and the Maskinonge Lake fault, also shown in Figure 1. These faults may isolate groundwater into different zones across the area of study. The site is underlain by gneisses forming stacked assemblages consisting of an overlying and underlying garnet-poor assemblage (granitic and grandioritic gneiss) and a central garnet-rich assemblage (monzonitic gneiss and quartzfeldspathic gneiss) (3). Fracture infilling materials consist of magmatic intrusions (including diabase dykes of the Grenville mafic dyke swarm) and a range of low-temperature minerals dominated by chlorite and calcite as well hematite, clays and other minerals, depending upon the host rock (4). The groundwater age coincides with the opening of the North Bay outlet when overflow was diverted into the Ottawa Valley about 10,500 BP (before present)) (5).

References

1. **Kay GM.** 1942. Ottawa-Bonnechere Graben and Lake Ontario homocline. Geological Society of America Bulletin **53:**585-646.

2. **Sikorsky RI, Thivierge RH, Siddiqui J.** 2011. Geologic Characterization of the Deep Gneissic Bedrock at Chalk River Laboratories (Ontario) using Oriented Drill Core and Integrated Borehole Surveys, Canadian Nuclear Society, Waste Management, Decommissioning and Environmental Restoration for Canada's Nuclear Activities, Current Practices and Future Needs. Canadian Nuclear Society, Toronto, Canada.

3. **Thompson P, Baumgartner P, Beaton ED, Chan T, Kitson C, Kozak E, Man A, Martino JB, Sharp K, Stroes-Gascoyne S, Thivierge RH.** 2011. An Investigation of the Suitability of the Chalk River Site to Host a Geologic Waste Management Facility for AECL's Low and Intermediate Level Wastes., Waste Management, Decommissioning and Environmental Restoration for Canada's Nuclear Activities, Current Practices and Future Needs. Canadian Nuclear Society, Toronto, Canada.

4. **Dugal JJB, Kamineni DC.** 1989. Lithology, Fracture Intensity, and Fracture Filling of Drill Core From Chalk River Research Area, Ontario. *In* Thomas MD, Dixon, D. F. (ed.), Proceedings of a Workshop On Geophysical and Related Geoscientific Research At Chalk River, Ontario.

5. **Karrow PF, Anderson TW, Clarke AH, Delorme LD, Sreenivasa MR.** 1975. Stratigraphy, paleontology, and age of Lake Algonquin sediments in southwestern Ontario, Canada. Quaternary Research **5:**49-87.

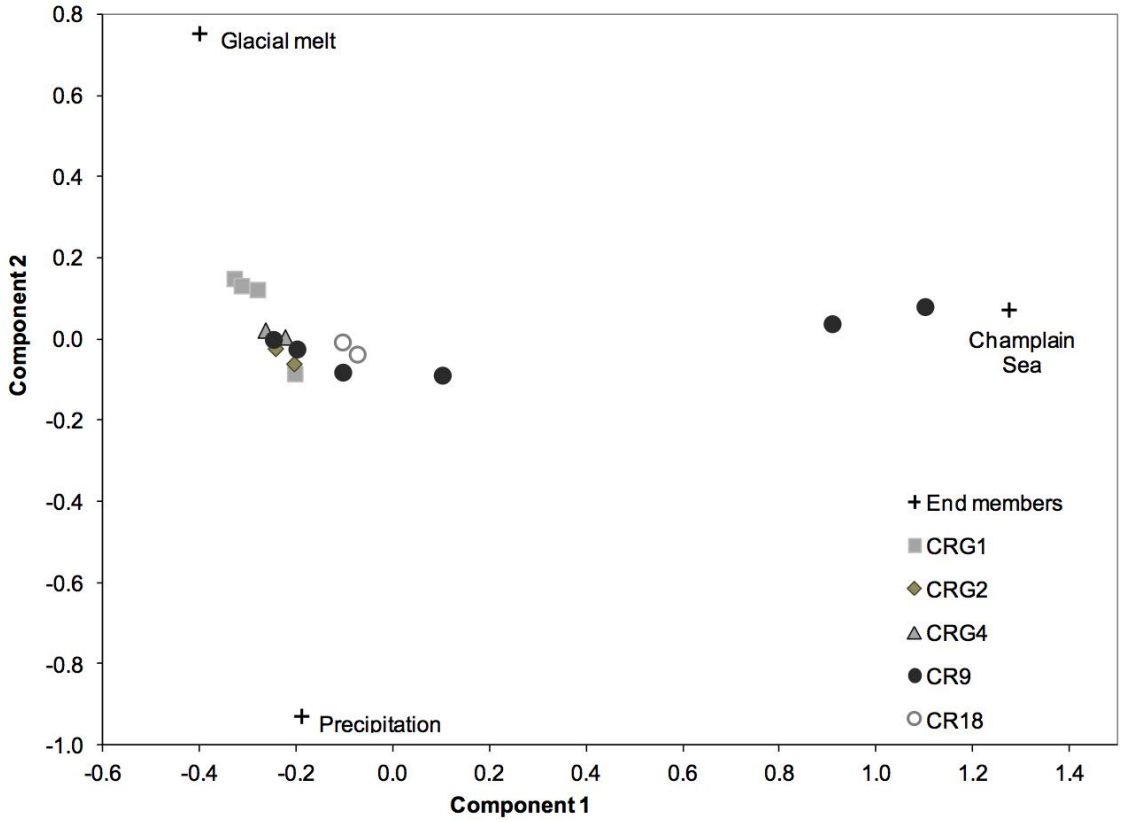


Figure S1. Principle component analysis based on major ions and stable isotopes of oxygen and hydrogen. From the Multivariate, Mixing and Mass Balance (M3) Modeling of fracture zones. CR-9 (filled circles), CR-18 (open circles), CRG-1 (squares), CRG-2 (diamonds) and CRG-4A (triangles). Compositional end members (crosses) are for glacial melt water, Champlain Sea water and modern precipitation

Table S1. Estimated concentrations[a] of bacterial 16S rRNA genes and cells (x10^5 per mL) in sampled groundwater.

Sample	Gene copies		Direct cell counts	
CRG1.3	84.02	(0.51)	1.30	(0.08)
CRG1.6	38.23	(5.77)	2.44	(0.48)
CRG1.8	65.54	(1.42)	1.77	(0.49)
CRG-1.14	69.36	(12.05)	4.78	(1.56)
CRG2.2	9.97	(0.50)	3.68	(0.34)
CRG2.4	10.15	(0.36)	1.14	(0.52)
CRG4A.9	126.11	(9.17)	11.0	(1.71)
CR9.2	2.33	(0.14)	0.55	(0.02)
CR9.5	3.04	(0.29)	0.52	(0.03)
CR9.12	4.27	(0.60)	0.67	(0.10)
CR18	73.25	(88.71)	1.94	(0.34)

[a] mean values (n = 3), standard deviation in parentheses

Table S3. Taxa Richness and Phylogenetic diversity (PD) within sampling locations.

Sampling Location		Taxa Richness	Phylogenetic Diversity, PD			
Borehole	Interval		Observed	Null ± sd	z-value	p-value
CR-18	--	221	14.7	17.5±0.4	-6.4	<0.001
CR-9	2	319	22.5	23.2±0.4	-1.7	<0.034
	5	218	16.2	17.4±0.4	-2.5	<0.001
	12	223	15.9	17.7±0.4	-4.2	<0.001
CRG-1	3	221	13.5	17.6±0.4	-9.3	<0.001
	6	253	15.8	19.5±0.4	-8.5	<0.001
	8	224	13.8	17.8±0.4	-9.3	<0.001
	14	195	11.8	16.0±0.4	-9.8	<0.001
CRG-2	2	262	15.4	20.0±0.4	-11	<0.001
	4	262	15.8	20.0±0.4	-9.6	<0.001
CRG-4A	9	180	11.3	15.0±0.4	-9.1	<0.001

Table S4. Mean pairwise distance (MPD) within sampling locations

Sampling Location		Mean Pairwise Distance (MPD				
Borehole	Interval	Observed	Null ± sd	z-value	p-value	NTI
CR-18	--	0.25	0.26 ± 0.02	-0.5	0.18	0.5
CR-9	2	0.29	0.28 ± 0.01	1.0	0.81	-1.0
	5	0.24	0.26 ± 0.02	-1.0	0.09	1.0
	12	0.26	0.27 ± 0.02	-0.5	0.23	0.5
CRG-1	3	0.22	0.23 ± 0.02	0.5	0.57	-0.5
	6	0.25	0.27 ± 0.02	-2.0	0.02	2.0
	8	0.24	0.26 ± 0.02	-1.0	0.07	1.0
	14	0.24	0.24 ± 0.02	0.0	0.47	0.0
CRG-2	2	0.25	0.24 ± 0.02	0.5	0.72	-0.5
	4	0.28	0.27 ± 0.02	0.5	0.73	-0.5
CRG-4A	9	0.24	0.26 ± 0.02	-1.0	0.19	1.0

Table S5. Mean nearest taxon distance (MNTD) within sampling locations

Sampling Location		Mean Nearest Taxon Distance, MNTD				
Borehole	**Interval**	**Observed**	**Null ± sd**	**z-value**	**p-value**	**NTI**
CR-18	--	0.09	0.12±0.02	-1.5	0.069	1.5
CR-9	2	0.08	0.11±0.01	-2.4	0.006	2.4
	5	0.07	0.12±0.02	-2.0	0.005	2.0
	12	0.07	0.12±0.01	-3.1	0.001	3.1
CRG-1	3	0.04	0.12±0.03	-2.9	0.001	2.9
	6	0.06	0.11±0.01	-3.6	0.001	3.5
	8	0.04	0.12±0.02	-3.8	0.001	3.8
	14	0.04	0.12±0.03	-3.1	0.001	3.1
CRG-2	2	0.06	0.12±0.01	-4.1	0.001	4.1
	4	0.05	0.11±0.02	-3.2	0.001	3.2
CRG-4A	9	0.05	0.12±0.03	-2.2	0.001	2.2

Relative Influences of Ecological Processes on the Meta-Community

Table S6. β Nearest Taxon Index distances between paired sampling locations

Borehole	CR18	CR9	CR9	CR9	CRG1	CRG1	CRG1	CRG1	CRG2	CRG2
Interval		2	5	12	14	3	6	8	2	4
CR9.2	3.84									
CR9.5	0.65	2.22								
CR9.12	2.41	-0.06	1.06							
CRG1. 14	0.03	4.90	3.28	1.36						
CRG1. 3	2.62	2.37	3.48	1.2	-1.39					
CRG1. 6	1.67	3.39	2.5	0.64	0.91	-1.46				
CRG1. 8	1.67	4.01	2.28	0.58	0.52	-2.39	1.51			
CRG2. 2	1.67	2.95	-0.46	-0.83	0.01	-1.96	0.03	-1.57		
CRG2. 4	2.26	3.52	-0.06	-0.97	-0.31	-0.95	2.28	-0.76	1.01	
CRG4A. 9	2.09	5.78	0.16	0.54	-1.11	-2.33	1.28	-0.95	-0.52	-1.54

Table S7. Raup-Crick Bray-Curtis, RC_{bray}, distances between paired sampling locations

Borehole	CR18	CR9	CR9	CR9.	CRG1	CRG1	CRG1	CRG1	CRG2	CRG2
Interval		2	5	12	14	3	6	8	2	4
CR9.2	1.00									
CR9.5	1.00	1.00								
CR9.12	1.00	1.00	0.04							
CRG1.14	1.00	1.00	1.00	1.00						
CRG1.3	1.00	1.00	1.00	1.00	0.60					
CRG1.6	1.00	1.00	1.00	1.00	0.82	0.87				
CRG1.8	1.00	1.00	1.00	1.00	0.46	0.59	0.56			
CRG2.2	1.00	1.00	1.00	1.00	0.98	1.00	1.00	0.96		
CRG2.4	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.99	0.48	
CRG4A.9	1.00	1.00	1.00	1.00	0.92	0.74	1.00	1.00	1.00	1.00

	βNTI	RC_{bray}
Selection (variable)	> +2.0	--
Selection (homogeneous)	< -2.0	--
Dispersal (limiting, drift)	null	> +0.95
Stochastic	null	null
Dispersal (homogenising)	null	< -0.95

Modelling the Meta-Community to Identify Significant Environmental and Spatial Variables

S1. Spatial variables explain taxa abundances of the meta-community

standardize the independent variables for modeling and calculate the scores of the principle components (PCA) for the environmental and spatial variables.

```
env.mem.stand=decostand(env.mem, method = "standardize")
pca.mem = principal(env.mem.stand, nfactors = 6,rotate = "none", covar = T, scores = TRUE)
scores = pca.mem$scores # axes scores for future use
```

perform a redundancy analysis (RDA) comparing the Hellinger transformed taxa abundances with the PCA scores for the environmental and spatial variables.

```
rda.all.mem = rda(sp.hel, scores)
anova.cca(rda.all.mem)  # to check the overall explanatory power of pca axes

        Permutation test for rda under reduced model
        Permutation: free
        Number of permutations: 999

        Model: rda(X = sp.hel, Y = scores)
              Df Variance      F Pr(>F)
        Model     6  0.43388 1.6952  0.017 *
        Residual  4  0.17063
        ---
        Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

        r2.mem = RsquareAdj(rda.all.mem)
        r2.mem
                $r.squared
                [1] 0.7177373

                $adj.r.squared
                [1] 0.2943433  ##Global Adjusted R²
```

#spatial and environmental variables are significant, with a global adjusted R^2 of 29.4%.
create a reduced model to identify the significant explanatory variables.

```
mod0=rda(sp.hel~1,env.mem)  #model with intercept only
mod1=rda(sp.hel~DOC+HCO3+sulfte+iron+manganese+MEM1+MEM2+MEM3,env.mem)
#model with all explanatory variable

#Forward selection
sp.model.mem=ordistep(mod0, scope = formula(mod1), direction="forward", perm.max = 999)
anova(sp.model.mem)
        Permutation test for rda under reduced model
        Permutation: free
```

```
Number of permutations: 999

Model: rda(formula = sp.hel ~ MEM1 + MEM2, data = env.mem)
        Df Variance     F Pr(>F)
Model    2  0.24644 2.753  0.001 ***
Residual  8  0.35807
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

calculate adjusted R^2 for the reduced model

```
step.res.mem <- ordiR2step(mod0, scope = formula(sp.model.mem), direction="forward")
step.res.mem$anova  # Summary table

                R2.adj Df     AIC      F Pr(>F)
+ MEM1         0.20367  1 -6.2493 3.5577  0.002 **
<All variables> 0.25959
```

the spatial variables, MEM1 and MEM2 account for 25.9% of the adjusted R^2.
the spatial variable, MEM1 accounts for 20.4% of the variation. No environmental variables were identified as significant.

```
library(ade4)
library(adegraphics)
library(adespatial)
library(psych)
library(simba)
library(fossil)
library(vegan)
library(psych)
library(picante)


#load phy and com or build phy and comm form seq file
#match phy and comm in picante
#perform ses.pd and ses.mntd in picante
#perform bNTI calc and rcbray calc with Stegen's code
#normalize bNTI and RCbray -- Stegen code
#quantitative ecology distance decay calculation, with long lat
#GuniFrac has code for weighted Unifrac


#comm file tranfored with decostand specifying hellinger transformation == sp.hel
#env.mem file stadnardised by deconstand specifying stand == env.mem.stand

nbtri <- tri2nb(as.matrix(x.y[1:12,]))##Long Lat neighbour matrix
lwB <- nb2listw(nbtri, style = "B")  ##listw object
lwW <- nb2listw(nbtri, style = "W")  ##listw object, weighted
scB <- mem(lwB)
scW <- mem(lwW)
moran.bounds(lwB)


moran.randtest(env.mem[,#], listw, nrepet = 999
moranNP.randtest(x, listw, nrepet = 999, alter = c("greater", "less", "two-sided"), ...)
scores.listw(lwW, wt = rep(1, length(lwW$neighbours)),MEM.autocor ="positive")
positiveI=scores.listw(lwW, wt = rep(1, length(lwW$neighbours)),MEM.autocor ="positive")
negativeI=scores.listw(lwW, wt = rep(1, length(lwW$neighbours)),MEM.autocor ="negative")

env$MEM1=positiveI$MEM1
env$MEM2=positiveI$MEM2
env$MEM3=positiveI$MEM3
env.mem=env[,c(1:5,11:15)]


env.mem.stand=decostand(env.mem[,-6], method = "standardize")
pca.mem = principal(env.mem.stand, nfactors = 6,rotate = "none", covar = T, scores = TRUE)
```

```
scores = pca.mem$scores  # axes scores for future use
rda.all.mem = rda(sp.hel, scores)
anova.cca(rda.all.mem)  # to check the overall explanatory power of pca axes

r2.mem = RsquareAdj(rda.all.mem)
r2.mem$r.squared
r2mem$adj.r.squared


mod0=rda(sp.hel~1,env.mem)  #model with intercept only
mod1=rda(sp.hel~.,env.mem) #model with all explanatory variables, env and mem
sp.model=ordistep(mod0, scope = formula(mod1), direction="forward", perm.max = 999)
anova(sp.model)
step.res<- ordiR2step(mod0, scope = formula(sp.model), direction="forward")
RsquareAdj(step.res)



#repeat with env only, or with mem only

mod1.mem=rda(sp.hel~MEM1+MEM2+MEM3.....etc.   ,env.mem),
sp.model.mem=ordistep(mod0, scope = formula(mod1.mem), direction="forward", perm.max = 999)
anova(sp.model.mem),
step.res.mem <- ordiR2step(mod0, scope = formula(sp.model.mem), direction="forward", perm.max =
999),
anova(step.res.mem), # Summary table
step.res.mem$anova,


###


mod1.env=rda(sp.hel~env1+env2.....etc.   ,env.mem)
sp.model.env=ordistep(mod0, scope = formula(mod1.env), direction="forward", perm.max = 999)
anova(sp.model.env)
step.res.env <- ordiR2step(mod0, scope = formula(sp.model.env), direction="forward", perm.max = 999)
anova(step.res.env)  # Summary table
step.res.env$anova
```

4. THE INFLUENCES OF TAXA DIVERGENCE, DISPERSAL AND DRIFT ON MICROBIAL CO-OCCURRENCE AND ABUNDANCE PATTERNS WITHIN SUBSURFACE CRYSTALLINE FRACTURES[5]

4.1 Abstract

Microbes catalyze important biogeochemical cycles that can affect the speciation and transport of radionuclides. Safety assessments of geologic repositories rely on predictions of *in situ* microbial processes but the inaccessibility of the terrestrial subsurface means that this habitat is difficult to explore.

An opportunity to study a crystalline subsurface to depths of up to 900 m allowed for analysis of 16S rRNA gene sequence patterns using an ecological framework of dispersal, drift and diversification.

Within an extent relevant to a repository, the observed patterns reflect possible connectivity with the surface and dispersal by fracture flow. Transport of taxa between connected subsurface sampling locations occur at distances of up to 500 m.

Phylogenetic clustering and a high proportion of sister taxa reflect the influence of diversification on taxa adapted to the subsurface habitat.

The influences of limited dispersal and drift would be strongest between unconnected sampling locations. Drift, therefore, may account for the unique phylogenetic diversity represented within each assemblage. After colonization, a loss of diversity would occur by local random extinction of low abundant taxa.

Under this ecological framework, by knowing the influences of dispersal, drift and diversification, evaluation of selection on the resulting 16S rRNA gene patterns may be possible.

[5] Adapted from
Danielle Beaton, Bradley S. Stevenson, Karen J. King-Sharp, Blake W. Stamps, Heather S. Nunn, Marilyne Stuart, and Jack Cornett, 2018, The Influences of Taxa Divergence, Dispersal and Drift on Microbial Co-Occurrence and Abundance Patterns within Subsurface Crystalline Fractures, submitted to ISME J December 22, 2017.

4.2 Introduction

Microorganisms make up a large proportion of the Earth's biomass (Kallemeyer et al., 2012), including the terrestrial subsurface, and catalyse important biogeochemical cycles. Microbial communities also possess a range of catalytic activities that can be important for overall ecosystem functioning; even so, the nature of these functions and the connection between function and community are limited (Graham et al., 2016; Bier et al., 2015; Griebler and Lueders, 2009). The biodiversity and function of microorganisms that inhabit the terrestrial subsurface are less studied than other, more accessible, habitats. Given the adoption of underground facilities for nuclear waste management in many countries, including Canada, the catalytic activities that are possible within a subsurface hosting a repository means that the microbiology of the formation could influence the speciation and transport of the emplaced radionuclides over the duration of the geological confinement. Studies performed on fracture water from the Fennoscandian and Canadian Shields reveal metabolism and taxa with known relatives associated with nitrogen, iron and sulfur cycling (Jain et al., 1997; Itävaara et al., 2011; Konno et al., 2013; Pedersen, 2013; Mcmahon and Parnell, 2014; Purkamo et al., 2014, Beaton et al., 2016; Hubalek et al., 2016; Itävaara et al., 2016; Küsel et al., 2016; Purkamo et al., 2016; Wu et al., 2016). The broad redox range encompassed by these cycles raises a concern that over long residence times the subsurface microbial processes will enable the transport of aqueous and gaseous hazardous materials by groundwater and gas migrations.

The effect of neutral (Hubbell, 2001) and random ecological processes -- like dispersal, drift and diversification -- on community assembly can complicate predictions of the selective forces acting within an ecosystem. The ecological framework proposed by Vellend (Vellend, 2010; Vellend, 2016) can account for random processes on community patterns by recognizing that microbial taxa can occur in unvarying environmental conditions (with corresponding low rates of community turnover), and in variable environmental conditions (with corresponding high rates of community turnover). Depending on the concomitant rate of taxa dispersal between a pair of sampling locations, the homogenizing effect of a high dispersal rate can confound identification of the influences on biodiversity derived from a variable environment, even when selection pressure is strong (Vellend, 2010). When the dispersal rate between communities is

low, or limited by the flow restricted fractures of a crystalline subsurface, the dynamics of biodiversity turnover between communities can be subject to drift and random variations in population even when selection pressure is weak (Vellend, 2010). This last combination of influences can result in compositional variation between communities (Adler et al., 2007) and such a situation would make it difficult to link subsurface assemblages with repository-scale ecosystem processes.

Empirically, microbial communities may be comprised of hundreds or even thousands of taxa, although most taxa are present in low relative abundances (Nemergut et al., 2013). These communities also display phylogenetic clustering – or co-occurrence of related taxa (Horner-Devine and Bohoman, 2006; Goberna et al., 2014; Beaton et al., 2016) – and decay of similarity with distance (Martiny et al., 2011; Horner-Devine et al., 2004; Lear et al., 2014) spanning meters to kilometers. These community characteristics occur even within a continuous volume of water lacking barriers to dispersal (Horner-Devine et al., 2004; Lear et al., 2013; Lear et al., 2014).

In this study, we applied Vellend's ecological framework (Vellend, 2010; Vellend, 2016) to study 16S rRNA gene patterns from spatially distributed fracture water accessed by boreholes drilled into the crystalline subsurface underlying the site at Chalk River Laboratories (CRL), Ontario, Canada. We compared the relative contributions of the random process of dispersal, drift and diversification on the resulting 16S rRNA gene patterns.

4.3 Materials and Methods

Figure 4-1 shows a map of the borehole locations within the study site. The site is located within the Central Grenville Gneissic Belt within the Grenville province, the rock dates from 1475 ± 14 to 1045 ± 6 Ma. Two major faults bound the study site: the Mattawa fault (Ottawa River) and the Maskinonge Lake fault. Diabase dykes traversing the site form another boundary. Figure 4-1 also shows the three boreholes used for sampling fracture water. Borehole CRG-3 traverses one of these dykes, borehole CRG-6 traverses two dykes and borehole CRG-9 is wholly located between these dykes. Stacked crystalline rock underlies the

site and consists of an overlying and underlying garnet-poor region, and a central garnet-rich region.

4.3.1 Sampling Fracture Water

Figure S1 shows a schematic of a borehole with an installed Westbay system®. We sampled four locations (depths) per borehole using a Westbay Multilevel Groundwater Monitoring System (Nova Metrix Groundwater Monitoring (Canada) Ltd) as previously described (Beaton et al., 2016). Opening the filled tubes and dispensing the sampled fracture water took place inside of a glove box under an atmosphere of filtered nitrogen gas. Inside the glovebox, the groundwater pH (Beckman PHI 265 pH/Temp/mV meter (Beckman Coulter, Inc.)) and conductivity (YSI Model 30 Conductivity Meter (YSI Inc., Yellow Springs, OH, USA)) were measured. For elemental analysis , aliquots of 100 mL were filtered through a 0.45 µm filter (isopore polycarbonate, Merck Millipore, Billerica, MA, USA) and immediately acidified by adding 0.5 mL nitric acid (ultra-trace grade, SeastarTM, Baseline®, Fisher Scientific, Ottawa, ON, Canada). Elemental composition of the groundwater was determined by inductively coupled plasma-mass spectrometry (ICP-MS), using either a Varian 820-MS (Agilent Technologies, Inc.) or an Element XR (Thermo Scientific) and by inductively coupled plasma atomic emission spectroscopy (ICP-AES, Optima 3300, Perkin Elmer). Anion concentrations were determined using a Dionex 3000 ICS ion chromatograph (Dionex, Sunnyvale, CA, USA). Dissolved organic (DOC) and inorganic carbon (DIC) were determined using a Dohrmann, model Phoenix 8000-UV Persulfate TOC Analyzer (Teledyne Teckmar, Mason, OH, USA). The excitation/emission of fluorescein in the groundwater provided a measure of residual drill water.

4.3.2 Characterization of Bacterial Assemblages

4.3.2.1 Nucleic Acid Extraction and Creation of 16S rRNA Gene Libraries

Whole community nucleic acids were extracted using the UltraClean® DNA Isolation Kit (Qiagen). Within 2 hrs of sampling, fracture water samples were filtered each through a sterile 0.22 µm, 47 mm polyethersulfone filter within a sterile 100 mL filter housing (Merck Millipore Corp., Billerica, MA); filters were either processed immediately or stored overnight at -20°C and processed the next day. The volume filtered for DNA extraction was between 0.8 - 1 L.

4.3.2.2 16S rRNA Library Preparation and Sequencing

A 16S rRNA gene library was amplified by PCR from each DNA extraction. The primers used in the initial amplification generated amplicons that spanned the V4 region of the 16S rRNA gene between position 519 and 802 (*E. coli* numbering), and produced a PCR amplicon ~300 bp in length. The forward primer (M13-519F: 5'- **GTA AAA CGA CGG CCA G**CA CMG CCG CGG TAA -3') contains the M13 forward primer (in bold), followed by a spacer and the 16S-specific sequence (underlined). The reverse primer (785R: 5'-TAC NVG GGT ATC TAA TCC-3') was taken directly from the primer "S-D-Bact07850b-A-18" in Klindworth et al. (2013). Each 25 µL PCR consisted of 1x Taq buffer (Fermentas, Glen Burnie, Maryland), 1.5 mM $MgCl_2$, 0.2 mM each dNTP, 0.2 µM of the forward and reverse primer, 0.625 U of Taq DNA polymerase (Fermentas) and 2 µL of extracted DNA. Thermal cycling was carried out in a Techne TC-512 thermal cycler (Techne, Burlington, NJ) using the following conditions: initial denaturation for 3 min at 96°C; 30 cycles of 30 s at 96°C, 45 s at 52°C and 45 s at 72°C; and a final extension for 10 min at 72°C.

Duplicate PCRs for samples were combined and purified using Ampure® XP paramagnetic beads (Beckman Coulter, USA) according to manufacturer's protocol. A second, 6 cycle PCR was used to add a unique 12 bp barcode (Hamady et al., 2008) to each amplicon library using a unique forward primer containing the barcode+M13 forward sequence (5'-3') and the 16S reverse primer, 785R. The resulting barcoded PCR products were purified using Ampure® XP paramagnetic beads (Beckman Coulter), quantified using the Qubit® HS assay (Life Technologies, Carlsbad, CA, USA), pooled in equimolar amounts, and concentrated to a final volume of ≈80µL using two Amicon® Ultra-0.5 mL 30K Centrifugal Filters (Merck Millipore). The final pooled library was then submitted to the Oklahoma Medical Research Foundation Genomics Facility (Oklahoma City, OK, USA) for sequencing on the MiSeq platform (Illumina Inc., San Diego, CA) using PE250 V2 chemistry. All sequences have been deposited in the short read archive of GenBank under Accession No. SRA608335.

4.3.2.3 Sequence Analysis

After sequencing, reads were merged using Pair-End reAd merger (PEAR) (Zhang et al., 2014), de-multiplexed in Quantitative Insights Into Microbial Ecology (QIIME) (Caporaso et al., 2010b),

and clustered into operational taxonomic units (OTUs) using UPARSE (Edgar, 2013) at a cut-off of 97% sequence similarity. After de-replication in UPARSE, taxonomy was assigned using the RDP Naïve Bayesian classifier (Wang et al., 2007) and the SILVA database (Release 111) (Pruesse et al., 2007). After OTU clustering in UPARSE, OTUs were aligned with pyNAST (Caporaso et al., 2010a) against an aligned version of the SILVA r111 database, and filtered to remove uninformative bases. Finally, a tree was produced with FastTree (Price et al., 2010) for beta diversity analyses. Beta diversity was estimated using weighted and unweighted UniFrac index (Lozupone and Knight, 2005).

4.3.3 Diversity and Co-Occurence

The BIOM file produced by QIIME was imported into the Bioconductor package phyloseq 1.14.0 (McMurdie and Holmes, 2013). A plot of taxa abundances was generated using code provided within McMurdie and Holmes (2013) sourced within the R global environment (R Core Team, 2015) as plot_abundance(). Heatmaps of the distances between the top 50 most abundant taxa and of the abundances of taxa subset by phylogenetic classification were created to display the distribution and relatedness of the fracture water taxa. Normalization of abundances was by the regularized log transformation using functions within the DESeq2 1.10.1 package (Love et al., 2014).

Diversity metrics were calculated using functions with the R package 'picante' 1.6-2 (Kembel et al., 2010) and 'GUniFrac' 1.0 (Chen et al. 2012). These were Faith's phylogenetic diversity and phylogenetic diversity by the mean pairwise distance (MPD) and by the mean nearest taxon distance (MNTD). To account for phylogenetic diversity scaling with taxa richness, we used the standardized effect size with abundance.weighted = TRUE. For each metric the null model 'taxa.labels' was set at 999 randomizations. Faith's PD (Faith, 1992) was also partitioned into total PD and average PD within assemblages (PD_{α}); the difference between them accounted for the average PD that was not present within assemblages (PD_{β}). To calculate MPD and MNTD, a pair-wise distance matrix of the meta-community tree was first calculated using the function cophenetic().

Distance decay of similarity was determined by converting the geographic coordinates between sampling locations to geographic distance using earth.dist() from the R package 'fossil'. The community diversity metrics were then plotted against geographic distance and the data fit using the function scatter.smooth(). The significance of the distance decay of similarity were tested by Procrustes (repetitions = 999)

Sister pairs of taxa were also identified within the phylogeny by using functions within the R packages 'ape' 4.1 and 'geiger' 2.0.6. The tree was first converted to an ultrametric tree using the default value for the smoothing parameter, lambda, to calculate the tree and node ages. These age values were then used to estimate the time of divergence between the sister nodes.

4.4 Results

Figure 4-2 shows the distributions of taxa by Phylum and Class for each assemblage. Taxa prevalence and abundances were derived from over 132000 high quality reads that were retained after processing for quality and removal of chimeric sequences. Four hundred and thirty five OTUs were detected among all samples at a 97% sequence similarity. Thirty-two OTUs (7.4%) were designated as unclassified at the rank of Phylum. The abundant bacteria and Archaea in this study were composed of multiple of OTUs, as shown in Figure 4-2. Prevalent and abundant OTUs were from the *Proteobacteria*, *Firmicutes*, *Bacteroidetes*, *Chloroflexi*, *Acidobacteria* and *Actinobacteria*. The *Nitrospirae* were abundant but less prevalent; the *Planctomycetes* and *Cyanobacteria* were prevalent but less abundant. The Archaea included members from the *Thermoplasmata*, *Halobacteria*, *Methanobacteria* and the *Miscellaneous Crenoarchaeatoa Group*.

4.4.1 Local Patterns of Diversity

Diversity patterns were determined using the mean pairwise phylogenetic distance and the mean shortest phylogenetic distance for each in the distance matrix. The standardized effect size compares the measured values to randomly generated null mean values. We were, therefore, able to separate the co-variance of diversity and taxa richness from patterns of phylogenetic dispersion. Both the mean pairwise distance (MPD, Figure 4-3, top panel) and the mean nearest taxon distances (MNTD, Figure 4-3, bottom panel) show that all but one of the

assemblages were more than two standard deviations lower than the corresponding null distribution of the meta-community (red line in Figure 4-3, top panel for MPD and bottom panel for MNTD, respectively). The one assemblage that was within 2 standard deviations of the null distribution for MPD (indicted by an open triangle) had an effect size of 1.97, or only just within 2 standard deviations of the null mean. The assemblages, therefore, cluster phylogenetically and display co-occurrence of related taxa compared to the meta-community. The corresponding z-values of the standardized effect size for phylogenetic diversity, PD, are all negative (Table S2) and all but two of the PD z-values are more than two standard deviations lower than the null mean. Partitioning PD into total PD and average PD per assemblage indicates the meta-community represents more diversity than what is represented within each assemblage, PD_{α},Table S3, with approximately half the total PD not represented within the assemblages (PD_{β}), (Table S3).

These differences in local diversity may reflect selection by local conditions (Horner-Devine and Bohannan, 2006); however, 124 sister pairs occur within the 403 identified taxa and these pairs suggest divergence, or speciation, which is a regional process within the ecological framework (Vellend, 2016). Table S5 lists the identities of the first taxon of each pair; these included members for the abundant taxa listed above, *Firmicutes*, *Bacteroidetes*, *Proteobacteria, Chloroflexi*, *Acidobacteria, Actinobacteria, Nitrospirae, Cyanobacteria* and a few *Archaea*.

4.4.2 Distance Decay of Similarity and Beta-Diversity

Both taxonomic similarity (1-Bray-Curtis), and phylogenetic similarity (1-UniFrac), display similar decay with increased distance, when plotted along the geographic distances between sampling locations up to 0.5 km (Figure 4-4). A Procrustes analysis shows the decay with distance was significant based on UniFrac ($p = 0.002$), but was not significant based on the abundances ($p = 0.932$). The UniFrac β-diversity calculates the mean shortest phylogenetic distances (Lozupone et al., 2011) detecting uniqueness associated with closest relatives.

Procrustes analyses of additional phylogenetic metrics showed p-values for Dnn (akin to the mean nearest taxon distance, MNTD) of 0.007 and Dpw (akin to the mean pairwise distance, MPD) of 0.161, respectively (Table S4). A principle coordinate analysis of weighted UniFrac

distances (Figure 4-5) shows that assemblages cluster by borehole and suggests the decay of similarity, and dispersal, occurs within vertically orientated fractures – those fractures transected by boreholes. The dashed lines shown in Figure 4-4 refer to the distances between the shallowest and deepest sampling locations within each borehole. The decay of similarity with distance (Figure 4-4) occurs at distances less than 0.5 km.

4.4.3 Co-occurrence Patterns across Sampling Locations

Figure 4-6 shows the co-occurrence and distances of the top 50 abundant taxa across assemblages. The taxa abundances were first variance-stabilized by the regularized log transformation to minimize the influence of a few variable taxa. The heatmap shows that, among the top 50 taxa, some are common across all sampling locations while others cluster mainly within one borehole. The diversity present within borehole CRG-9 is not present within borehole CRG-3 or borehole CRG-6.

Heatmaps of taxa abundances provide additional insight into the co-occurrence patterns of taxa from within the same Family (Figures S2-S6). The distributions and regularized abundances of the *Deltaproteobacteria* and *Clostridia* (Figure S2) and of the *Betaproteobacteria* and *Nitrospirae* (Figure S3) are examples of taxa with known relatives that are involved in sulfur and nitrogen metabolism, respectively. We focus on these examples because both nitrogen and sulfur compounds were measured in extracted rock porewater (Peterman et al., 2016), while the fracture water had measurable sulfate but undetectable ammonia, nitrite or nitrate (Table S1). The distribution of taxa associated with sulfate reduction -- taxa within the *Clostridia* and the *Deltaproteobacteria* -- have some members within borehole CRG-9 and other members within borehole CRG-3 or borehole CRG-6 (Figure S2). The distribution of taxa associated with the nitrogen metabolism – taxa within the *Betaproteobacteria*, and the *Nitrospirae*, differ in co-occurrence patterns in that the *Betaproteobacteria* are distributed across each of the sampling locations, while the *Nitrospirae*, are prevalent and diverse within borehole CR-9 but are either not detected or are in low abundance within borehole CRG-3 and borehole CRG-6 (Figure S3). In association with these taxa are taxa associated with nitrogen fixation – such as the prevalent and abundant Clostridia (Figure 4-2, Figure 4-6) and the prevalent but low abundance *Cyanobacteria* (Figure 4-2). The Archaea also show a distribution

pattern within fracture water from borehole CR-9 in which several abundant taxa were related to deep-sea hydrothermal vent systems occur within fracture water from borehole CR-9 (Figure S6).

4.5 Discussion

Our findings suggest that ecologically similar surface taxa invaded the subsurface via fracture flow. We found that each assemblage is phylogenetically clustered (Figure 4-3) and that prevalent and abundant taxa across assemblages are from within the same taxonomic classes (Figure 4-2a). The overall phylogenetic diversity of the meta-community represents more diversity than is found within any one of the assemblages (Table S3). Close relatives accounted for 61% of the identified taxa. These pairs contribute to both the phylogenetic clustering within assemblages (Figure 4-3) and the 'missing' phylogenetic diversity, or diversity not present within individual assemblages that made up the regional community. The distance decay of similarity (Figure 4-4) and the pattern of assemblage clustering seen in the principle coordinate analysis (Figure 4-5) suggests that dispersal into fractures and transport of taxa between connected sampling locations is an assembly process shaping these communities.

Diversification, as a regional scale ecological process (Vellend, 2016), has also shaped these communities.

These patterns, therefore, suggest that the observed phylogenetic clustering was the result of regional processes combined with dispersal into the subsurface, and were not a result of local selection by abiotic factors such as pH or temperature. This outcome does not mean that local processes have not selected for the taxa present, but that the taxa present arrived at the subsurface locations with the needed adaptive characteristics for this habitat. Similar co-occurrence of related bacterial taxa is seen in other habitats such as in soils with increased productivity (Goberna et al., 2014B), in habitats successfully invaded by related plants that share ecological traits prior to invasion (Cadotte et al., 2009) and where there is spatial heterogeneity within a habitat that allows related taxa to co-exist (Lear, 2014).

For microbial assemblages characterized by sequence similarity of the 16S rRNA gene, clustering may be seen within assemblages that harbor taxa having multiple 16S rRNA gene

copies; an artifact created by the potential influence of intragenomic heterogeneity within this gene that could lead to an overestimation of microbial diversity (Suna et al., 2013; Pei et al., 2010). At the 97% similarity cut off, the V4-V5 region of the 16S rRNA gene from fully sequenced bacterial and archaeal genomes display the lowest intragenomic heterogeneity, and thus a lower associated risk of overestimating the number of related taxa (Suna et al., 2013). Even so, the phyla with greater numbers of gene copies, such as the *Proteobacteria* and *Firmicutes*, also had greater degrees of overestimation compared to phyla with fewer or only one gene copy (Suna et al., 2013). Some of the sister pairs, therefore, may reflect the 16S rRNA gene copy number of the taxa found within the subsurface.

A separate consideration in detecting co-occurrence of related taxa is that the metrics used to analyze assemblages can be sensitive to sampling. For example, by the data used to construct assemblages, and by the spatial grain used to define assemblages (Gonzales-Caro et al., 2012). The patterns of co-occurrence detected within the fractures, therefore, may reflect the sampling restricted by the borehole locations and depths. With these limitations in mind, the assembly of the subsurface communities appear to link to stochastic influences of surface recharge via vertical fracture flow and by horizontal dispersal limitation combined with apparent divergence of the regional taxa (Evans et al., 2017).

4.5.1 Community Structure and Possible Ecosystem Function

Microbial clades constructed from DNA sequence data may represent a compilation of ecotypes (species) whose members are ecologically similar, and so, genetic diversity within an ecotype is constrained by selection and drift (Cohan and Koeppel, 2008). An adaptive mutant arising from within an ecotype outcompetes, potentially to extinction, all other members from within the same ecotype; while members from other ecotypes are not outcompeted by the new mutant; and so at the level of genus may appear as co-occurrence of related taxa. Between periods of selection, therefore, diversity within an ecotype can increase until such a time when a new adaptive mutant arises, and ecotype diversity declines. Any lineage divergence occurring during the intervening period would be temporary unless diversification through cladogenesis had occurred (Koeppel et al., 2013) either through specialization for use of different

compounds than used by the original ancestral population (Østman et al., 2014; Treves et al., 1998), or through spatially correlated resource availability (Buckling et al., 2000).

The microbial assemblages from this study occur within fractured crystalline rock (Neymark et al., 2013). The combination of a lack of measurable nitrogen in the fracture water and the parts per million concentrations of ammonia, nitrite and nitrate in the small volume of extracted porewater (Peterman et al., 2016) suggest that there is spatial heterogeneity in resource availability. This heterogeneity further suggests that it may be possible to discern local selective processes that are separate from the 16S rRNA gene patterns created by dispersal, drift and diversification.

4.6 Conclusions

The ecological processes of dispersal drift and diversifications play major roles in the observed patterns of the subsurface community assembly. Patterns of the 16S rRNA geneV4 region found in this study reflect possible connectivity with the surface (Hubalek et al., 2016) in keeping with the meteoric source waters (Beaton et al., 2016). Transport of taxa between connected subsurface sampling locations occur at distances up to 500 m. This apparent connectivity suggests that is may be possible to discern influences of selection on the 16S rRNA gene patterns.

The processes of dispersal and diversification account for the number of lineages found across the sampling locations for a high percentage of related taxa. Diversification, as a regional scale process, would mean these taxa were pre-adapted to the subsurface habitat. The co-occurrence of multiple sister pairs suggest success by invader taxa that share ecological traits prior to invasion (Cadotte et al., 2009) into the subsurface.

The influences of limited dispersal and drift on 16S rRNA gene patterns would be strongest between unconnected sampling locations. Drift, therefore, may account for the unique local phylogenetic diversity resulting from random extinction of low abundant taxa; and thus act to lower the local phylogenetic diversity.

4.7 Author Contributions

All authors collected and contributed data sets for analysis as well as participated in the conceptual drafting and revision of this manuscript. Karen King-Sharp oversaw the groundwater sampling and geochemical analyses. Blake W. Stamps and Heather S. Nunn performed sequencing and sequence data analysis. E. Danielle Beaton conducted subsequent analyses on the sequence data and was the primary author in writing and revising the manuscript. Bradley Stevenson and Marilyne Stuart contributed significantly in manuscript development and revision.

4.8 Acknowledgements

We gratefully acknowledge the contributions from Sim Stroes-Gascoyne, Canadian Nuclear Laboratories, Pinawa, Manitoba and David Larssen, Westbay Instruments, Vancouver, British Columbia. We also acknowledge the contributions from Steve Rose, Lee Bellan, Dan Festarini, Amy Festarini, Lisa Shi, Grace Chen and the other contributors to the Geologic Waste Management Facility project, Canadian Nuclear Laboratories, Chalk River, Ontario.

4.9 Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

4.10 References

Adler, P.B., Hillerislambers, J., Levine, J.M., 2007, “A Niche for Neutrality”, Ecology Letters, 10, 95-104.

Beaton, E.D., Stevenson, B.S., King-Sharp, K.J., Stamps, B.W., Nunn, H.S., Stuart, M., 2016, “Local and Regional Diversity Reveals Dispersal Limitation and Drift as Drivers for Groundwater Bacterial Communities from a Fractured Granite Formation”, Frontiers in Microbiology, 7:1933, doi:10.3389/fmicb.2016.01933.

Bier, R.L., Bernhardt, E.S., Boot, C.M., Graham, E.B., Hall, E.K., Lennon, J.T. et al., 2015, “Linking Microbial Community Structure and Microbial Processes: An Empirical and Conceptual Overview”, FEMS Microbiology Ecology, 91: fiv113. doi: 10.1093/femsec/fiv113.

Bray, J.R., Curtis, J.T., 1957, “An Ordination of the Upland Forest Communities of Southern Wisconsin”, Ecological Monographs, 27:325.

Buckling, A., Kassen, R., Bell, G., Rainey, P.B., 2000, “Disturbance and Diversity in Experimental Microcosms”, Nature, 408, 961-964.

Cadotte, M.W., Hamilton, M.A., Murray, B.R., 2009, “Phylogenetic relatedness and plant invader success across two spatial scales”, Diversity and Distributions, 15, 481-488.

Cadotte, M.W., Davies, T.J., 2016, “Phylogenies in ecology: A guide to Concepts and Methods”, Princeton University Press, Chapter 7.

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L., Knight, R., 2010a, “PyNAST: a flexible tool for aligning sequences to a template alignment”, Bioinformatics, 26:266–267, doi: 10.1093/bioinformatics/btp636.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello et al., 2010b, “QIIME allows analysis of high-throughput community sequencing data”, Nat Methods, 7:335–336, doi:10.1038/nmeth.f.303.

Chen J., Bittinger K., Charlson E.S., Hoffmann, C., Lewis, J., Wu, G.D et al., 2012, “Associating microbiome composition with environmental covariates using generalized UniFrac distances”, Bioinformatics, 28(16):2106-2113, doi:10.1093/bioinformatics/bts342.

Cohan, F.M., Koeppel, A.F., 2008, “The Origins of Ecological Diversity in Prokaryotes, Current Biology”, 18, R1024-R1034.

Edgar, R.C., 2013, “UPARSE: highly accurate OTU sequences from microbial amplicon reads”, Nat Methods, 10:996–998, doi:10.1038/nmeth.2604.

Evans, S., Martiny, J.B.H., Allison, S.D., 2017, “Effects of Dispersal and Selection on Stochastic Assembly in Microbial Communities”, ISME Journal, 11, 176-185, doi:10.1038/ismej.2016.96.

Faith, D.P., 1992, “Conservation evaluation and phylogenetic diversity”, Biol. Conserv., 61, 1–10, doi: 10.1016/0006-3207(92)91201-3.

Goberna, M., Navarro-Cano, J.A., Valiente-Banuet, A., García, C., Verdú, M., 2014a, “Abiotic stress tolerance and competition related traits underlie phylogenetic clustering in soil bacterial communities”, Ecol. Lett., 17, 1191–1201, doi: 10.1111/ele.12341.

Goberna, M., García, C., Verdú, M., 2014b, “A role for biotic filtering in driving phylogenetic clustering in soil bacterial communities”, Global Ecology and Biogeography, 1346-1355, 10.1111/geb.12227.

González-Caro, S., Parra, J.L., Graham, C.H., McGuire, J.A., Cadena, C.D., 2012, “Sensitivity of Metrics of Phylogenetic Structure to Scale, Source of Data and Species Pool of Hummingbird Assemblages along Elevational Gradients”, PLoS One, e35472. doi: 10.1371/journal.pone.0035472.

Graham, E.B., Knelman, J.E., Schindlbacher, A., Siciliano, S., Breulmann, M., Yannarell, A et al. 2016, “Microbes as Engines of Ecosystem Function: When Does Community Structure Enhance Predictions of Ecosystem Processes?”, Frontiers in Microbiology, 7, 214, http://journal.frontiersin.org/article/10.3389/fmicb.2016.00214.

Griebler, C., Lueders, T., 2009, “Microbial Biodiversity in Groundwater Ecosystems”, Freshwater Biology, 54, 649–677, doi:10.1111/j.1365-2427.2008.02013.x.

Hamady, M., Walker, J.J., Harris, J.K., Gold, N.J., Knight, R., 2008, “Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex”, Nat Methods, 5:235–237, doi: 10.1038/nmeth.1184.

Holloway, J.M., Dahlgren, R.A., Hansen, B., Casey, W.H., 1998, “Contribution of Bedrock Nitrogen to High Nitrate Concentrations in Stream Water”, Nature, 395, 785-788.

Holloway, J.M., 2002, “Nitrogen in Rock: Occurrences and Biogeochemical Implications”, Global Geochemical Cycles, 16, 1118, doi:1029/2002GB001862.

Horner-Devine, M.C., Carney, K.M., Bohannan, B.J., 2004, “An ecological perspective on bacterial biodiversity”, Proc Biol Sci., 271, 113-22.

Horner-Devine, M.C., Bohannan, B.J.M., 2006, “Phylogenetic clustering and overdispersion in bacterial communities”, Ecology, 87, S100–S108, doi: 10.1890/ 0012-9658(2006)87[100:PCAOIB]2.0.CO;2.

Hubalek, V., Wu, X., Eiler, A., Buck, M., Heim, C., Dopson, M., Bertilsson, S., Lonescu, D., 2016, “Connectivity to the surface determines diversity patterns in subsurface aquifers of the Fennoscandian shield”, ISME Journal, 10 (10), pp. 2447-2458.

Hubbell, S.P., 2001, “The unified neutral theory of biodiversity and biogeography,” in Monographs in Population Biology, eds S.A. Levin and H.S. Horn (Princeton, NJ: Princeton University Press), 32.

Itävaara, M., Nyyssönen, M., Kapanen, A., Nousiainen, A., Ahonen, L., Kukkonen, I., 2011, “Characterization of bacterial diversity to a depth of 1500 m in the Outokumpu deep borehole, Fennoscandian Shield”, FEMS Microbiology Ecology, 77 (2), pp. 295-309.

Itävaara, M., Salavirta, H., Marjamaa, K., Ruskeeniemi, T., 2016, “Geomicrobiology and Metagenomics of Terrestrial Deep Subsurface Microbiomes”, Advances in Applied Microbiology, 94, pp. 1-77.

Jain, D.K., Providenti, M., Tanner, C., Cord, I., Stroes-Gascoyne, S., 1997, “Characterization of microbial communities in deep groundwater from granitic rock”, Can. J. Microbiol., 43, 272–283, doi: 10.1139/m97-038.

Kallmeyer, J., Pockalny, R., Adhikari, R.R., Smith, D.C., D’Hondt, S., 2012, “Global distribution of microbial abundance and biomass in subseafloor sediment”, Proc. Natl. Acad. Sci. U.S.A., 109, 16213–16216, doi: 10.1073/pnas.1203849109.

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly et al., 2010, “Picante: R tools for integrating phylogenies and ecology”, Bioinformatics, 26, 1463–1464, doi: 10.1093/bioinformatics/btq166.

King-Sharp, K.J., Frape, S.K., Peterman, Z., Gwynne, R., Tian, L., Gurban, I., 2016, “Synthesis of Geochemical and Fracture Mineral Studies Relevant to a Deep Geological Repository for Nonfuel Wastes at Chalk River”, Canadian Nuclear Review, 10.12943/CNR.2016.00015.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M et al., 2013, “Evaluation of General 16S Ribosomal RNA Gene PCR Primers for Classical and Next-Generation Sequencing-Based Diversity Studies”, Nucleic Acids Res., 41:e1, doi: 10.1093/nar/gks808.

Koeppel, A.F., Wertheim, J.O., Barone, L., Gentile, N., Krizanc, D., Cohan, F.M., 2013, “Speedy Speciation in a Bacterial Microcosm: New Species can Arise as Frequently as Adaptations within a Species”, ISME Journal, 7, 1080-1091, doi: 10.1038/ismej.2013.3.

Konno, U., Kouduka, M., Komatsu, D.D., Ishii, K., Fukuda, A., Tsunogai, U et al., 2013, “Novel Microbial Populations in Deep Granitic Groundwater from Grimsel Test Site, Switzerland”, Microbial Ecology, 65 (3), pp. 626-637.

Küsel, K., Totsche, K.U., Trumbore, S.E., Lehmann, R., Steinhäuser, C., Herrmann, M., 2016, “How Deep Can Surface Signals be Traced in the Critical Zone? Merging Biodiversity with Biogeochemistry Research in a Central German Muschelkalk Landscape”, Frontiers in Earth Science, 4, art. no. 32.

Lear, G., Washington, V., Neale, M., Case, B., Buckley, H., Lewis, G., 2013, “The Biogeography of Stream Bacteria”, Global Ecology and Biogeography, 22: 544–554, doi:10.1111/geb.12046.

Lear, G., Bellamy, J., Case, B.S., Lee, J.E., Buckley, H.B., 2014, “Finescale Spatial Patterns in Bacterial Community Composition and Function within Freshwater Ponds”, ISME Journal, 8, 1715–1726, doi: 10.1038/ismej.2014.21.

Love, M.I., Huber, W., Anders, S., 2014, “Moderated Estimation of Fold Change and Dispersion for RNA-Seq data with DESeq2.”, Genome Biology, 15, pp. 550, doi: 10.1186/s13059-014-0550-8.

Lozupone, C., Knight, R., 2005, “UniFrac: A New Phylogenetic Method for Comparing Microbial Communities”, Applied and Environmental Microbiology, 71: 8228–8235.

Lozupone, C., Lladser, M.E., Knights, D., Stombaugh, J., Knight, R., 2011, “UniFrac: an Effective Distance Metric for Microbial Community Comparison”, ISME Journal, 5: 169–172.

Martiny, J.B.H., Eisen, J.A., Penn, K., Allison, S.D., Horner-Devine, M.C., 2011, "Drivers of Bacterial β-diversity Depend on Spatial Scale", Proceedings of the National Academy of Sciences, 108, 7850-7854, doi:10.1073/pnas.1016308108.

Mayfield, M.M., Levine, J.M., 2010, "Opposing Effects of Competitive Exclusion on the Phylogenetic Structure of Communities", Ecol. Lett., 13, 1085–1093, doi: 10.1111/j.1461-0248.2010.01509.x.

McMahon, S., Parnell, J., 2014, "Weighing the deep continental biosphere", FEMS Microbiology Ecology, 87 (1), pp. 113-120.

McMurdie, P.J., Holmes, S., 2013, "Phyloseq: An R Package for reproducible interactive analysis and graphics of microbiome census data", PLoS ONE, 8(4), pp. e61217, http://dx.plos.org/10.1371/journal.pone.0061217.

Muyzer, G., de Waal, E.C., Uitterlinden, A.G., 1993, "Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA", Appl. Environ. Microbiol., 59, 695–700.

Nemergut, D.R., Schmidt, S.K., Fukami, T., O'Neill, S.P., Bilinski, T.M., Stanish, L.F et al., 2013, "Patterns and processes of microbial community assembly", Microbiol. Mol. Biol. Rev., 77, 342–356, doi: 10.1128/MMBR.00051-12.

Neymark, L.A., Peterman, Z.E., Moscati, R.J., Thivierge, R.H., 2013, "U–Pb, Rb–Sr, and U-Series Isotope Geochemistry of Rocks and Fracture Minerals from the Chalk River Laboratories Site, Grenville Province, Ontario, Canada", Applied Geochemistry, 36, 10-33.

Østman, B., Lin, R., Adami, C., 2014, "Trade-Offs Drive Resource Specialization and the Gradual Establishment of Ecotypes", BMC Evolutionary Biology, 14:113, doi: 10.1186/1471-2148-14-113.

Pedersen, K., 2013, "Metabolic activity of subterranean microbial communities in deep granitic groundwater supplemented with methane and H 2", ISME Journal, 7 (4), pp. 839-849.

Pei, A.Y., Oberdorf, W.E., Nossa, C.W, Agarwal, A., Chokshi, P., Gerz, E.A. et al., 2010, "Diversity of 16S rRNA Genes within Individual Prokaryotic Genomes", Applied and Environmental Microbiology, 76, 3886-3897, doi: 10.1128/AEM.02953-09.

Peterman, Z., Neymark, L., King-Sharp, K.J., Gascoyne, M., 2016, "Isotope hydrology of the Chalk River Laboratories Site, Ontario, Canada", Applied Geochemistry, 66, 149-161.

Price, M.N., Dehal, P.S., Arkin, A.P., 2010, "FastTree 2 – approximately maximum-likelihood trees for large alignments", PLOS One, 5:e9490, doi: 10.1371/journal.pone.0009490.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J et al., 2007, "SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB", Nucleic Acids Res., 35:7188–7196, doi: 10.1093/nar/gkm864.

Purkamo, L., Bomberg, M., Nyyssönen, M., Kukkonen, I., Ahonen, L., Itävaara, M., 2014, "Heterotrophic Communities Supplied by Ancient Organic Carbon Predominate in Deep Fennoscandian Bedrock Fluids", Microbial Ecology, 69 (2), pp. 319-332.

Purkamo, L., Bomberg, M., Kietäväinen, R., Salavirta, H., Nyyssönen, M., Nuppunen-Puputti, M., et al., 2016, "Microbial co-occurrence patterns in deep Precambrian bedrock fracture fluids", Biogeosciences, 13 (10), pp. 3091-3108.

R Core Team, 2015, "R: A Language and Environment for Statistical Computing", Vienna: R Foundation for Statistical Computing, available at: http://www.Rproject.org.

Suna, D.L., Jianga, X., Wub, Q.L., Zhoua, N-Y., 2013, "Intragenomic Heterogeneity of 16S rRNA Genes Causes Overestimation of Prokaryotic Diversity", Applied and Environmental Microbiology, 79, 5962-5969.

Treves, D.S., Manning, S., Adams, J., 1998, "Repeated Evolution of an Acetate Cross-feeding Polymorphism in Long-Term Populations of Eschericia coli", Molecular Biology and Evolution, 15, 789-797.

Vellend, M., Cornwell, W.K., Magnuson-Ford, K., Mooers, A., 2010, "Measuring phylogenetic biodiversity," in Biological Diversity: Frontiers in Measurement and Assessment, Chap. 14, eds A.E. Magurran and B.J. McGill (Oxford: Oxford University Press).

Vellend, M., 2016, "The Theory of Ecological Communities", Monograph in Population Biology, 57, Princeton University Press.

Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007, "Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy", Appl Environ Microbiol, 73:5261–5267, doi: 10.1128/AEM.00062-07.

Wu, X., Holmfeldt, K., Hubalek, V., Lundin, D., Åström, M., Bertilsson, S et al., 2016, "Microbial metagenomes from three aquifers in the Fennoscandian shield terrestrial deep biosphere reveal metabolic partitioning among populations", ISME Journal, 10 (5), pp. 1192-1203.

Zhang, J., Kobert, K., Flouri, T., Stamatakis, A., 2014, "PEAR: A fast and accurate Illumina Paired-End reAd merger", Bioinformatics, 30:1–7, doi: 10.1093/bioinformatics/btt593.

4.11 List of Figures

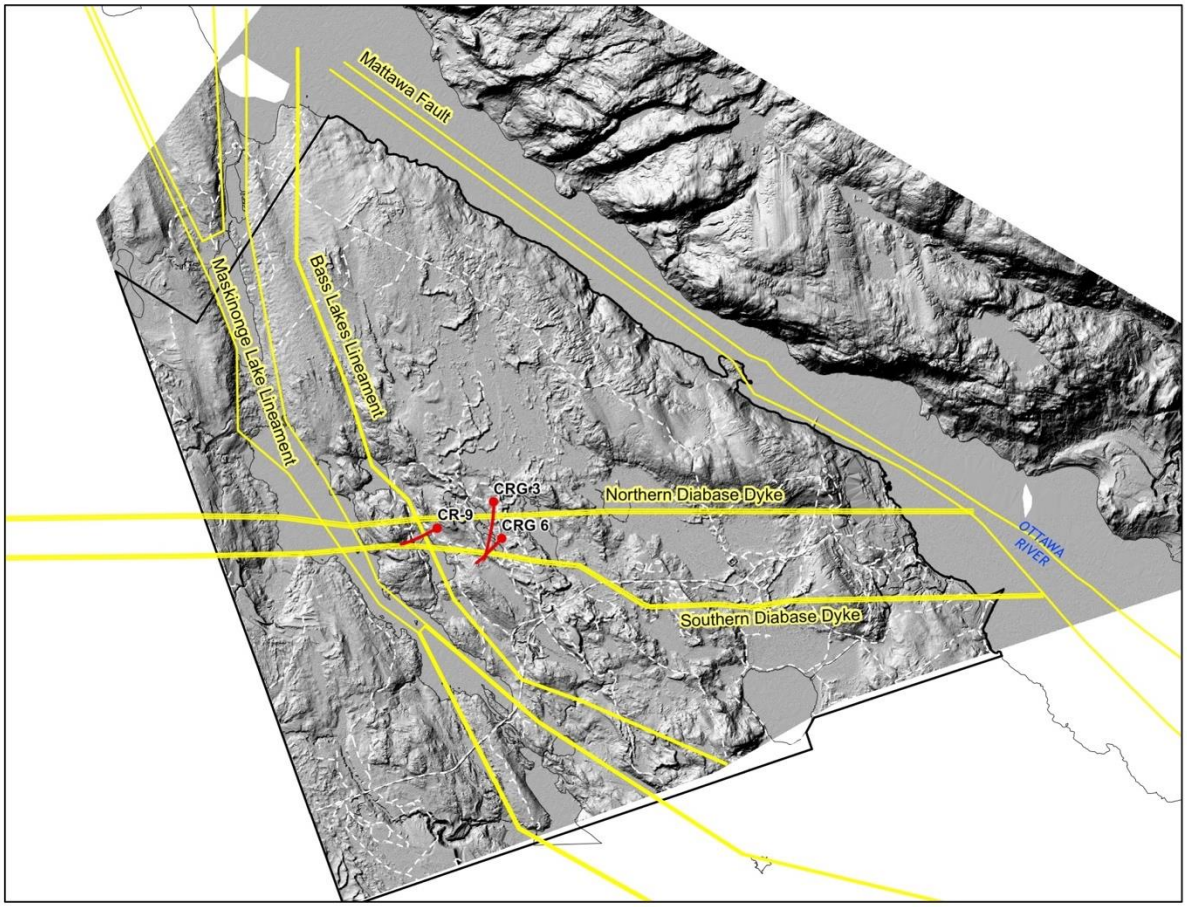


Figure 4-1 Surface topography of the Chalk River site and the approximate locations of boreholes CR-9, CRG-3 and CRG-6 relative to the site geologic boundaries: the Mattawa fault, the Maskinonge/Bass Lake faults and the East-West trending diabase dykes.

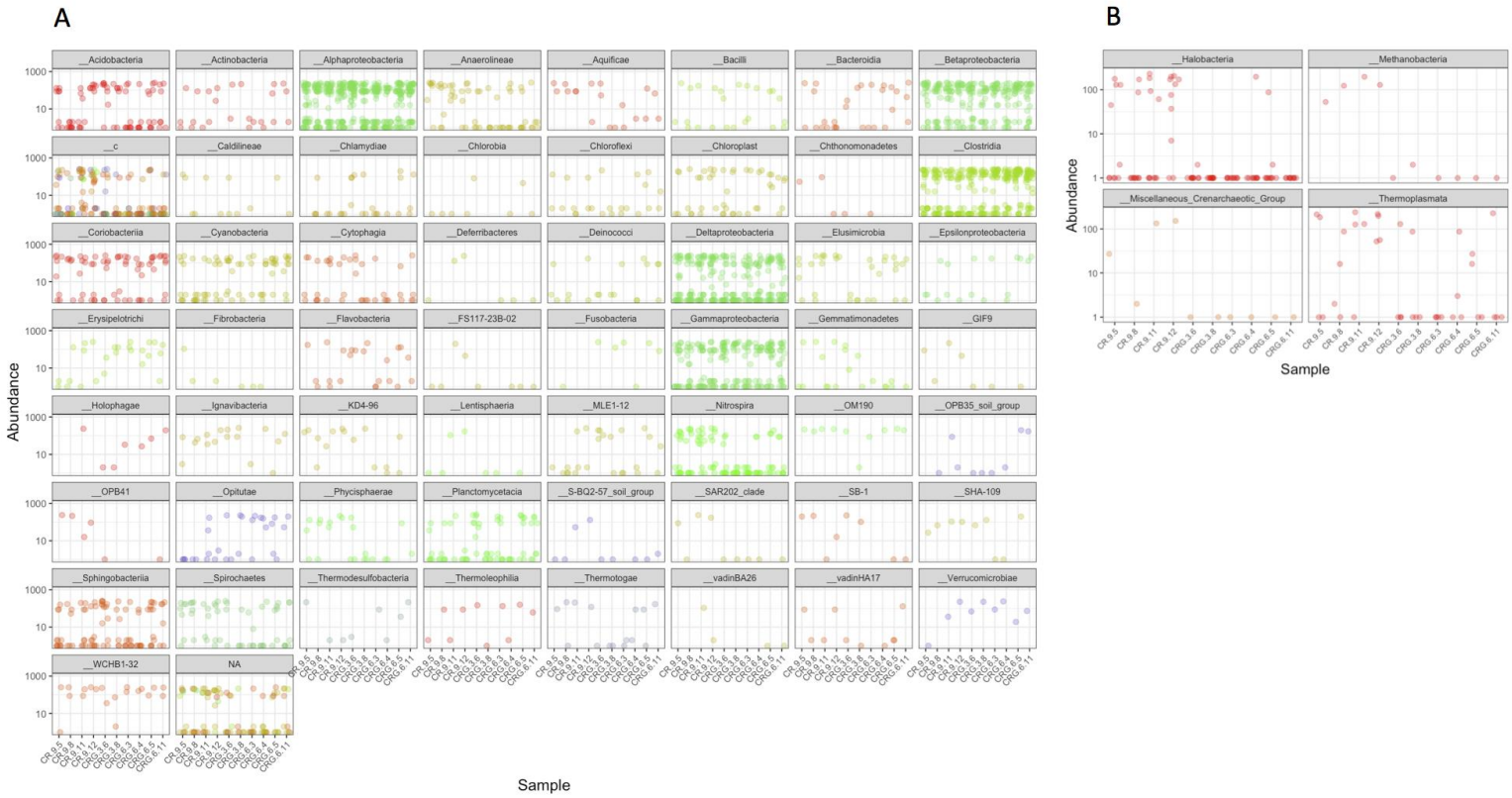


Figure 4-2 Distribution of OTUs for bacteria (A) and Archaea (B). Abundances are by Phylum and Class within the fracture water sampling locations. The abundance data was rarified to a common sequence depth of 2941.

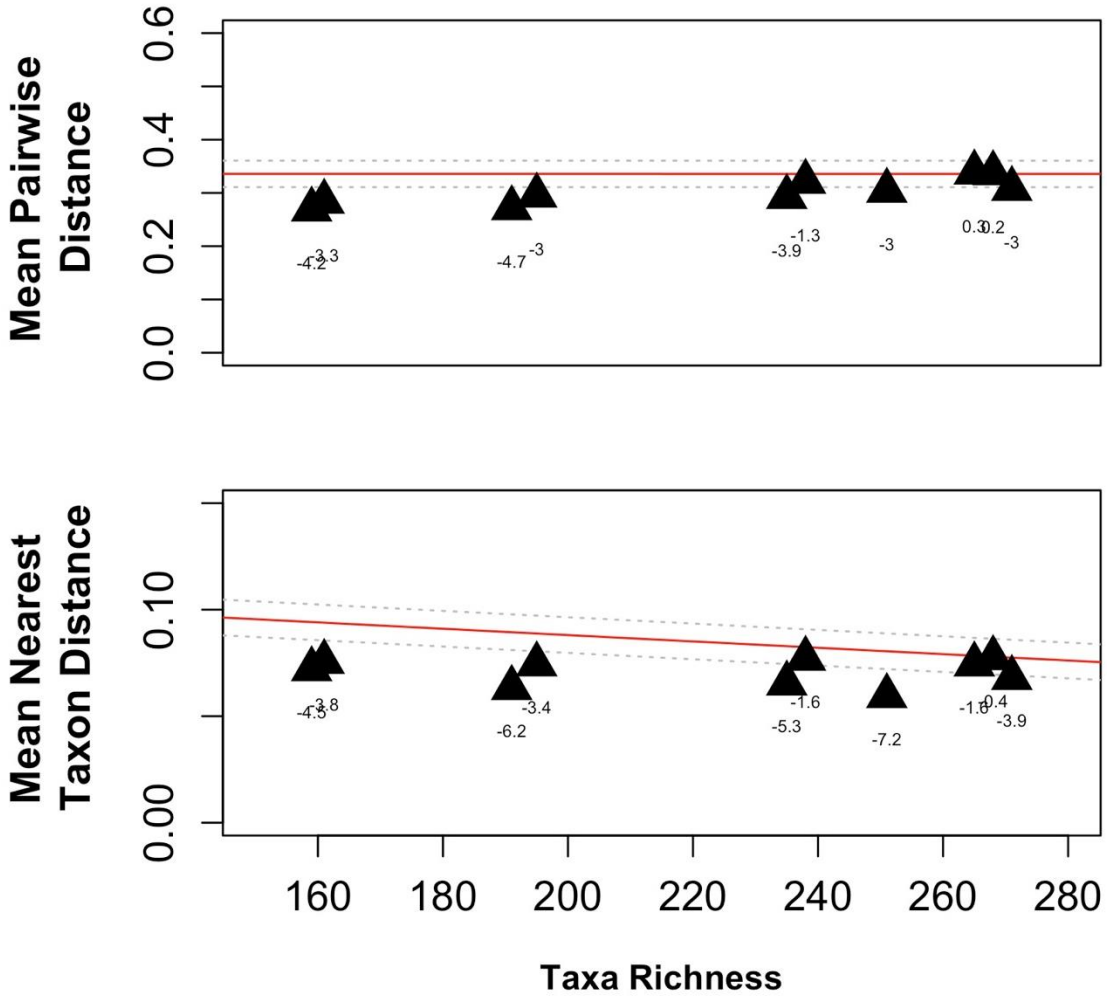


Figure 4-3 Co-occurrence of related taxa. Observed values (black triangles) for mean pairwise distance (MPD) and the mean nearest taxon distance (MNTD) of assemblages relative to the taxa richness. The solid red lines are the regression lines for the null mean ± two standard deviations (dotted lines). Triangles represents the values for each sampling location (assemblage); a filled triangle inside of an open triangle represents values that were within two standard deviations from the null mean; all other data points were more than 2 standard deviations from the null mean. Values below the red line indicate phylogenetic clustering. Abundance data were variance stabilized by the regularized log transform.

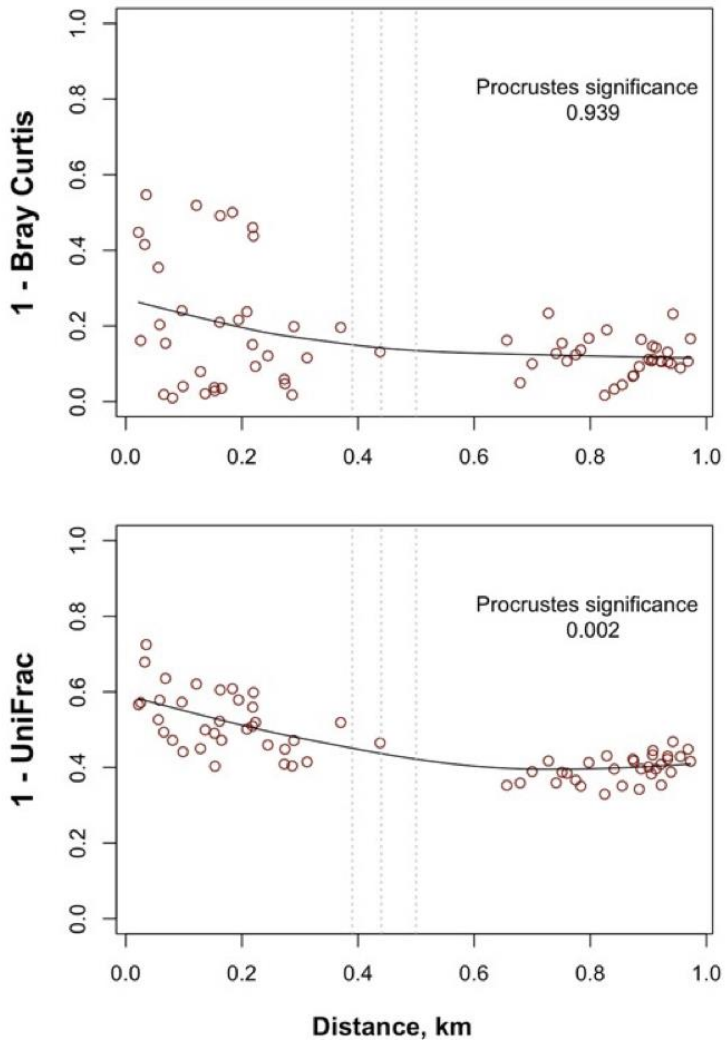


Figure 4-4 Decay of similarity for Bray Curtis, 1-UniFrac dissimilarity spanning a 1 km geographic distance between sampling locations. The grey lines show the maximum distance between the shallowest and deepest sampling locations within each borehole: CR-9, ~500 m; CRG-3, ~390 m; CRG-6, ~440 m

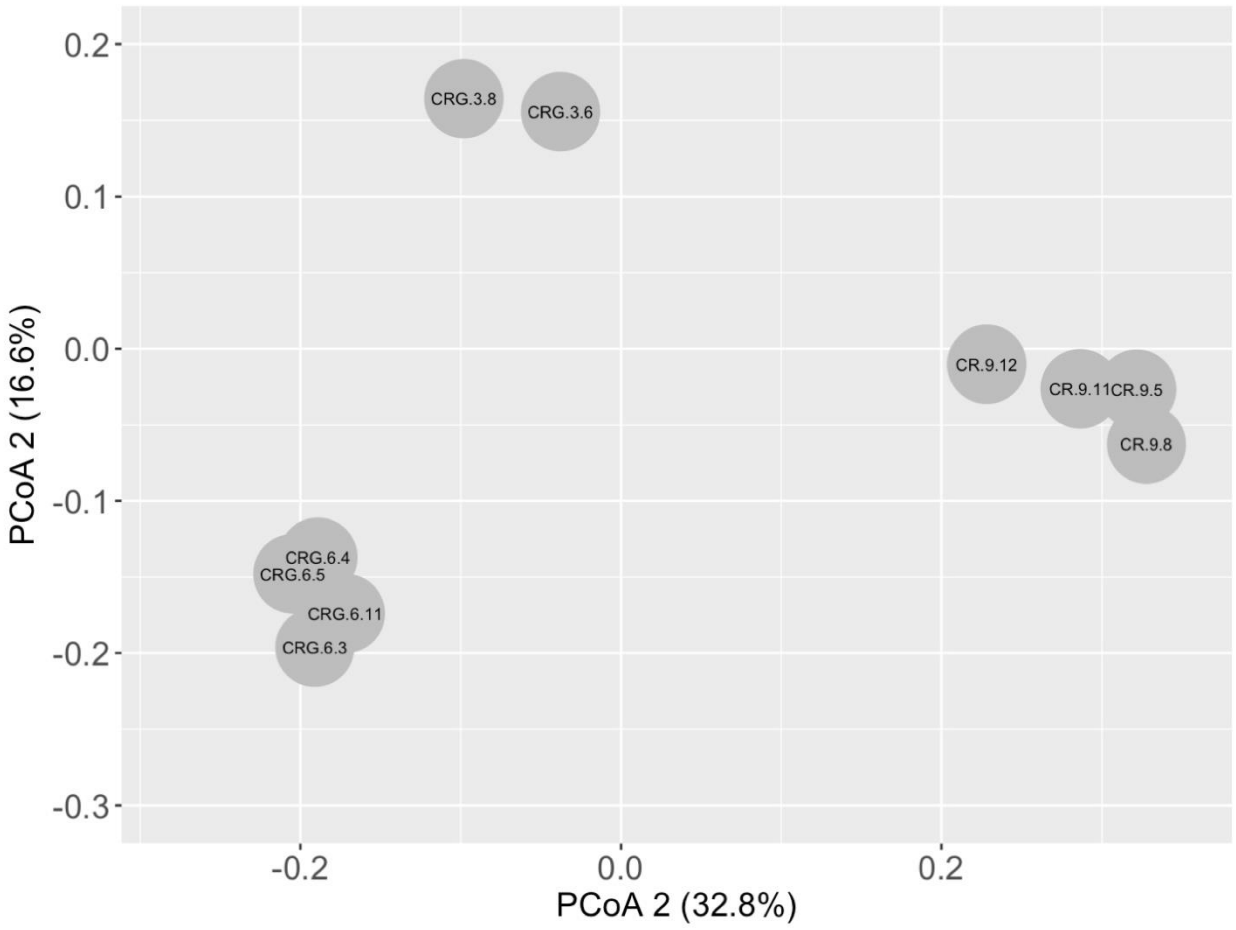


Figure 4-5 Principle coordinate analysis of weighted UniFrac distances. Labels for sampling locations are by borehole and interval (see Table S1).

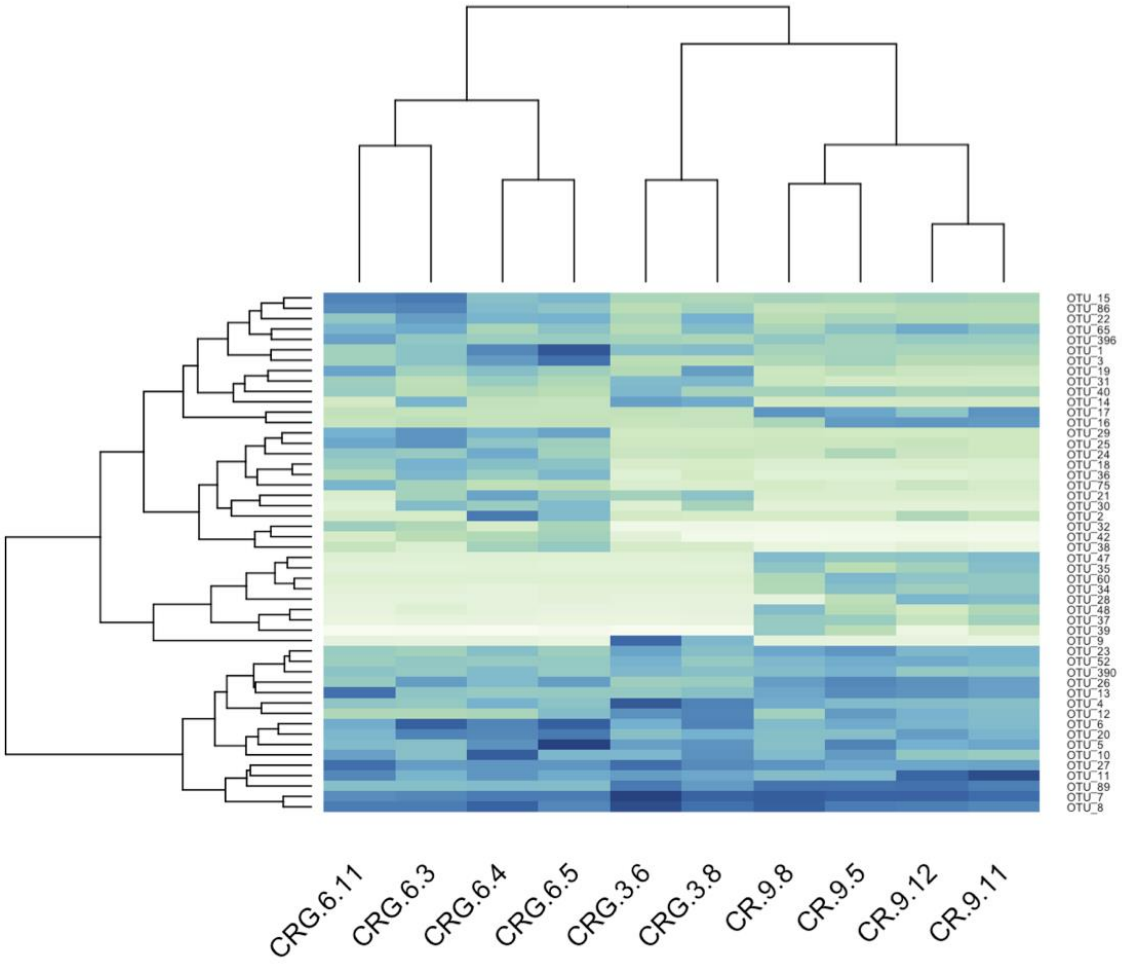


Figure 4-6 Heatmap of the 50 most abundant 16s rRNA genes. Counts were variance stabilized using the regularized log transformation.

4.12 Supplemental Information

TABLE S1. Groundwater physicochemical and borehole descriptive parameters

Borehole	Interval	Sampling Date	Elevation*, m collar	Elevation*, m interval	Easting**	Northing m
CRG-3	6	23/9/2013	163	-122	313344.3	5102404
	8	25/9/2013		-247	313335.7	5102336
	11	26/9/2013		-417	313318.3	5102241
	14	27/9/2013		-617	313293.2	5102122
CRG-6	3	6/5/2013	171	48	313391.2	5102135
	4	12/5/2013		-4	313373.8	5102116
	5	14/5/2013		-68	313351.6	5102092
	11	13/5/2013		-390	313234.8	5101980
CRG-9	5	5/5/2014	124.8	-154	312617.4	5102207
	8	6/5/2014		-364	312466.6	5102145
	11	12/5/2014		-485	312449.8	5102130
	12	13/5/2014		-542	312416.1	5102120

* elevation relative to sea level

** Zone 18 of the Universal Transverse Mercator coordinate system

TABLE S2 Taxa richness and Faith's phylogenetic diversity (PD) of assemblages, including assemblages from within sampling locations CRG-3 intervals 11 and 14.

Borehole	Interval	Number of Taxa	Phylogenetic Diversity, PD			
			Observed	Null ± sd	z-value	p-value
CRG-3	6	240	14.8	17.7 ± 0.33	-8.8	0.001
	8	158	10.3	12.7 ± 0.34	-7.3	0.001
	11	71	5.5	6.6 ± 0.27	-4.2	0.001
	14	41	3.3	4.2 ± 0.23	-3.8	0.001
CRG-6	3	158	10.8	12.7 ± 0.32	-6.0	0.001
	4	260	17.2	18.9 ± 0.32	-5.2	0.001
	5	225	14.6	16.9 ± 0.33	-7.0	0.001
	11	189	12.0	14.7 ± 0.32	-8.3	0.001
CRG-9	5	259	18.3	18.8 ± 0.33	-1.6	0.057
	8	260	18.5	18.9 ± 0.33	-1.1	0.143
	11	236	16.4	17.5 ± 0.33	-3.2	0.003
	12	190	13.1	14.7 ± 0.35	-4.7	0.001

Repetitions =999

TABLE S3 Partitioning phylogenetic diversity across the meta-community

PD total	PD alpha assemblage average	PD beta PD missing from the assemblage average
100%	48%	52%

TABLE S4. Procrustes significance values for four measures of phylogenetic beta-diversity and Bray-Curtis; testing for environmental and spatial relationships

	Significance of Beta-Diversity Metrics				
	UniFrac unweighted	**UniFrac weighted**	**Dnn**	**Dpw**	**Bray-Curtis**
Environmental Distance	0.099	0.128	0.232	0.187	0.763
Spatial Distance	0.003	0.002	0.007	0.161	0.932
	terminal	terminal	terminal	basal	

Repetitions =999

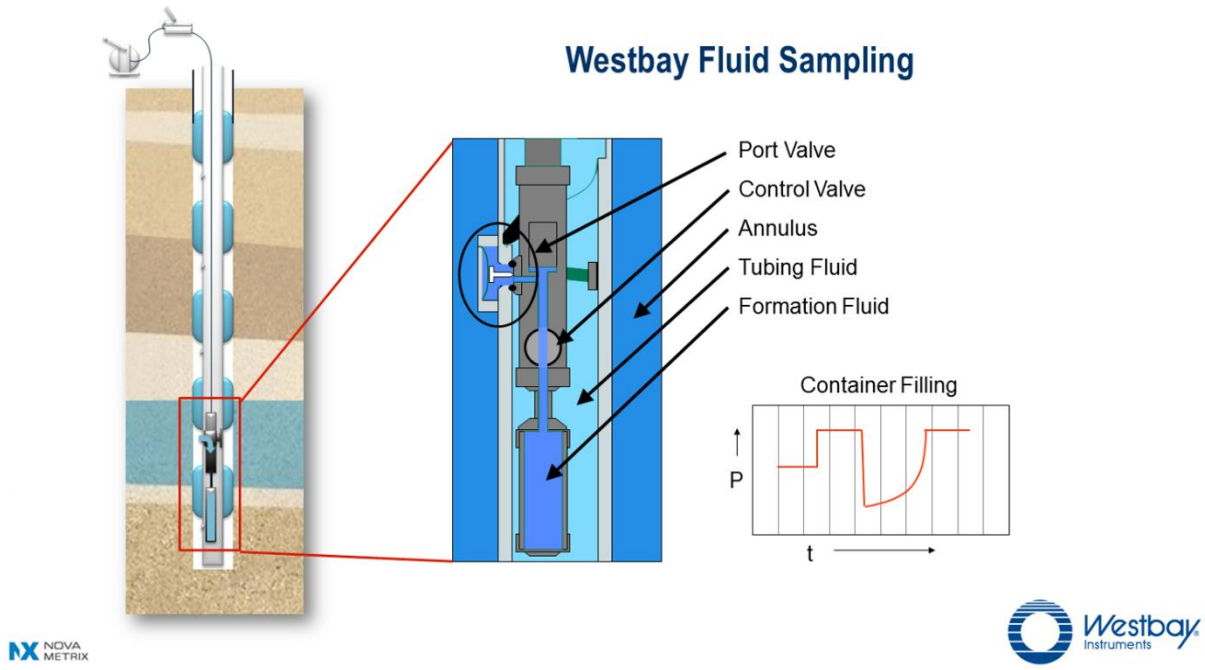


Figure S1. Schematic of the Westbay System

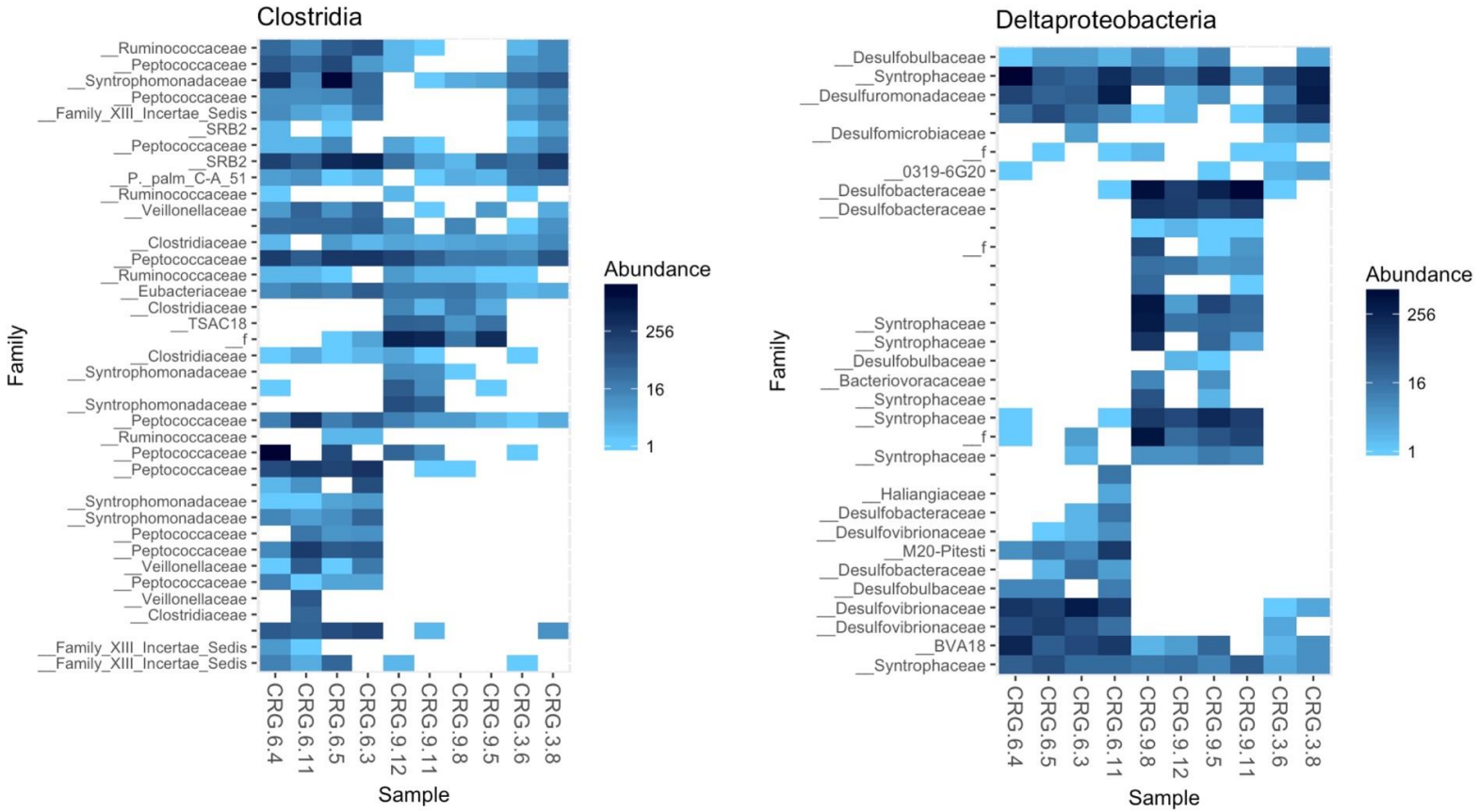


Figure S2. Heatmap showing the relative abundances of OTUs identified as *Clostridia* and *Deltaproteobacteria* and with known relative associated with sulfur metabolism.

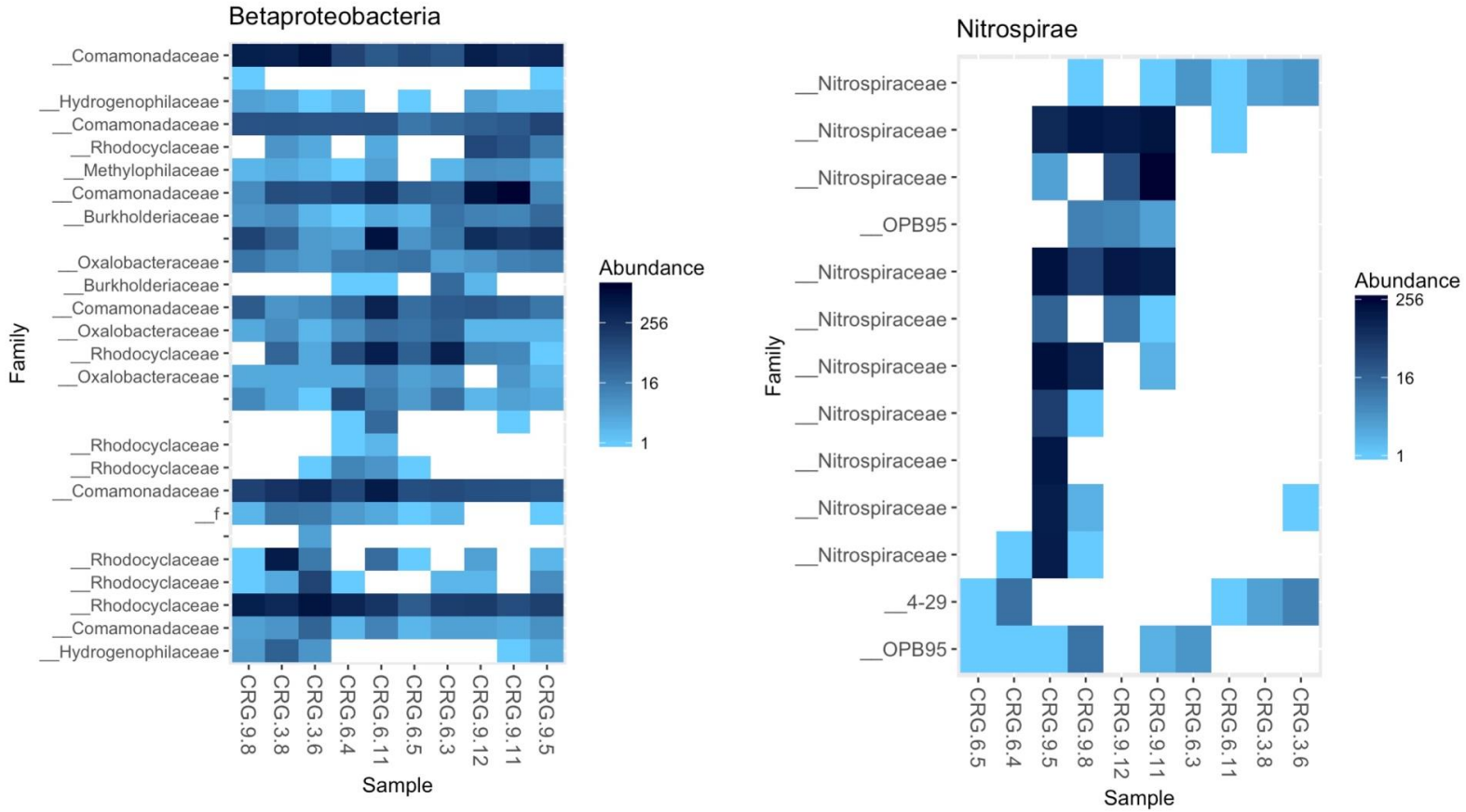


Figure S3. Heatmap showing the relative abundances of OTUs identified as *Betaproteobacteria* and *Nitrospirae* and with known relative associated with nitrogen metabolism.

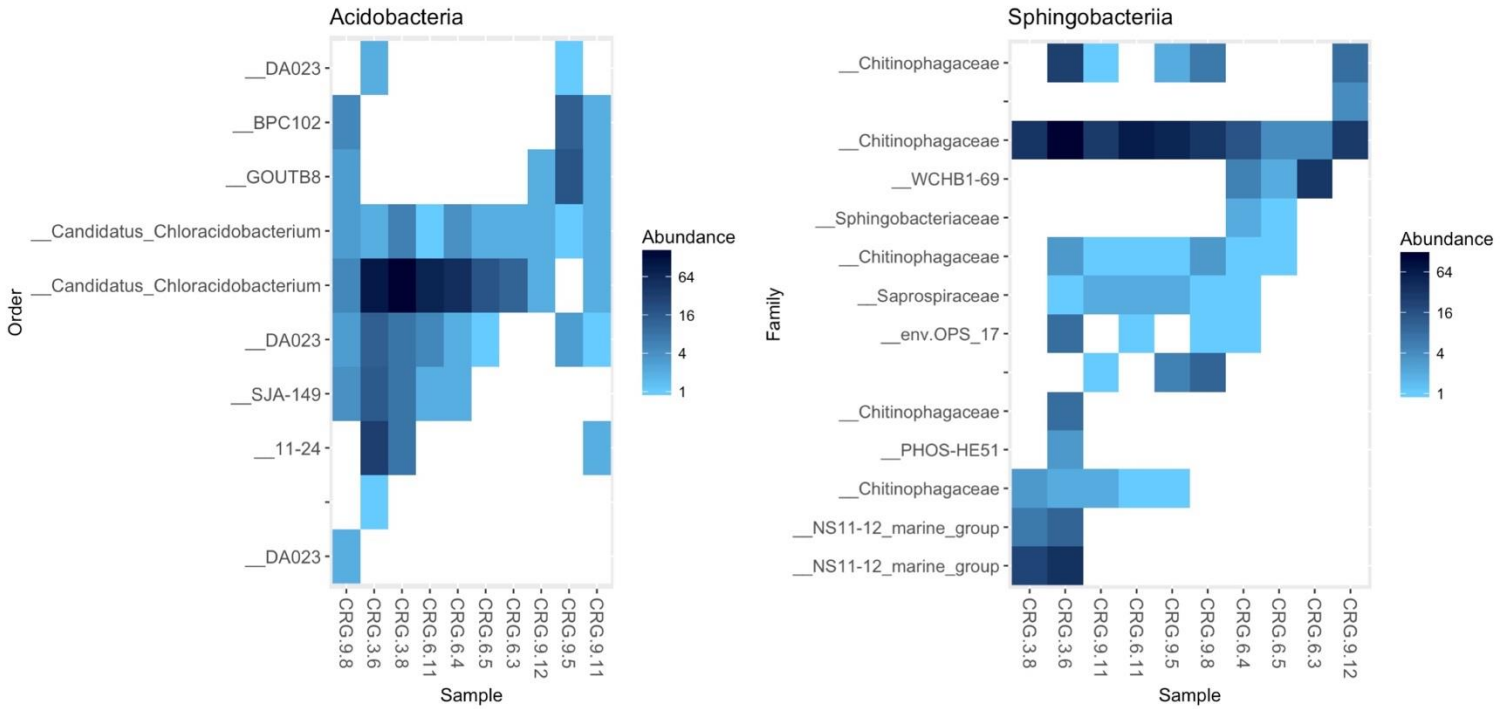


Figure S4 Diagram showing the relative abundances of operational taxonomical units (OTUs) within each sample (heatmap) representing Acidobacteria and Sphingobacteriia.

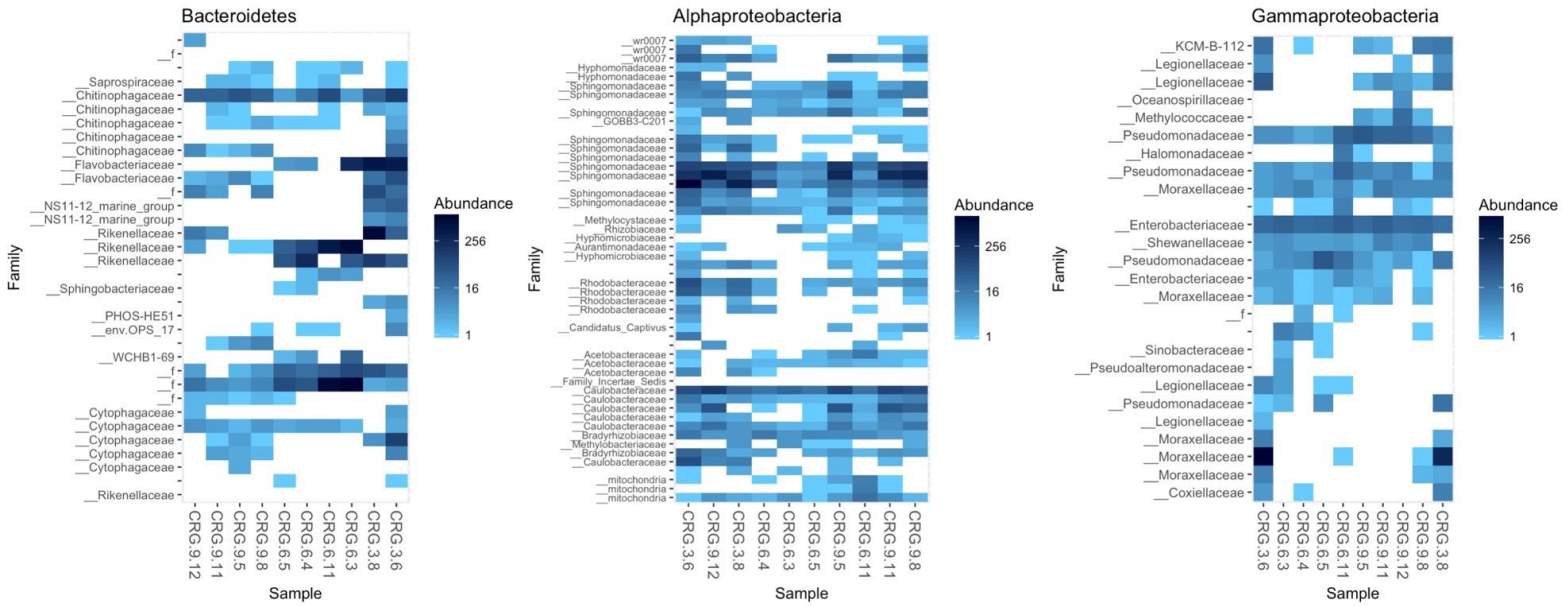


Figure S5 Diagram showing the relative abundances of operational taxonomical units (OTUs) within each sample (heatmap) representing *Bacteroidetes*, *Alphaproteobacteria* and *Gammaproteobacteria*.

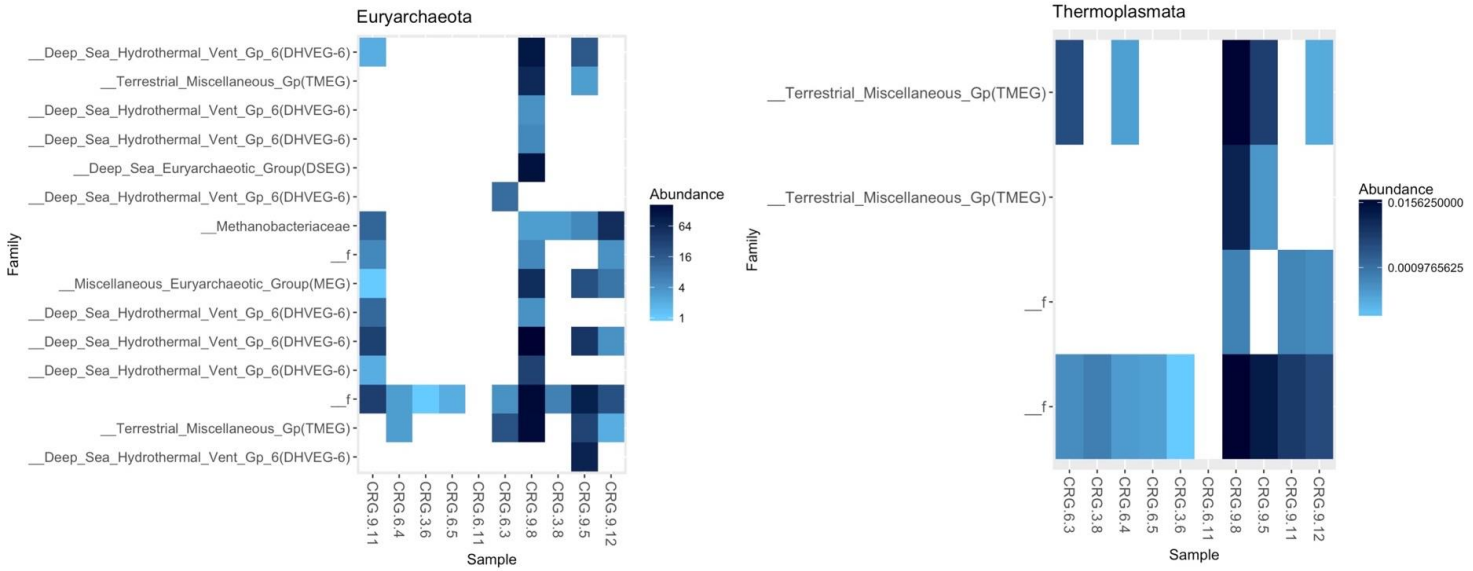


Figure S6. Heatmap showing the relative abundances of (OTUs) identified as *Euyarchaeota* and *Themoplasmata*.

TABLE S5. Identities of the sister pairs obtained from the phylogenetic tree for the meta-community. The OTUs that displayed differential abundance (Chapter 4) are marked by the name of the biome.

OTU sister A	OTU sister B	Kingdom	Phylum	Class	Displayed differential abundance
OTU_77	OTU_178	Bacteria	Proteobacteria	Betaproteobacteria	CRG6
OTU_3	OTU_324	Bacteria	Elusimicrobia	Elusimicrobia	
OTU_244	OTU_310	Bacteria	Bacteroidetes	Sphingobacteriia	
OTU_390	OTU_396	Bacteria	Proteobacteria	Betaproteobacteria	CRG6
OTU_204	OTU_419	Bacteria	Proteobacteria	Alphaproteobacteria	
OTU_240	OTU_65	Bacteria	Chloroflexi	Anaerolineae	
OTU_24	OTU_135	Bacteria	Proteobacteria	Deltaproteobacteria	CRG6
OTU_354	OTU_173	Bacteria	Cyanobacteria	Cyanobacteria	
OTU_423	OTU_375	Bacteria	Proteobacteria	Betaproteobacteria	
OTU_115	OTU_384	Bacteria	Cyanobacteria	Chloroplast	
OTU_15	OTU_156	Bacteria	Bacteroidetes	WCHB1-32	
OTU_215	OTU_232	Bacteria	Proteobacteria	Alphaproteobacteria	
OTU_285	OTU_332	Bacteria	Chloroflexi	Anaerolineae	
OTU_205	OTU_394	Bacteria	Proteobacteria	Alphaproteobacteria	
OTU_249	OTU_52	Bacteria	Proteobacteria	Alphaproteobacteria	
OTU_312	OTU_129	Bacteria	Cyanobacteria	MLE1-12	
OTU_23	OTU_259	Bacteria	Proteobacteria	Alphaproteobacteria	
OTU_262	OTU_336	Bacteria	Proteobacteria	Betaproteobacteria	
OTU_81	OTU_157	Bacteria	Cyanobacteria	Cyanobacteria	
OTU_223	OTU_55	Bacteria	Proteobacteria	Deltaproteobacteria	
OTU_227	OTU_416	Bacteria	NA	NA	
OTU_372	OTU_386	Bacteria	Cyanobacteria	Cyanobacteria	
OTU_277	OTU_279	Bacteria	Chlamydiae	Chlamydiae	
OTU_357	OTU_95	Bacteria	Chloroflexi	Anaerolineae	CR9
OTU_216	OTU_318	Bacteria	Chloroflexi	NA	
OTU_195	OTU_317	Bacteria	Chloroflexi	KD4-96	
OTU_319	OTU_64	Bacteria	Proteobacteria	Alphaproteobacteria	
OTU_217	OTU_121	Bacteria	Bacteroidetes	Sphingobacteriia	
OTU_20	OTU_75	Bacteria	Firmicutes	Clostridia	
OTU_337	OTU_131	Bacteria	Firmicutes	Clostridia	
OTU_305	OTU_148	Bacteria	Proteobacteria	Deltaproteobacteria	
OTU_25	OTU_44	Bacteria	Bacteroidetes	Bacteroidia	
OTU_288	OTU_63	Bacteria	Proteobacteria	Gammaproteobacteria	
OTU_292	OTU_363	Bacteria	Proteobacteria	Alphaproteobacteria	
OTU_193	OTU_143	Bacteria	Firmicutes	Clostridia	
OTU_42	OTU_18	Bacteria	Proteobacteria	Deltaproteobacteria	
OTU_308	OTU_162	Bacteria	Proteobacteria	Alphaproteobacteria	
OTU_283	OTU_261	Bacteria	Proteobacteria	Deltaproteobacteria	
OTU_380	OTU_160	Bacteria	Acidobacteria	Acidobacteria	
OTU_180	OTU_168	Archaea	Euryarchaeota	Halobacteria	
OTU_190	OTU_391	Bacteria	Bacteroidetes	Sphingobacteriia	
OTU_72	OTU_145	Bacteria	Actinobacteria	Coriobacteriia	
OTU_103	OTU_164	Bacteria	NA	NA	
OTU_229	OTU_427	Bacteria	Planctomycetes	Planctomycetacia	
OTU_181	OTU_351	Bacteria	Proteobacteria	Betaproteobacteria	
OTU_251	OTU_266	Bacteria	Proteobacteria	Alphaproteobacteria	
OTU_270	OTU_411	Bacteria	Proteobacteria	Gammaproteobacteria	
OTU_201	OTU_202	Bacteria	Planctomycetes	Planctomycetacia	

OTU sister A	OTU sister B	Kingdom	Phylum	Class	Displayed differential abundance
OTU_28	OTU_120	Bacteria	Actinobacteria	Coriobacteriia	CR9
OTU_338	OTU_238	Bacteria	Nitrospirae	Nitrospira	
OTU_231	OTU_281	Bacteria	Proteobacteria	Epsilonproteobacteria	
OTU_298	OTU_328	Bacteria	Chloroflexi	GIF9	
OTU_352	OTU_99	Bacteria	Proteobacteria	Gammaproteobacteria	
OTU_53	OTU_186	Bacteria	Gemmatimonadetes	Gemmatimonadetes	
OTU_8	OTU_429	Bacteria	Proteobacteria	Betaproteobacteria	
OTU_287	OTU_401	Bacteria	Acidobacteria	Acidobacteria	
OTU_220	OTU_152	Bacteria	Actinobacteria	Actinobacteria	
OTU_368	OTU_151	Bacteria	Cyanobacteria	Chloroplast	
OTU_260	OTU_342	Bacteria	Bacteroidetes	Cytophagia	
OTU_379	OTU_85	Bacteria	Proteobacteria	Betaproteobacteria	
OTU_276	OTU_127	Bacteria	Proteobacteria	Alphaproteobacteria	
OTU_408	OTU_194	Bacteria	Proteobacteria	Alphaproteobacteria	
OTU_236	OTU_69	Bacteria	Firmicutes	Erysipelotrichi	
OTU_101	OTU_93	Bacteria	Proteobacteria	Alphaproteobacteria	
OTU_112	OTU_14	Bacteria	Bacteroidetes	Flavobacteria	CRG6
OTU_329	OTU_340	Bacteria	Bacteroidetes	Sphingobacteriia	
OTU_253	OTU_45	Bacteria	Firmicutes	Clostridia	CRG6
OTU_356	OTU_406	Bacteria	Chloroflexi	Caldilineae	
OTU_33	OTU_381	Bacteria	Proteobacteria	Gammaproteobacteria	
OTU_2	OTU_214	Bacteria	Firmicutes	Clostridia	
OTU_359	OTU_82	Bacteria	Proteobacteria	Gammaproteobacteria	
OTU_415	OTU_76	Bacteria	WCHB1-60	c	
OTU_361	OTU_79	Bacteria	Proteobacteria	Deltaproteobacteria	
OTU_200	OTU_289	Bacteria	NA	NA	
OTU_255	OTU_369	Bacteria	GOUTA4	c	
OTU_252	OTU_31	Bacteria	Acidobacteria	Acidobacteria	
OTU_280	OTU_140	Bacteria	Proteobacteria	Gammaproteobacteria	
OTU_104	OTU_163	Archaea	Euryarchaeota	Thermoplasmata	CR9
OTU_1	OTU_315	Bacteria	Firmicutes	Clostridia	
OTU_365	OTU_9	Bacteria	Proteobacteria	Gammaproteobacteria	
OTU_293	OTU_425	Bacteria	Planctomycetes	Phycisphaerae	
OTU_239	OTU_341	Bacteria	TM6	c	
OTU_245	OTU_334	Bacteria	Planctomycetes	Planctomycetacia	
OTU_254	OTU_177	Unclassified	NA	NA	
OTU_393	OTU_212	Bacteria	Bacteroidetes	vadinHA17	
OTU_256	OTU_269	Bacteria	Proteobacteria	Deltaproteobacteria	
OTU_395	OTU_409	Bacteria	Deinococcus-Thermus	Deinococci	
OTU_199	OTU_316	Bacteria	Armatimonadetes	c	
OTU_78	OTU_137	Bacteria	BD1-5	c	CR9
OTU_192	OTU_225	Bacteria	Planctomycetes	OM190	
OTU_208	OTU_300	Bacteria	Nitrospirae	Nitrospira	
OTU_26	OTU_59	Bacteria	Chlorobi	Ignavibacteria	CRG6
OTU_398	OTU_182	Bacteria	Proteobacteria	Deltaproteobacteria	
OTU_203	OTU_29	Bacteria	Firmicutes	Clostridia	
OTU_113	OTU_142	Archaea	Euryarchaeota	Halobacteria	
OTU_153	OTU_420	Bacteria	Proteobacteria	Deltaproteobacteria	
OTU_330	OTU_282	Bacteria	Verrucomicrobia	c	
OTU_263	OTU_73	Bacteria	Spirochaetes	Spirochaetes	
OTU_243	OTU_105	Bacteria	Aquificae	Aquificae	
OTU_353	OTU_399	Bacteria	Chloroflexi	Chloroflexi	
OTU_32	OTU_149	Bacteria	Firmicutes	Clostridia	

OTU sister A	OTU sister B	Kingdom	Phylum	Class	Displayed differential abundance
OTU_119	OTU_322	Bacteria	Firmicutes	Clostridia	
OTU_417	OTU_165	Bacteria	Candidate_division_OP3	c	
OTU_114	OTU_136	Bacteria	Elusimicrobia	Elusimicrobia	
OTU_422	OTU_432	Bacteria	Proteobacteria	Alphaproteobacteria	
OTU_290	OTU_385	Bacteria	Deinococcus-Thermus	Deinococci	
OTU_222	OTU_274	Bacteria	NA	NA	
OTU_198	OTU_350	Bacteria	Thermodesulfobacteria	Thermodesulfobacteria	
OTU_242	OTU_49	Bacteria	Candidate_division_WS3	c	
OTU_87	OTU_98	Bacteria	Nitrospirae	Nitrospira	CR9
OTU_219	OTU_273	Bacteria	TA06	c	
OTU_364	OTU_335	Bacteria	Candidate_division_OD1	c	
OTU_355	OTU_70	Bacteria	Candidate_division_SR1	c	
OTU_197	OTU_57	Bacteria	Spirochaetes	Spirochaetes	
OTU_117	OTU_172	Bacteria	Firmicutes	Clostridia	
OTU_228	OTU_275	Bacteria	NA	NA	
OTU_237	OTU_106	Bacteria	Proteobacteria	Gammaproteobacteria	
OTU_309	OTU_175	Bacteria	Firmicutes	Clostridia	
OTU_295	OTU_362	Bacteria	NA	NA	
OTU_430	OTU_139	Bacteria	Firmicutes	Bacilli	
OTU_320	OTU_37	Bacteria	Fibrobacteres	Fibrobacteria	
OTU_271	OTU_374	Bacteria	NA	NA	
OTU_61	OTU_158	Bacteria	Verrucomicrobia	Opitutae	

5. SELECTION INFLUENCES MICROBIAL CO-OCCURRENCE AND ABUNDANCE PATTERNS WITHIN SUBSURFACE CRYSTALLINE FRACTURES[6]

5.1 Abstract

Long-term safety predictions for geologic confinement of radioactive waste rely on understanding the natural microbial processes that occur within a host formation. In addition to selection, local random influences like dispersal and drift can contribute to observed patterns of microbial diversity and abundances; in such cases, predictions that are based solely on environmental conditions may not be relevant.

A study of a crystalline subsurface at Chalk River Laboratories, located in Chalk River Ontario allowed for the analysis of patterns in gene sequences under an ecological framework of selection, dispersal, drift and diversification. Within the extent relevant to a repository, phylogenetic clustering and decay of similarity with distance suggest that assembly may governed by trait-based biotic selection and dispersal. The redundancy in functional genes for energy metabolism across sampling locations suggest common traits enable co-existence. Only a few taxa displayed a growth advantage linked to sulfate, organic carbon, manganese and an unknown but spatially correlated parameter.

The suite of genes for nitrogen metabolism suggest a potential for complete nitrogen cycling including nitrogen fixation. Demonstrated metabolism by nitrate reduction was common across all sampling locations. Demonstrated metabolism by sulfate reduction was observed within a subset of sampling locations.

The ecological framework makes it possible to distinguish between selection and randomness in subsurface community assembly. Models for repository performance and safety will be improved by differentiating between the four main processes: selection, dispersal, drift and diversification.

5.2 Introduction

Underground repositories for nuclear waste management involve long-term geologic confinement of radioactive waste. Predictions of natural subsurface processes over long residence times within host formations, such as terrestrial crystalline formations, rely on knowledge of the component taxa, functional genes and demonstrated traits based on growth in selective media (Jain et al., 1997; Itävaara et al., 2011; Konno et al., 2013; Pedersen, 2013; Mcmahon and Parnell, 2014; Purkamo et al., 2014; Beaton et al., 2016; Hubalek et al., 2016;

[6] Adapted from
Danielle Beaton, Bradley S. Stevenson, Karen J. King-Sharp, Blake W. Stamps, Heather S. Nunn, Marilune Stuart, Sim Stroes-Gascoyne, and Jack Cornett, 2018, Trait-Based Biotic Selection Influences Microbial Co-Occurrence and Abundance Patterns within Subsurface Crystalline Fractures, submitted to ISME J January, 2018.

Itävaara et al., 2016; Küsel et al., 2016; Purkamo et al., 2016; Wu et al., 2016). Terrestrial crystalline formations support microbial communities composed of as few as a single taxon (Chivian et al., 2008) to hundreds of taxa (Itävaara et al., 2011; Hallbeck and Pedersen, 2012; Nyyssönen et al., 2014; Stegen et al., 2015; Beaton et al., 2016; Beaton et al., submitted). Evaluating the role of selection on these community patterns will provide evidence for whether a subsurface fractured environment represents a structured ecosystem and whether the corresponding ecosystem processes can be incorporated into long term deterministic models for repository safety.

Component individuals within an assemblage encompass a range of functional traits for cycling nutrients; and these traits contribute to ecosystem processes. Apparent biogeography in the phylogenetic and metabolic patterns of microbial communities (Horner-Devine et al., 2004; Horner-Devine and Bohoman, 2006; Martiny et al., 2011; Nemergut et al., 2011; Lear et al., 2013; Goberna et al., 2014; Lear et al., 2014; Beaton et al., 2016) challenge the emphasis placed on selection as being a main driving force in microbial community assembly and ecosystem function (Baas-Becking, 1934). This further challenges our ability to predict energy flow through an ecosystem based on community composition (Graham et al., 2016; Bier et al., 2015; Griebler and Lueders, 2009). A non-random phylogenetic pattern would suggest selection as a governing process on community assembly (Horner-Devine and Bohannan, 2006; Emerson and Gillespie, 2008); however, the influence of dispersal can obscure the influence of selection on the resulting community patterns (Emerson and Gillespie, 2008; Vellend, 2010).

The numerically dominant bacteria detected within the crystalline environment underlying the Chalk River Laboratories (CRL) site in Ontario, Canada, have closest cultivated representatives from the phenotypically diverse Betaproteobacteria, Deltaproteobacteria, Bacteroidetes, Actinobacteria, Nitrospirae and Firmicutes (Beaton et al., 2016; Beaton et al., submitted). Patterns of the 16S rRNA gene at this site revealed that the ecological processes (Vellend, 2010; Vellend, 2016) of dispersal, drift and diversification explain much of phylogenetic diversity and co-occurrence patterns of the component taxa (Beaton et al., submitted). Although nitrogen compounds were not detected within the fracture water, active nitrogen metabolism was detected within this formation (Stroes-Gascoyne et al., 2011). Geochemical analyses of the

drilled cores revealed that the rock porewater contained ammonia, nitrite and nitrate in parts per million concentrations (Figure 5-1) (King-Sharp et al., 2016; Peterman et al., 2016); sulfate was generally more concentrated within the porewater, and there were comparable concentrations of organic and inorganic carbon within the porewater and fracture water.

In this study, we extend our analysis of ecological processes on subsurface microbial community assembly by evaluating the role of selection on the observed community patterns. To do this we compared multiple subsurface sampling locations for evidence of demonstrated metabolism, differential abundance of the 16S rRNA gene and differential abundance of annotated functional genes associated with energy metabolism.

5.3 Materials and Methods

5.3.1 Sampling Groundwater from Fractures

We sampled four locations each from three drilled boreholes -- CR-9 (May 2014), CRG-3 (September 2013) and CRG-6 (May 2013) (Table S1). These boreholes are located between two major faults that bound the study site; these are the Mattawa fault (Ottawa River) and the Maskinonge Lake fault; shown in Figure 5-2. Diabase dykes traversing the site form another boundary. Borehole CRG-3 traverses one of these dykes, borehole CRG-6 traverses two dykes and borehole CR-9 is wholly located between these dykes. Underlying the site are stacked crystalline assemblages consisting of an overlying and underlying garnet-poor assemblage, and a central garnet-rich assemblage. The study site is located within the Central Grenville Gneissic Belt within the Grenville province, the rock drilled from these boreholes dates from 1475 ± 14 to 1045 ± 6 Ma before present (Neymark et al., 2013). The fracture water dates from 12 ka before present (King-Sharp et al., 2016)

Groundwater was collected using a Westbay Multilevel Groundwater Monitoring System (Nova Metrix Groundwater Monitoring (Canada) Ltd.), as described previously (Beaton et al., 2016). Figure S1 shows a schematic of a borehole and an installed Westbay System® that was used to isolate multiple zones within the boreholes; thus, preventing unnatural fracture water flow within the borehole itself. The groundwater sampler consists of four 250 mL stainless steel tubes connected in series by tubing and Swagelok fittings. Prior to each sampling, the tubes

were sterilized by autoclave and the fittings were sterilized by washing them with 70% ethanol. Validation of the sterilization and transport procedures was performed using sterilized water and amplification by polymerase chain reactions (PCR) of the bacterial 16S rRNA gene. Since the tube assemblies contacted only the interior of the casing surface, the probability of introducing surface microbes into the sampled volumes was minimal.

Opening the filled tubes and dispensing the sampled groundwater took place inside of a glove box under an atmosphere of filtered nitrogen gas. Inside the glovebox, the groundwater pH (Beckman PHI 265 pH/Temp/mV meter (Beckman Coulter, Inc.)) and conductivity (YSI Model 30 Conductivity Meter (YSI Inc., Yellow Springs, OH, USA) were measured and 100 mL aliquots were filtered through a 0.45 μm filter (isopore polycarbonate, Merck Millipore, Billerica, MA, USA) and immediately acidified by adding 0.5 mL nitric acid (ultra-trace grade, Seastar™, Baseline®, Fisher Scientific, Ottawa, ON, Canada) for elemental analysis. Elemental composition of the groundwater was determined by inductively coupled plasma-mass spectrometry (ICP-MS, using either a Varian 820-MS (Agilent Technologies, Inc.) or an Element XR (Thermo Scientific)) and by inductively coupled plasma atomic emission spectroscopy (ICP-AES, Optima 3300, Perkin Elmer). Anion concentrations were determined using a Dionex 3000 ICS ion chromatograph (Dionex, Sunnyvale, CA, USA). Dissolved organic carbon (DOC) and dissolved inorganic carbon (DIC) were determined using a Dohrmann, model Phoenix 8000-UV Persulfate TOC Analyzer (Teledyne Teckmar, Mason, OH, USA). The fluorescent dye, fluorescein (ex 488 nm/ em 530 nm), that had been added to the drilling fluid tank to a final concentration of 1000 μg/L, provided a measure of residual drill water.

5.3.2 Characterization of Bacterial Assemblages

Fracture water that was collected for microbial analyses was transferred from the Westbay sample tubes to a sterile glass container and maintained at 5-8 °C till the water was evaluated for most probable number, direct cell count and nucleic acids extraction.

5.3.2.1 Nucleic Acid Extraction and Creation of 16S rRNA Gene Libraries

Whole community nucleic acids were extracted using the UltraClean® DNA Isolation Kit (Qiagen). Within 2 hrs of sampling, fracture water samples were filtered each through a sterile

0.22 µm, 47 mm polyethersulfone filter within a sterile 100 mL filter housing (Merck Millipore Corp., Billerica, MA); filters were either processed immediately or stored overnight at -20°C and processed the next day. The volume filtered for DNA extraction was between 0.8 - 1 L.

5.3.3 16S rRNA Library Preparation and Sequencing

A 16S rRNA gene library was amplified by PCR from each DNA extraction. The primers used in the initial amplification generated amplicons that spanned the V4 region of the 16S rRNA gene between position 519 and 802 (*E. coli* numbering), and produced a PCR amplicon ~300 bp in length. The forward primer (M13-519F: 5'- **GTA AAA CGA CGG CCA G**CA CMG CCG CGG TAA -3') contains the M13 forward primer (in bold), followed by a spacer and the 16S-specific sequence (underlined). The reverse primer (785R: 5'-TAC NVG GGT ATC TAA TCC-3') was taken directly from the primer "S-D-Bact07850b-A-18" in Klindworth et al. (2013). Each 25 µL PCR reaction consisted of 1 x Taq buffer (Fermentas, Glen Burnie, Maryland), 1.5 mM $MgCl_2$, 0.2 mM each dNTP, 0.2 µM of the forward and reverse primer, 0.625 U of Taq DNA polymerase (Fermentas) and 2 µL of extracted DNA. Thermal cycling was carried out in a Techne TC-512 thermal cycler (Techne, Burlington, NJ) using the following conditions: initial denaturation for 3 min at 96°C; 30 cycles of 30 s at 96°C, 45 s at 52°C and 45 s at 72°C; and a final extension for 10 min at 72°C.

Duplicate PCR reactions for samples were combined and purified using Ampure® XP paramagnetic beads (Beckman Coulter, USA) according to the manufacturer's protocol. A second, 6-cycle PCR was used to add a unique 12 bp barcode (Hamady et al., 2008) to each amplicon library using a unique forward primer containing the barcode+M13 forward sequence (5'-3') and the 16S reverse primer, 785R. The resulting barcoded PCR products were purified using Ampure® XP paramagnetic beads (Beckman Coulter), quantified using the Qubit® HS assay (Life Technologies, Carlsbad, CA, USA), pooled in equimolar amounts, and concentrated to a final volume of ≈80 µL using two Amicon® Ultra-0.5 mL 30K Centrifugal Filters (Merck Millipore). The final pooled library was then submitted to the Oklahoma Medical Research Foundation Genomics Facility (Oklahoma City, OK, USA) for sequencing on the MiSeq platform using PE250 V2 chemistry (Illumina Inc., San Diego, CA). All sequences have been deposited in the short read archive of GenBank under Accession No. SRA608335.

5.3.4 Sequence Analysis

5.3.4.1 Sequence Processing and Taxon Identification

After sequencing, reads were merged using Paired End reAd mergeR (PEAR) (Zhang et al., 2014), de-multiplexed in Quantitative Insights Into Microbial Ecology (QIIME) (Caporaso et al., 2010b) then UPARSE (Edgar, 2013) was used to filter the sequences by quality and to cluster them into operational taxonomic units (OTUs) at a 97% cut-off for sequence similarity. After clustering in UPARSE, taxonomy was assigned using the Ribosome Database Project (RDP) Naïve Bayesian classifier (Wang et al., 2007) and the SILVA database (Release 111) (Pruesse et al., 2007). OTUs were aligned using pyNAST (Nearest Alignment Space Termination) (Caporaso et al., 2010a) against an aligned version of the SILVA r111 database, and filtered to remove uninformative bases. Finally, a tree was produced with FastTree (Price et al., 2010) for beta diversity analyses. Beta diversity was estimated using weighted and unweighted UniFrac index (Lozupone and Knight, 2005).

5.3.5 Preparation of Metagenomes

Samples were prepared for metagenome sequencing using the Nextera XT kit from Illumina. First, the DNA was tagmented (tagged and fragmented) by the Nextera XT transposome that simultaneously fragmented the input DNA and added adapter sequences to the ends. In the next step, the tagmented DNA was amplified via a limited-cycle PCR program. The PCR step also added index sequences required for cluster formation. The PCR sample was then cleaned with AMPureXP beads to purify the library DNA and remove short library fragments. Then, each library was normalized to ensure more equal library representation in the pooled sample. Finally, equal volumes of the normalized libraries were combined and then sent to the Oklahoma Medical Research Foundation Genomics Facility (Oklahoma City, OK, USA). All metagenome sequences and annotations are available on the MG-RAST server under project mgp11978.

Further analyses were performed within the R environment (R Core Team, 2015) using packages from the Comprehensive R Archive network (CRAN) and Bioconductor repositories.

5.3.6 Spatial Descriptors

A matrix of Moran's eigenvector maps (MEM) were created from Easting and Northing, zone 18, Universal Transverse Mercator coordinates for each borehole interval sampling location (Table S2) by the method of principle coordinates of neighbor matrices (Borcard and Legendre, 2002; Legendre et al., 2009) using functions from within the R packages 'spdep' (version 0.6-13) and 'adespatial' (version 0.0-8). The functions used to create the spatial weightings matrix were nbtri(), that converts the spatial coordinates of the sampling locations into a distance neighbors map, and the function nb2listw() that creates the weightings matrix from the neighbors map. The eigenvectors with positive Moran's I values reveal different spatial structures over the entire range of scales encompassed by the geographical sampling area. The first positive MEM values generated in the analyses represent broader spatial structures, and the last positive MEM values represent finer spatial structures.

5.3.7 Differential Abundance of the 16S rRNA Gene

The Biological Observation Matrix (BIOM) file produced by QIIME was imported into the Bioconductor package 'phyloseq' (version 1.14.0) (McMurdie and Holmes, 2013). The environmental and spatial data were merged into the phyloseq object as sample data. The datasets within this object were subset to remove samples with low numbers of 16S rRNA gene sequences; the data from CRG-3-11 and CRG-3-14, each with fewer than 600 sequences (Figure S2A), were therefore removed and not included in subsequent analyses. Depending on the analysis being performed, separate phyloseq objects were created; one containing all assemblages and three containing borehole-specific assemblages. Comparisons were made between the environment and spatial variables with the OTU abundances using the function bioenv() (Clarke and Ainsworth, 1993) within the package 'vegan' (version 2.4-3); the correlation was set to Spearman, dissimilarity was set to Bray-Curtis.

To evaluate differential abundances across sampling locations, the borehole specific phyloseq objects were converted into DESeq2 objects using the phyloseq_to_deseq2() function; this conversion allowed the community abundance data to be processed by a function call to the DESeq pipeline within the package Bioconductor 'DESeq2' (version 1.10.1) (Love at al., 2014);

the function call to the pipeline involves normalization, an estimation of dispersion for each gene and fitting to a generalized linear model. This package was developed to perform analyses on RNASeq data but has also been applied to microbiome data (McMurdie and Holmes, 2014). Normalization involves dividing the raw read counts for each taxa by the median count across the samples (creating a virtual reference sample) then computing the size factor as the median of ratios of each sample to the reference sample. The default correction for the false discovery rate was Benjamini-Hochberg (1995). The significance level, alpha, for differential abundance, was set to 0.01. The variables identified as significant by the bioenv analysis were tested as variables in the linear model.

The DESeq function call also calculates a Cook's distance as a diagnostic test for outliers and as a measure of how much a single sample is influencing the fitted coefficients; a large value for Cook's distance indicates an outlier count. The pipeline automatically flags taxa with Cook's distance above a threshold, the value of which is estimated from the sample size and the number of parameters being estimated.

The DESeq2 package was also used to normalize the Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology (KO) annotated genes for energy metabolism that were obtained from the MG-RAST server (Meyer et al., 2008,). Heatmaps of the normalized gene abundances were created using heatmap.2 from the gplots package.

5.3.8 Direct Cell Count and Most Probable Number

Epifluorescent direct cell counting was used to enumerate total cell densities within the groundwater. Triplicate 1 mL volumes of the groundwater were incubated with a DNA intercalating dye then filtered onto black polycarbonate filters (Fisher Scientific, 25 mm, 0.22 µm pore size) for viewing and counting with either a Nikon E600 microscope or a Zeiss Axiophot microscope; at least fifteen fields of view and at least 300 cells were counted per filter for a coefficient of variation of 5.8% per filter. The most probable number method was used to determine viable counts for sulfate and nitrate reduction. Serial dilutions of the fracture water were prepared in triplicate and then mixed with selective media for detecting sulfate reduction (degassed ATCC Medium: 1249 Modified Baar's Medium for Sulfate Reducers) and for partial

and full denitrification (nitrate broth, Difco) – these are referred to as sulfate reducing bacteria (SRB), nitrate utilizing bacteria (NUB) and nitrate reducing bacteria (NRB), respectively. The presence of SRB were determined by the formation of an iron sulfide precipitate after a 4-week incubation period at ambient room temperature, the presence of partial denitrification was determined by the formation of nitrite as detected by its reaction with zinc and for full denitrification by the formation of nitrogen gas that had collected inside inverted Durham tubes throughout the 4-week incubation period at 28°C.

5.4 Results

5.4.1 Abundances and Distributions

Total cell densities amongst sampling locations ranged from 4.7×10^4 to 5.7×10^5 counts/mL (Table S3). All samples displayed nitrogen metabolism as nitrogen reduction and denitrification, with most probable numbers spanning from 28 to 1×10^5 mpn for nitrate reducing bacteria to form nitrite (NUB, Table S3 and Figure 5-3) and from 1 to 150 mpn for denitrifying bacteria (NRB, Table S3 and Figure 5-3). Only samples taken from borehole CRG-6 displayed sulfate reduction to hydrogen sulfide, with most probable numbers spanning from 46 to 460 mpn (SRB, Table S3, Figure 5-3); for the samples taken from boreholes CRG-3 and CR-9, sulfate reduction was either undetected or was low abundance at less than 10 mpn.

The sampling location with the highest mpn for nitrate reduction to nitrite was borehole CRG-3, interval 6 (Table S3, Figure 5-3). This sampling location was sampled three times in succession; the first sampling displayed nitrite mpn that was 19% of the total direct cell density for this sampling (1.1×10^5 mpn and 5.7×10^5 cells/mL, respectively). Direct cell counts and mpn declined with each sampling with the proportion of mpn to total cell density highest for the second sampling at 0.29 mpn/total cell count and lowest for the third sampling at 2 mpn/total cell count.

Sequencing of the 16S rRNA gene and processing of these sequences through the QIIME pipeline identified 435 taxa across all sampling locations. The resulting mean-variance distribution for each taxon is shown in Figure S2B in relation to the expected mean-variance for a Poisson distribution -- where the variance is equal to the mean (solid black line), and in

relation to a fitted mean-variance distribution using the relationship for a negative binomial distribution -- that includes an additional variance term (solid red line). The negative binomial distribution is used within the DESeq2 package to estimate genes that display significant differential expression (Love et al., 2014). We applied this relationship to assess differential abundances of the 16S rRNA gene for taxa within the spatial extent of the sampling locations. Assemblages from each borehole were subset into separate biomes. The basis for creating these biomes were by phylogenetic relatedness of the 16S rRNA gene and decay of similarity with distance between sampling locations (Beaton et al., submitted) that suggested the selected locations were connected by dispersal. Of the 435 total taxa across all sampling locations, 345 were present within the CR-9 biome, 257 were present in the CRG-3 biome and 311 were present within the CRG-6 biome (Table S4). The frequency distribution of the identified taxa (Figure 5-4) shows that there are more taxa at low abundance within each assemblage and only a few taxa are present at higher abundance. Multiple taxa that are present within the whole biome are also differently distributed amongst the corresponding sampling locations with up to 150 taxa at zero frequency, depending on the assemblage (Figure 5-4). The taxa that are unique to each of the three biomes are listed by phylum in Table S4 – the CR-9 biome displays the most taxa per biome and has the highest proportion of unique taxa compared to assemblages within the other two biomes. The drivers for community assembly within the CR-9 biome may differ from the drivers within the CRG-3 or CRG-6 biomes.

5.4.2 Differential Taxa Abundances

Bio-Env analysis (Clarke and Ainsworth, 1993) provides a means for comparing multivariate community data with the multivariate environmental and spatial data. The analysis output lists the variable combinations that provided the best fit based on the corresponding Spearman's rank correlation coefficient. The environmental and spatial data that co-varied with the CR-9 biome and the CRG-6 biome are listed in Table S1 and S2, respectively. For the CR-9 biome, the Bio-Env analysis identified two positive Moran's eigenvector map coefficients, labelled as MEM4 and MEM11, and fracture water manganese; the corresponding Spearman's rank correlation coefficient was 0.94 (Table 5-1). For the CRG-6 biome, the Bio-Env analysis identified fracture water DOC, fracture water sulfate and fracture water manganese; and a

corresponding Spearman's rank correlation coefficient was 1.00 (Table 5-1). The variables identified as being correlated with the whole meta-community biome are listed in Table S5; these were four Moran's eigenvector map coefficients, fracture water organic carbon and fracture water inorganic carbon – with a Spearman's rank correlation coefficient of 0.62. This full biome included the two assemblages from borehole CRG-3 and the four assemblages, each from boreholes CR-9 and CRG-6.

Differential abundances of the 16S rRNA gene within the CR-9 and CRG-6 biomes were analyzed directly from the sequences read counts (McMurdie and Holmes, 2014; Love et al., 2014). The number of sequences obtained for each sample are provided beside the dispersion plot in Figure S2A. The taxa within the CR-9 biome that displayed differential abundance are shown by the OTU number in Figure 5-5 and by taxonomy in Table 5-2. The DESeq2 pipeline identified 3.8% -- 13 out of the 345 taxa -- as being differentially abundant. Six taxa were linked to manganese; these were a *Chlorflexi* (OTU_95), two Nitrospirae (OTU_91 and 58), a *Deltaproteobacteria* (OTU_10), a Candidate_division_BRC1 (OUT_58) and BD-15 (OTU_78); one taxon was linked with the spatial coefficient, MEM11 (OTU_28, an *Actinobacteria* (Table 5-2)). The remaining taxa were linked to both manganese and MEM11; the direction of fold changes were opposed for each variable and the fold change was higher for MEM11.

The spatial coefficient, MEM11, may indicate an unmeasured characteristic of the environment that influences these abundances. Three of the taxa increased in abundance with MEM11 and decreased with manganese: these were a *Nitrospirae* (OTU_67), a BD1-5 (OTU_137) and a *Euryarchaeota* of the Terrestrial_Miscellaneous_Gp (TMEG) (OTU_104). The remaining three taxa decreased in abundance with MEM11 and increased with manganese: these were a *Nitrospirae* (OTU_87), a *Firmicutes* (OTU_88) and a *Betaproteobacteria* (OTU_11).

The taxa within the CRG-6 biome that displayed differential abundance are shown in Figure 5-6 by OTU number and in Table 5-3 by taxonomy. For this biome, the DESeq pipeline identified 3.5% - 11 out of the 311 taxa -- as being differentially abundant. The link between differential abundance and sulfate within this biome corroborates the mpn for sulfate reduction shown in Figure 5-3. Two of the taxa within this biome displayed a fold change with sulfate only, five of the taxa displayed a fold change with DOC only and four of the taxa displayed a fold change

with both sulfate and DOC; the direction of fold change was the same for each variable and the magnitude of the fold change was higher for DOC. The taxa from *Bacteroidetes* (OTU_14 and 21) each decreased with DOC and sulfate; three of the four *Betaproteobacteria* (OTU_13, 77 and 369, but not the OTU_45), the *Deltaproteobacteria* (OTU_135 and 321) and the *Clostridia* (OTU_141 and 268) each increased with DOC and sulfate. The one *Betaproteobacteria* (OTU_45) and the remaining taxon, a *Chlorobi* (OTU_26), decreased with sulfate.

5.4.3 Annotated Functional Genes for Energy Metabolism

The metagenomes that were obtained from processing the shotgun sequence data through the MG-RAST pipeline are available on the MG-RAST server under study 11978. Within the MG-RAST pipeline, the KEGG orthology (KO) annotation database was used to identify those sequences with similarity to evolutionary conserved genes. The KO annotated read counts for functional genes involved in energy metabolism were normalized using the same DESeq procedure (Love et al., 2014) as was used for the 16S rRNA gene amplicons by creating a virtual reference sample and then computing the size factor as the median of ratios of each sample to the reference sample. The resulting data were plotted as a heatmap for genes within the main KO categories for energy metabolism (Figure S5). The most abundant genes within this category (Figure S5) were annotated to proteins involved in oxidative phosphorylation then methane metabolism, nitrogen metabolism, carbon fixation and sulfur metabolism. Annotation to energy metabolism associated with photosynthesis genes ([PATH:ko00195], [PATH:ko00196] and [PATH:ko00710]) were <2500 reads (Figure S5); these annotations were not analyzed further.

For the remainder of the annotations, separate heatmaps were created for the most abundant functional gene categories for energy metabolism: Oxidative phosphorylation [PATH:ko00190] (Figure S6), Methane metabolism [PATH:ko00680] (Figure S7), Carbon fixation pathways in prokaryotes [PATH:ko00720] (Figure S8), Nitrogen metabolism [PATH:ko00910] (Figure 5-7) and Sulfur metabolism [PATH:ko00920] (Figure 5-8). We find that each sampling location displays very similar metabolic capabilities with the most genes annotated to the KO pathway for methane metabolism [PATH:ko00680] at 118 genes annotated out of a possible 171 (Figure S7). This pathway includes genes of the Woods Ljungdahl pathway (Ragsdale and Peirce, 2008)

[PATH:ko00720] at 11 of a possible 103 genes annotated (Figure S8). The annotations within these two pathways means that there is a full complement of genes for acetogenesis in which carbon dioxide and small organic molecules (Ragsdale and Peirce, 2008; Schuchmann and Müller, 2014) are combined with hydrogen gas to form acetate.

The next most highly annotated set of genes found within each sampling location were the genes associated with oxidative phosphorylation, [PATH:ko00190] at 79 of a possible 203 genes annotated (Figure S6). The most abundant genes within this KEGG pathway belong to the subunits forming the proton translocating NADH oxidoreductase of the nuo genes, EC:1.6.5.3; the most abundant representatives being the nuoF, nuoG, nuoL, nuoM, nuoN, nuoD and nuoH (Figure S6). The next most abundant annotated genes involved in oxidative phosphorylation are two F-type H^+ transporting ATPase subunits, EC:3.6.3.14 and an inorganic pyrophosphatase, EC:3.6.1.1 (Figure S6). This combination of genes support the capability of proton motive force generation needed for oxidative phosphorylation coupled to respiration.

Inorganic compounds like nitrate, sulfate and carbon dioxide are reduced by many organisms as sources of cellular nitrogen, sulfur and carbon, respectively. The metabolic processes can be assimilatory, where the reduced nitrogen, sulfur or carbon are incorporated into biomolecules or dissimilatory, where the free energy change of these reduction reactions are coupled to oxidation of more reduced compounds and harnessed in the form of a proton motive force. Inorganic nitrogen compounds are common electron acceptors in anaerobic respiration. Of the most highly annotated set of genes that were associated with nitrogen metabolism [PATH:ko00910], at 50 of a possible 60 genes annotated within this KEGG pathway, the most abundant was a periplasmic nitrate reductase, napA (Figure 5-7,EC:1.7.99.4). This enzyme catalyses the conversion of nitrate to nitrite and thus may support respiration (Stewart et al., 2002). The second most abundant gene associated with nitrogen metabolism was a nitrite reductase, nirB (Figure 5-7, EC:1.7.1.4), that catalyzes a six-electron reduction of nitrite to form ammonia. Both napA and nirB genes form part of the nitrate reduction IV pathway (Liu and Peck, 1981; Liu et al., 1988).

Annotated genes for both a hydroxylamine reductase (Figure 5-7, EC:1.7.99.1) and a hydroxylamine oxidase, hao (Figure 5-7, EC:1.7.3.4) suggest that nitrogen conversion can occur

in the direction from nitrate to ammonia and from ammonia back to nitrite. There are also genes for the conversion of nitrite to nitrogen gas via nitric oxide and nitrous oxide (Figure 5-7), and multiple annotated genes for nitrogenases, nif (Figure 5-7, EC:1.18.6.1) that can fix nitrogen to form ammonia. The presence and abundance of these genes suggest that there may be a capability for a complete nitrogen cycle within each of the sampling locations. The measured formation of nitrite from nitrate within all of the sampling locations by most probable number (Figure 5-3) may therefore represent a subset of the *in situ* nitrogen metabolizing capability.

Inorganic sulfur compounds are also common electron acceptors in anaerobic respiration. Of the most highly annotated set of genes that were associated with sulfur metabolism [PATH:ko00910], at 11 of a possible 99 genes annotated within this KEGG pathway, the most abundant were genes associated with substrate level phosphorylation (Figure 5-8) for an adenylylsulfate reductase gene, aprA, EC:1.8.99.2 and a sulfate adenylyltransferase subunit 2 gene, EC:2.7.7.4. The first step in this reaction is the adenosine monophosphate (AMP) dependent oxidation of sulfite to form adenosine phosphosulfate (APS) (Figure 5-8); next the AMP is reacted with a diphosphate to form adenosine triphosphate (ATP) (Peck, 1968); this reaction can also operate in the reducing direction, consuming ATP to reduce sulfate to form APS and then sulfite. Two annotated genes for sulfite reduction are also abundant, the cysL/cysJ genes (Figure 5-8, EC:1.8.1.2) and the sir gene (Figure 5-8, EC:1.8.7.1). The assimilatory sulfite reductases catalyze a six-electron reduction of sulfite to form sulfide, this sulfide is then used in the biosynthesis of the amino acid cysteine (Kredich, 2008).

5.5 Discussion

Three factors that influence community composition and diversity are niche-based biotic and abiotic competition (Keddy, 1992; Chesson, 2000; Adler et al., 2007; Mayfield and Levine, 2010), neutral assembly of ecologically equivalent component taxa (Hubbell, 2001), and prior events such as dispersal (Ricklefs, 1987). The phylogenetic information within the 16S rRNA gene allows for the distributions and co-occurrences of multiple microbial taxa to be compared with corresponding environmental and spatial information under an ecological framework of community dynamics (Vellend, 2010; Vellend, 2016); the four main processes are: selection,

ecological drift, dispersal and diversification. At the spatial extent of the meta-community, randomness accounts for close to 50% of the variance in the phylogenetic beta diversity.

On large spatial scales, microbial community diversity can correlate with environmental factors such as salinity, pH and with environments (Lozupone and Knight, 2007; Lauber et al., 2009; Fierer et al., 2012; Stamps et al., 2016). At local scales, ecosystem functions such as the dynamics of carbon, nitrogen and phosphorous can also vary with community diversity (Wagg et al., 2014). In our studies, we found that much of the observed differences in community diversity across the extent of the sampling locations underlying CRL were explained by the ecological processes of dispersal, drift and diversification (Beaton et al., submitted). The local clustering of related taxa suggests that the taxa that enter the subsurface with recharge share ecological traits and fitness characteristics (Cadotte et al., 2009; Mayfield and Levine, 2010) and that local spatial heterogeneity could allow related taxa to co-exist (Lear, 2014). In this study, we explored the role of selection on community patterns by testing the fracture water for culturability of the component taxa on selective media (Figure 5-3); by testing for co-variance of the 16S rRNA gene with the environmental and spatial data (Table 5-1); and by comparing differential abundances of 16S rRNA gene between sampling locations connected by dispersal (Figure 5-5 and Table 5-2, Figure 5-6 and Table 5-3). We also compared the abundances of annotated genes from metagenomes constructed from the fracture water samples (Figure S5-Figure S8 and Figure 5-7and Figure 5-8).

Selection associated with differential abundance of the 16S rRNA gene identified environmental and spatial factors: sulfate and organic carbon plus manganese and a spatial coefficient. Selection by differential abundance, however, was not a major driver of community assembly as it accounted for only ~8% of the total abundances. At a spatial scale of less than the 1 km^3 from this study, it may be possible to identify either greater significance for sulfate and manganese or to additional environmental factors.

5.5.1 Nitrogen Metabolism

Fracture water from all sampling locations displayed nitrogen metabolism; nitrate reduction to nitrite was the dominant reaction (blue bars, Figure 5-3). The KO annotated genes further

suggest that complete nitrogen cycling is possible, including nitrogen fixation by multiple nif genes (Figure 5-7). The most abundant gene detected for nitrogen metabolism was a gene for the catalytic subunit of nitrate reductase that localizes to the periplasm of Proteobacteria -- the napA gene (Figure 5-7); the Proteobacteria were prevalent and abundant across all sampling locations. Other abundant nap genes, the ferredoxin-type proteins -- napH and napG -- and the cytochrome c-type proteins -- napC and napB -- would likely form part of an electron transfer complex with the catalytic subunit.

Periplasmic nitrate reductases are functionally diverse proteins that catalyze the dissimilatory conversion of nitrate to nitrite within both the denitrification pathway and the ammonia production pathway (Spararacino-Watkins et al., 2014); these proteins also help to maintain the cell redox potential and to scavenge nitrate. The periplasmic location of the nitrate reductase means that its role in respiration is limited. Respiration linked nitrate reduction at the sampling locations appears to be unlikely since the membrane associated near nitrate reductase genes, that are associated with the cell membrane and therefore is involved in respiration, were not among the abundant genes.

The diverse roles of nap proteins can also mean that the presence of a napA gene within a genome does not necessarily relate to a particular function; for example, the napA gene copy numbers remained constant along a gradient of nitrate and ammonia contaminated sediment -- but the gene copies for narG, that is associated with respiration, declined with distance from the contaminant source (Smith et al., 2007). The proteins coded for by the abundant nitrite reducing nir genes and nitrogen fixing nif genes consume reducing equivalents by converting nitrite and nitrogen gas to ammonia, respectively (Figure 5-7) (Wang and Gunsalus, 2000). It is possible, therefore, that within the subsurface of CRL, nitrogen compounds are used by multiple phyla to dissipate reducing equivalents instead of reducing nitrogen compounds as part of energy conservation or nitrogen assimilation.

5.5.2 Sulfur Metabolism

Sulfur is an essential component of microbial metabolism. In our tests, sulfur metabolism was detected within a subset of sampling locations (Figure 5-3). Sulfate is unreactive under

physiological conditions, and therefore must be activated before entering cellular metabolism. The selective medium used for most probable number contains iron that is used to detect sulfate reduction to sulfide by the formation of a black iron sulfide precipitate; formation of this precipitate is interpreted as evidence for dissimilatory sulfate reduction. At all sampling locations, the most abundant genes annotated to the sulfur pathway were for genes associated with sulfur assimilation; the aprA, cys and sir genes (Figure 5-8). Part of this pathway can also produce ATP by substrate level phosphorylation (Peck, 1968). Genes associated with dissimilatory sulfite reductase were not represented within the abundant annotated genes, suggesting that the observed sulfur metabolism within the fracture water was not linked to a proton motive force.

The sulfur assimilation pathway is activated by environmental signals that are mediated by guanosine triphosphate (GTP) binding (G) proteins; these proteins act as molecular switches regulating gene expression (Caldon et al., 2001; Wittinghofer and Vetter, 2011). The cysN protein in assimilatory sulfur metabolism is a G protein that, in a complex with a cysD protein, forms the functional sulfate adenylyltransferase (Leyh, 1988); the activity of this complex activates sulfate to adenosine phosphosulfate (APS) in the first step of sulfate assimilation. For example, upregulation of cys proteins is observed when cells are under stress from fermentation products and from oxidation (Pinto et al., 2004; Brown et al., 2006; Wang et al., 2009; Wang et al., 2013). The episode of observed sulfate reduction (Figure 5-3) might reflect a stress response stimulating cysteine and methionine biosynthesis.

5.5.3 Differential Abundance

A primary role of microbial taxa within the biosphere is as catalysts of biogeochemical cycles. The rate of energy flow through an ecosystem would depend on the size of the microbial population and their activity. In this study, the fracture subsurface holds a high degree of functional redundancy in genes associated with energy metabolism; it also hosts a high proportion phylogenetically related taxa. Trait-based biotic selection, therefore, may be a driver forming these communities.

Ecological drift--from deterministic expectations--causes species abundances to fluctuate randomly and can result in local extinction, especially for species at low relative abundance. Microbial communities display distributions in which most of the taxa tend to be found in low relative abundances (Nemergut et al., 2013). Natural communities like these are thought to be structured by niche differences and competition among taxa; and that these differences are thought to modulate the effects of drift (Gilbert and Levine, 2017). A study of plant communities found that competitive differences between taxa created conditions in which drift played a strong role--rather than being lessened--by favoring a subset of species and leaving most other species at small population sizes (Gilbert and Levine, 2017). This effect of drift occurred despite stabilizing niche differences that would favor local low abundant taxa through greater population growth rates. Density dependent traits, such as quorum sensing in bacteria, would also influence the balance between drift and selection.

The abundance distributions observed within the fracture water biomes (Figure 4) shows that there are numerous lower abundant taxa and numerous zero occurrences within the individual sampling locations that comprise each biome. A negative binomial distribution was used to calculate differential abundances of individual taxa within the CR-9 biome (Figure 5-5) and the CRG-6 biome (Figure 5-6); the log-ratio-average abundance (MA) plot (Figure S3) reveals that the differentially abundant taxa among each biome were few -- with less than 4% of the component taxa displaying significant changes in abundance; this finding contrasts with overall viability that ranged from 10-75% based on reduction of the 5-Cyano-2,3-di-(p-tolyl) tetrazolium chloride (CTC) (Beaton et al., 2017). The differentially abundant taxa spanned the range of average abundances across all the taxa, so growth advantage may be important but was limited to a few taxa at the time of sampling.

The significant environmental parameters associated with abundance were sulfate and organic carbon within the CRG-6 biome. Despite the prevalence and abundance of the *Proteobacteria* within this biome, only a few taxa from the *Betaproteobacteria* and the *Deltaproteobacteria* were found to be differently abundant (Figure 5-5 and Figure 5-6); the abundances of taxa from the *Betaproteobacteria* were associated with both sulfate and organic carbon, and the abundances of taxa from the *Deltaproteobacteria* were associated with organic carbon only;

two taxa from the prevalent and abundant *Clostridia* were also differently abundant with sulfate and organic carbon. Considering the metabolic genes for sulfur metabolism (Figure 5-8); these connections between taxa abundance may be due to sulfur assimilation and not to dissimilatory sulfate reduction.

The significant environmental parameters associated with abundance within the CR-9 biome were manganese and an unknown but spatially correlated parameter: a Moran eigenvector map coefficients referred to as MEM11. The abundance fold changes within this biome for taxa from the *Nitrospira* (Figure 5-5) suggest that an unknown spatially correlated parameter and soluble manganese may be involved in nitrification reactions within the fracture water. Coupling between manganese and the nitrogen redox cycle has been observed in other environments; for example, a decrease in the number of gene copies for ammonia oxidation in soils (Xin et al., 2016), the coupling nitrification and denitrification in marine sediments (Hulth et al., 1999) and denitrification of the nitrate produced when ammonia is formed by the degradation of organic matter (Mogollón et al., 2016). In this study, fracture water manganese and the CTC viable populations across sampling locations displayed similar distributions (Beaton et al., 2017); manganese in the fracture water is either directly involved in microbial processes or it co-varies with, for example, weathering reactions.

5.5.4 Conclusions

Given the role of randomness on subsurface community assembly, long term predictions of natural subsurface microbial process that are based solely on selection risk misattributing ecosystem functions.

The functional redundancy of genes for energy metabolism combined with the diversity and prevalence of representatives from, for example, *Proteobacteria*, *Clostridia*, *Actinobacteria* and *Acidobacteria*, suggest that trait-based biotic selection supports multiple related taxa.

Niche-based biotic selection plays a limited role in the observed patterns of subsurface community assembly; accounting for 8% of the differential abundance of the 16S rRNA gene V4 regions and 1% or less of the cultivable fraction. The differentially abundant taxa that were observed correlated with the fracture water organic carbon and sulfate, and with the fracture

water manganese plus a potentially unknown (unmeasured) environmental parameter associated with a spatial coefficient.

Nitrogen metabolism was prevalent across all sampling locations. The porewater composition of nitrogen compounds and the observed differential abundances of *Nitrospira* suggest that nitrogen cycling may be a major biogeochemical process within this subsurface. Sulfate metabolism was detected within one biome, suggesting that sulfate reduction is either localized within the extent of the study, or its reduction is episodic. These data provide a basis for studying subsurface redox dynamics relevant to proposed geologic repositories.

5.6 Author Contributions

All authors collected and contributed data sets for analysis as well as participated in the conceptual drafting and revision of this manuscript. Karen J. King-Sharp oversaw the groundwater sampling and geochemical analyses. Sim Stroes-Gascoyne oversaw the microbial most probable number analysis. Blake W. Stamps and Heather S. Nunn performed sequencing and sequence data analysis. Danielle Beaton conducted subsequent analyses on the sequence data and was the primary author in writing and revising the manuscript. Bradley S. Stevenson and Marilyne Stuart contributed significantly in manuscript development and revision.

5.7 Acknowledgements

We gratefully acknowledge the contributions from David Larssen, Westbay Instruments, Vancouver, British Columbia.

5.8 Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

5.9 References

Adler, P.B., Hillerislambers, J. and Levine, J.M., 2007, “A niche for neutrality”, Ecology Letters, 10: 95–104, doi:10.1111/j.1461-0248.2006.00996.x.

Baas-Becking, L.G.M., 1934, “Geobiologie of inleiding tot de milieukunde”, WP Van Stockum & Zoon, The Hague, The Netherlands.

Beaton, E.D., Stevenson, B.S., King-Sharp, K.J., Stamps, B.W., Nunn, H.S., Stuart, M. and Cornett, J., Submitted, "The Influences of Taxa Divergence, Dispersal and Drift on Microbial Co-Occurrence and Abundance Patterns within Subsurface Crystalline Fractures", Manuscript submitted.

Beaton, E.D., Stevenson, B.S., King-Sharp, K.J., Stamps, B.W., Nunn, H.S. and Stuart, M., 2016, "Local and regional diversity reveals dispersal limitation and drift as drivers for groundwater bacterial communities from a fractured granite formation". Front. Microbiol., 7:1933 10.3389/fmicb.2016.01933.

Benjamini, Y. and Hochberg, Y., 1995, "Controlling the False Discovery Rate: A Practical and Powerful Approach to multiple testing", Journal of the Royal Statistical Society, series B, 57,289-300.

Bier, R.L., Bernhardt, E.S., Boot, C.M., Graham, E.B., Hall, E.K., Lennon, J.T., Nemergut, D.R., Osborne, B.B., Ruiz-González, C., Schimel, J.P., Waldrop, M.P. and Wallenstein, M.D., 2015, "Linking Microbial Community Structure and Microbial Processes: An Empirical and Conceptual Overview", FEMS Microbiology Ecology, 2015, 91, doi: 10.1093/femsec/fiv113.

Borcard, D. and Legendre, P., 2002, "All-Scale Spatial Analysis of Ecological Data by Means of Principal Coordinates of Neighbour Matrices", Ecological Modelling, 153. 51-68, DOI: 10.1016/S0304-3800(01)00501-4.

Brown, S.D., Thompson, M.R., VerBerkmoes, N.C, Chourey, K., Shah, M., Zhou, J., Hettich, R.L. and Thompson, D.K., 2006, "Molecular Dynamics of the Shewanella oneidensis Response to Chromate Stress", Mol Cell Proteomics, 5: 1054-1071, doi:10.1074/mcp.M500394-MCP200.

Cadotte, M.W., Hamilton, M.A. and Murray, B.R., 2009, "Phylogenetic relatedness and plant invader success across two spatial scales", Diversity and Distributions, 15, 481-488.

Caldon, C.E., Yoong, P. and March, P.E., 2001, “Evolution of a molecular switch: universal bacterial GTPases regulate ribosome function”, Molecular Microbiology, 41: 289–297, doi:10.1046/j.1365-2958.2001.02536.x.

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L. and Knight, R., 2010a, “PyNAST: a flexible tool for aligning sequences to a template alignment”, Bioinformatics, 26 266–267, 10.1093/bioinformatics/btp636.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsuenenko, T., Zaneveld, J. and Knight, R., 2010b, “QIIME allows analysis of high-throughput community sequencing data”, Nat Methods, 7:335–336, doi:10.1038/nmeth.f.303.

Chesson, P., 2000, “Annual Review of Ecology and Systematics”, Annual Reviews, 31:1, 343-366.

Chivian, D., Brodie, E.L., Alm, E.J., Culley, D.E., Dehal, P.S., Desantis, T.Z., Gihring, T.M., Lapidus, A., Lin, L-H, Lowry, S.R., Moser, D.P., Richardson, P.M., Southam, G., Wanger, G., Pratt, L.M., Andersen, G.L., Hazen, T.C., Brockman, F.J., Arkin, A.P., Tullis, C. and Onstott, T.C., 2008, “Environmental Genomics Reveals a Single-Species Ecosystem Deep Within Earth”, Science, 10, 275-278, DOI: 10.1126/science.1155495.

Clarke, K.R and Ainsworth, M., 1993, “A method of linking multivariate community structure to environmental variables”, Marine Ecology Progress Series, 92, 205–219.

Edgar, R.C., 2013, “UPARSE: highly accurate OTU sequences from microbial amplicon reads”, Nat Methods, 10:996–998. doi:10.1038/nmeth.2604.

Emerson, B.C. and Gillespie, R.G., 2008, “Phylogenetic analysis of community assembly and structure over space and time”, Trends Ecol. Evol., 23 619–630, 10.1016/j.tree.2008.07.005.

Fierer, N., Leff, J.W., Adams, B.J., Nielsen, U.N., Bates, S.T., Lauber, C.L., Owens, S., Gilbert, J.A., Wall, D.H. and Caporaso, J.G., 2012, “Cross-biome metagenomic analyses of soil

microbial communities and their functional attributes", Proceedings of the National Academy of Sciences, 109, 21390-21395, doi:10.1073/pnas.1215210110.

Gilbert, B. and Levine, J.M., 2017, "Ecological drift and the distribution of species diversity", Proc. Biol. Sci., 284 (1855), doi: 10.1098/rspb.2017.0507.

Goberna, M., Navarro-Cano, J.A., Valiente-Banuet, A., García, C. and Verdú, M., 2014, "Abiotic stress tolerance and competition related traits underlie phylogenetic clustering in soil bacterial communities", Ecol. Lett., 17 1191–1201, 10.1111/ele.12341.

Graham, E.B., Knelman, J.E., Schindlbacher, A., Siciliano, S., Breulmann, M., Yannarell, A., Beman, J.M., Abell, G., Philippot, L., Prosser, J., Foulquier, A., Yuste, J.C., Glanville, H.C., Jones, D.L., Angel, R., Salminen, J., Newton, R.J., Bürgmann, H., Ingram, L.J., Hamer, U., Siljanen, H.M.P., Peltoniemi, K., Potthast, K., Bañeras, L., Hartmann, M., Banerjee, S., Yu, R-Q., Nogaro, G., Richter, A., Koranda, M., Castle, S.C., Goberna, M., Song, B., Chatterjee, A., Nunes, O.C., Lopes, A.R., Cao, Y., Kaisermann, A., Hallin, S., Strickland, M.S., Garcia-Pausas, J., Barba, J., Kang, H., Isobe, K., Papaspyrou, S., Pastorelli, R., Lagomarsino, A., Lindström, E.S., Basiliko, N. and Nemergut, D.R., 2016, "Microbes as Engines of Ecosystem Function: When Does Community Structure Enhance Predictions of Ecosystem Processes?", Frontiers in Microbiology, 7, 214, http://journal.frontiersin.org/article/10.3389/fmicb.2016.00214.

Griebler, C. and Lueders, T., 2009, "Microbial Biodiversity in Groundwater Ecosystems", Freshwater Biology, 54, 649–677. doi:10.1111/j.1365-2427.2008.02013.x.

Hallbeck, L. and Pedersen, K., 2012, "Culture-dependent comparison of microbial diversity in deep granitic groundwater from two sites considered for a Swedish final repository of spent nuclear fuel", FEMS Microbiol. Ecol., 81 66–77. 10.1111/j.1574-6941.2011.01281.x.

Hamady, M., Walker, J.J., Harris, J.K., Gold, N.J. and Knight, R., 2008, "Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex", Nat Methods, 5:235–237, doi: 10.1038/nmeth.1184.

Horner-Devine, M.C., Carney, K.M. and Bohannan, B.J., 2004, “An ecological perspective on bacterial biodiversity”, Proceedings of Biological Sciences, 271, 113-22.

Horner-Devine, M.C. and Bohannan, B.J.M., 2006, “Phylogenetic clustering and overdispersion in bacterial communities”, Ecology, 87 S100–S108, 10.1890/0012-9658.

Hubalek, V., Wu, X., Eiler, A., Buck, M., Heim, C., Dopson, M., Bertilsson, S. and Lonescu, D., 2016, “Connectivity to the surface determines diversity patterns in subsurface aquifers of the Fennoscandian shield”, ISME Journal, 10 (10), pp. 2447-2458.

Hulth, S., Aller, R.C. and Gilbert, F., 1999, “Coupled anoxic nitrification/manganese reduction in marine sediments”, Geochimica et Cosmochimica Acta, 63, 49-66, doi:org/10.1016/S0016-7037(98)00285-3.

Itävaara, M., Nyyssönen, M., Kapanen, A., Nousiainen, A., Ahonen, L. and Kukkonen, I., 2011, “Characterization of bacterial diversity to a depth of 1500 m in the outokumpu deep borehole, fennoscandian shield”. FEMS Microbiol. Ecol., 77 295–309, 10.1111/j.1574-6941.2011.01111.x.

Itävaara, M., Salavirta, H., Marjamaa, K. and Ruskeeniemi, T., 2016, “Geomicrobiology and Metagenomics of Terrestrial Deep Subsurface Microbiomes”, Advances in Applied Microbiology, 94, pp. 1-77.

Jain, D.K., Providenti, M., Tanner, C., Cord, I. and Stroes-Gascoyne, S., 1997, “Characterization of microbial communities in deep groundwater from granitic rock”, Can. J. Microbiol., 43, 272–283, doi: 10.1139/m97-038.

Keddy, P.A., 1992, “Assembly and response rules: two goals for predictive community ecology”, Journal of Vegetation Science, 3: 157–164, doi:10.2307/3235676.

Kredich, N.M., 2008, “Biosynthesis of Cysteine”, EcoSal Plus, 3(1), DOI: 10.1128/ecosalplus.3.6.1.11.

King-Sharp, K.J., Frape, S.K., Peterman, Z., Gwynne, R., Tian, L. and Gurban, I., 2016, “Synthesis of Geochemical and Fracture Mineral Studies Relevant to a Deep Geological Repository

for Nonfuel Wastes at Chalk River", Canadian Nuclear Review, 10.12943/CNR.2016.00015.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, Q., Horn, M. and Glöckner, F.O., 2013, "Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies", Nucleic Acids Research, 41, e1, doi: 10.1093/nar/gks808.

Konno, U., Kouduka, M., Komatsu, D.D., Ishii, K., Fukuda, A., Tsunogai, U., Ito, K. and Suzuki, Y., 2013, "Novel Microbial Populations in Deep Granitic Groundwater from Grimsel Test Site, Switzerland", Microbial Ecology, 65 (3), pp. 626-637.

Küsel, K., Totsche, K.U., Trumbore, S.E., Lehmann, R., Steinhäuser, C. and Herrmann, M., 2016, "How Deep Can Surface Signals be Traced in the Critical Zone? Merging Biodiversity with Biogeochemistry Research in a Central German Muschelkalk Landscape", Frontiers in Earth Science, 4, art. no. 32.

Lauber, C.L., Hamady, M., Knight, R. and Fierer, N., 2009, "Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale", Applied Environmental Microbiology, 75 5111-5120, doi:10.1128/AEM.00335-09.

Lear, G., Washington, V., Neale, M., Case, B., Buckley, H. and Lewis, G., 2013, "The Biogeography of Stream Bacteria", Global Ecology and Biogeography, 22: 544–554, doi:10.1111/geb.12046.

Lear, G., Bellamy, J., Case, B.S., Lee, J.E. and Buckley, H.B., 2014, "Finescale Spatial Patterns in Bacterial Community Composition and Function within Freshwater Ponds" ISME J., 8, 1715–1726, doi: 10.1038/ismej.2014.21.

Legendre, P., Mi, X, Ren, H., Ma, K., Yu, M., Sun, I.F. and He, F., 2009, "Artitioning beta diversity in a subtropical broad-leaved forest of China", Ecology, 90:663-7.

Liu, M.C. and Peck, H.D., 1981, "The isolation of a hexaheme cytochrome from Desulfovibrio desulfuricans and its identification as a new type of nitrite reductase", Journal of Biological Chemistry, 256, 13159-13164, PMID: 7309757.

Liu, M.C., Bakel, B.W., Liu, M.Y. and Dao, T.N., 1988, “Purification of Vibrio fischeri nitrite reductase and its characterization as a hexaheme c-type cytochrome”, Arch Biochem Biophys, 262(1); 259-65, PMID: 2833168.

Love, M.I., Huber, W. and Anders, S., 2014, “Moderated Estimation of Fold Change and Dispersion for RNA-Seq data with DESeq2” Genome Biology, 15, pp. 550. doi: 10.1186/s13059-014-0550-8.

Lozupone, C. and Knight, R., 2005, “UniFrac: A New Phylogenetic Method for Comparing Microbial Communities”, Applied and Environmental Microbiology, 71: 8228–8235.

Lozupone, C.A. and Knight, R., 2007, “Global patterns in bacterial diversity”, Proceedings of the National Academy of Sciences, 104: 11436–11440, doi: 10.1073/pnas.0611525104.

Lozupone, C., Lladser, M.E., Knights, D., Stombaugh, J. and Knight, R., 2011, “UniFrac: an Effective Distance Metric for Microbial Community Comparison”, The ISME Journal, 5: 169–172.

Martiny, J.B.H., Eisen, J.A, Penn, K., Allison, S.D. and Horner-Devine, M.C., 2011, “Drivers of Bacterial β-diversity Depend on Spatial Scale”, Proceedings of the National Academy of Sciences, 108, 7850-7854, doi:10.1073/pnas.1016308108.

Mayfield, M.M. and Levine, J.M., 2010, “Opposing effects of competitive exclusion on the phylogenetic structure of communities”, Ecology Letters, 13, 1085-1093.

McMahon, S. and Parnell, J., 2014, “Weighing the deep continental biosphere”, FEMS Microbiology Ecology, 87 (1), pp. 113-120.

McMurdie, P.J. and Holmes, S., 2013, “Phyloseq: An R Package for reproducible interactive analysis and graphics of microbiome census data”, PLoS ONE, 8(4), pp. e61217, http://dx.plos.org/10.1371/journal.pone.0061217.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Wilke, A., Wilkening, J. and Edwards, R.A., 2008, “The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes”, BMC Bioinformatics, 20089-386, DOI: 10.1186/1471-2105-9-386.

Mogollón, J.M.,Mewes, K. and Kasten, S., 2016, “Quantifying manganese and nitrogen cycle coupling in manganese-rich, organic carbon-starved marine sediments: Examples from the Clarion-Clipperton fracture zone”, Geophysical Research Letters, 43, 7114-7123, doi:10.1002/2016GL069117.

Nemergut, D.R., Costello, E.K., Hamady, M., Lozupone, C., Jiang, L., Schmidt, S.K., Fierer, N., Townsend, A.R., Cleveland, C.C., Stanish, L. and Knight, R., 2011, “Global patterns in the biogeography of bacterial taxa”, Environmental Microbiology, 13, 135-44, doi: 10.1111/j.1462-2920.2010.02315.x.

Neymark, L.A., Peterman, Z.E., Moscati, R.J. and Thivierge, R.H., 2013, “U–Pb, Rb–Sr, and U-series isotope geochemistry of rocks and fracture minerals from the Chalk River Laboratories site, Grenville Province, Ontario, Canada”, Applied Geochemistry, 36, 10-33, https://doi.org/10.1016/j.apgeochem.2013.06.004.

Nyyssönen, M., Bomberg, M., Kapanen, A., Nousiainen, A., Pitkänen, P. and Itävaara, M., 2012, “Methanogenic and sulphate-reducing microbial communities in deep groundwater of crystalline rock fractures in Olkiluoto, Finland”, Geomicrobiol. J., 29 863–878, 10.1080/01490451.2011.635759,

Nyyssönen, M., Hultman, J., Ahonen, L., Kukkonen, I., Paulin, L., Laine, P., Itävaara, M. and Auvinen, P., 2014, “Taxonomically and functionally diverse microbial communities in deep crystalline rocks of the Fennoscandian shield”, ISME J., 8 126–138, 10.1038/ismej.2013.125.

Peck, H.D., 1968, “Energy-coupling mechanisms in chemolithotrophic bacteria”, Annual Reviews in Microbiology, 22, 489-518.

Pedersen, K., 2013, “Metabolic activity of subterranean microbial communities in deep granitic groundwater supplemented with methane and H 2”, ISME Journal, 7 (4), pp. 839-849.

Peterman, Z., Neymark, L., King-Sharp, K.J. and Gascoyne, M., 2016, “Isotope hydrology of the Chalk River Laboratories site, Ontario, Canada”, Applied Geochemistry, 66, 149-161.

Pinto, R., Tang, Q., Britton, W., Leyh, T. and Triccas, J., 2004, "The Mycobacterium tuberculosis cysD and cysNC genes form a stress-induced operon that encodes a tri-functional sulfate-activating complex", Microbiology, 150(6):1681-1686, doi:10.1099/mic.0.26894-0.

Price, M.N., Dehal, P.S. and Arkin, A.P., 2010, "FastTree 2 – approximately maximum-likelihood trees for large alignments", PLOS One, 5:e9490, doi: 10.1371/journal.pone.0009490.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J. and Glöckner, F.O., 2007, "SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB". Nucleic Acids Res, 35:7188–7196, doi: 10.1093/nar/gkm864.

Purkamo, L., Bomberg, M., Nyyssönen, M., Kukkonen, I., Ahonen, L. and Itävaara, M., 2014, "Heterotrophic Communities Supplied by Ancient Organic Carbon Predominate in Deep Fennoscandian Bedrock Fluids", Microbial Ecology, 69 (2), pp. 319-332.

Purkamo, L., Bomberg, M., Kietäväinen, R., Salavirta, H., Nyyssönen, M., Nuppunen-Puputti, M., Ahonen, L., Kukkonen, I. and Itävaara, M., 2016, "Microbial co-occurrence patterns in deep Precambrian bedrock fracture fluids", Biogeosciences, 13 (10), pp. 3091-3108.

R Core Team, 2015, "R: A language and environment for statistical computing", R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/.

Ragsdale, S.W. and Pierce, E., 2008, "Acetogenesis and the Wood-Ljungdahl pathway of CO(2) fixation", Biochim Biophys Acta., 84(12):1873-98, doi: 10.1016/j.bbapap.2008.08.012.

Ricklefs, R.E., 1987, "Community diversity: relative roles of local and regional processes", Science, 235(4785):167-71, DOI: 10.1126/science.235.4785.167.

Rodriguez, G.G., Phipps, D., Ishiguro, K. and Ridgway, H.F., 1992, "Use of a fluorescent redox probe for direct visualization of actively respiring bacteria", Applied Environmental Microbiology, 58, 1801-1808.

Sahl, J.W., Schmidt, R., Swanner, E.D., Mandernack, K.W., Templeton, A.S., Kieft, T.L., Smith, R.L., Sanford, W.E., Callaghan, R.L., Mitton, J.B. and Spear, J.R., 2008, "Subsurface

microbial diversity in deep-granitic-fracture water in Colorado", Appl. Environ. Microbiol., 74, 143–152, 10.1128/AEM.01133-07.

Schuchmann, K. and Müller, V., 2014, "Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria", Nature Reviews Microbiology, 12, 809-821.

Smith, C.J., Nedwell, D.B., Dong, L.F. and Osborn, A.M., 2007, "Diversity and Abundance of Nitrate Reductase Genes (narG and napA), Nitrite Reductase Genes (nirS and nrfA), and Their Transcripts in Estuarine Sediments", Applied and Environmental Microbiology, 73:11, 3612-3622, doi:10.1128/AEM.02894-06.

Spararacino-Watkins, C., Stiltz, J.F. and Basu, P., 2014, "Nitrate and Periplasmic Nitrate Reductases", Chemical Society Reviews, 43, 676-706, doi:10.1039/c3cs60249d.

Stamps, B.W., Lyles, C.N., Suflita, J.M., Masoner, J.R., Cozzarelli, I.M., Kolpin, D.W. and Stevenson, B.S., 2016, "Municipal Solid Waste Landfills Harbor Distinct Microbiomes", Frontiers in Microbiology, 7, 534, DOI=10.3389/fmicb.2016.00534.

Stegen, J.C., Lin, X., Fredrickson, J.K., Chen, X., Kennedy, D.W. and Murray, C.J., Rockland, M.L. and Konopka, A., 2013, "Quantifying community assembly processes and identifying features that impose them" ISME J., 7(11), 2069–2079, 10.1038/ismej.2013.93.

Stewart, V., Lu, Y. and Darwin, A.J., 2002, "Periplasmic nitrate reductase (NapABC enzyme) supports anaerobic respiration by Escherichia coli K-12", Journal of Bacteriology, 184, 1314-23.

Strange, R.W., Dodd, F.E., Abraham, Z.H.L., Grossmann, J.G., Brüser, T., Eady, R.R., Smith, B.E. and Hasnain, S.S., 1995, "The substrate-binding site in Cu nitrite reductase and its similarity to Zn carbonic anhydrase", Nature Structural Biology, 2, 287-292, doi:10.1038/nsb0495-287.

Stroes-Gascoyne, S., Hamon, C.J., Audette-Stuart, M., Beaton, E.D., King-Sharp, K., Festarini, A., Serran, M., McMullin, D., Kramer-Tremblay, S., Rose, S. and Bellan, L., 2011, "Microbial Characterization of Groundwater from Boreholes CR9 and CR18 at CRL (2007–2009) –

Implications for a Possible Future Repository for Radioactive Non-fuel Waste", Canadian Nuclear Society, Toronto, ON.

Vellend, M., Cornwell, W.K., Magnuson-Ford, K. and Mooers, A., 2010, "Measuring phylogenetic biodiversity," in Biological Diversity: Frontiers in Measurement and Assessment, Chap. 14, eds A.E. Magurran and B.J. McGill (Oxford: Oxford University Press).

Vellend, M., 2016, "The Theory of Ecological Communities", Monograph in Population Biology, 57, Princeton University Press.

Wagg, C., Bender, S.F., Widmer, F. and van der Heijden, M.G.A., 2014, "Soil biodiversity and soil community composition determine ecosystem multifunctionality", Proceedings of the National Academy of Sciences, 111, 5266-5270, doi: 10.1073/pnas.1320054111.

Wang, H. and Gunsalus, R.P., 2000, "The nrfA and nirB Nitrite Reductase Operons in Escherichia coli are Expressed Differently in Response to Nitrate than to Nitrite", Journal of Bacteriology, 182(20), 5813-5822.

Wang, Q., Garrity, G.M., Tiedje, J.M. and Cole, J.R., 2007, "Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy", Appl Environ Microbiol, 73:5261–5267, doi: 10.1128/AEM.00062-07.

Wang, Q., Venkataramanan, K.P., Huang, H., Papoutsakis, E.T. and Wu, C.H., 2013, "Transcription factors and genetic circuits orchestrating the complex, multilayered response of Clostridium acetobutylicum to butanol and butyrate stress", BMC Systems Biology, 7, 120, http://doi.org/10.1186/1752-0509-7-120.

Wang, S., Deng, K., Zaremba, S., Deng, X., Lin, C., Wang, Q. and Zhang, W., 2009, "Transcriptomic Response of Escherichia coli O157:H7 to Oxidative Stress", Applied and Environmental Microbiology, 75(19), 6110–6123, http://doi.org/10.1128/AEM.00914-09.

Wittinghofer, A. and Vetter, I.R., 2011, "Structure-Function Relationships of the G Domain, a Canonical Switch Motif", Annual Review of Biochemistry, 80:1, 943-971.

Wu, X., Holmfeldt, K., Hubalek, V., Lundin, D., Åström, M., Bertilsson, S. and Dopson, M., 2016, "Microbial metagenomes from three aquifers in the Fennoscandian shield terrestrial deep biosphere reveal metabolic partitioning among populations", ISME Journal, 10 (5), pp. 1192-1203.

Xin, X., Jiang, X, Su, J, Yan, X, Ni, J., Faeflen, S.J., Huang, X. and Wright, A.L., 2016, "Manganese oxide affects nitrification and ammonia oxidizers in subtropical and temperate acid forest soils", CATENA, Volume 137, 2016, Pages 24-30, ISSN 0341-8162, doi:org/10.1016/j.catena.2015.09.004.

Zhang, J., Kobert, K., Flouri, T. and Stamatakis, A., 2014, "PEAR: A fast and accurate Illumina Paired-End reAd merger", Bioinformatics, 30:1–7, doi: 10.1093/bioinformatics/btt593.

5.10 List of Tables

Table 5-1

Bio-Env results for CR-9 and CRG-6 biomes. The abundance data from each biome was variance stabilised using functions from the DESeq2 package (Love et al., 2014). The results from this analysis were used as input to the differential abundance analyses shown in Figure 5-5 and Figure 5-6.

Sampling Locations	Best Model Parameters	Spearman's Rank Correlation
CR-9 biome	Manganese, MEM4, MEM11	0.94
CRG-6 biome	Manganese, DOC, Sulfate	1.00

Table 5-2
Taxa identifications of the differentially abundant 16S rRNA gene displayed in Figure 5-5.

OTU	Kingdom	Phylum	Class	Order	Family	Genus	Direction of Fold Change	
							Manganese	MEM11
OTU_95	Bacteria	Chloroflexi	Anaerolineae	Anaerolineales	Anaerolineaceae	g	↑	--
OTU_78	Bacteria	BD1-5	c	o	f	g	↑	--
OTU_10	Bacteria	Proteobacteria	Deltaproteobacteria	Syntrophobacterales	Syntrophaceae	NA	↓	--
OTU_51	Bacteria	Candidate_division _BRC1	c	o	f	g	↓	--
OTU_91	Bacteria	Nitrospirae	Nitrospira	Nitrospirales	Nitrospiraceae	g	↓	--
OTU_58	Bacteria	Nitrospirae	Nitrospira	Nitrospirales	Nitrospiraceae	g	↓	--
OTU_28	Bacteria	Actinobacteria	Coriobacteriia	Coriobacteriales	Coriobacteriaceae	NA	--	↓
OTU_137	Bacteria	BD1-5	c	o	f	g	↓	↑
OTU_104	Archaea	Euryarchaeota	Thermoplasmata	Thermoplasmatales	Terrestrial_Miscellaneous_Gp (TMEG)	g	↓	↑
OTU_67	Bacteria	Nitrospirae	Nitrospira	Nitrospirales	Nitrospiraceae	g	↓	↑
OTU_88	Bacteria	Firmicutes	Clostridia	Clostridiales	Syntrophomonadaceae	g	↑	↓
OTU_87	Bacteria	Nitrospirae	Nitrospira	Nitrospirales	Nitrospiraceae	g	↑	↓
OTU_11	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	NA	↑	↓

MEM = Moran's eigenvector map, NA = not available, -- no effect with the variable, c = class, o = order, f = family, g = genus.

Table 5-3
Taxa identifications of the differentially expressed 16S rRNA gene displayed in Figure 5-6.

OTU#	Kingdom	Phylum	Class	Order	Family	Genus	Direction of Fold Change	
							DOC	Sulfate
OTU_26	Bacteria	Chlorobi	Ignavibacteria	Ignavibacteriales	SR-FBR-L83	g	--	↓
OTU_45	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	NA	--	↓
OTU_21	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	NA	↓	--
OTU_77	Bacteria	Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae	NA	↑	--
OTU_321	Bacteria	Proteobacteria	Deltaproteobacteria	Desulfuromonadales	NA	NA	↑	--
OTU_135	Bacteria	Proteobacteria	Deltaproteobacteria	Desulfuromonadales	M20-Pitesti	g	↑	--
OTU_268	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Caminicella	↑	--
OTU_14	Bacteria	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae	Maritimimonas	↓	↓
OTU_396	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	NA	↑	↑
OTU_13	Bacteria	Proteobacteria	Betaproteobacteria	NA	NA	NA	↑	↑
OTU_141	Bacteria	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	g	↑	↑

DOC = dissolved organic carbon, NA = not available, -- no effect with the variable, c = class, o = order, f = family, g = genus.

5.11 List of Figures

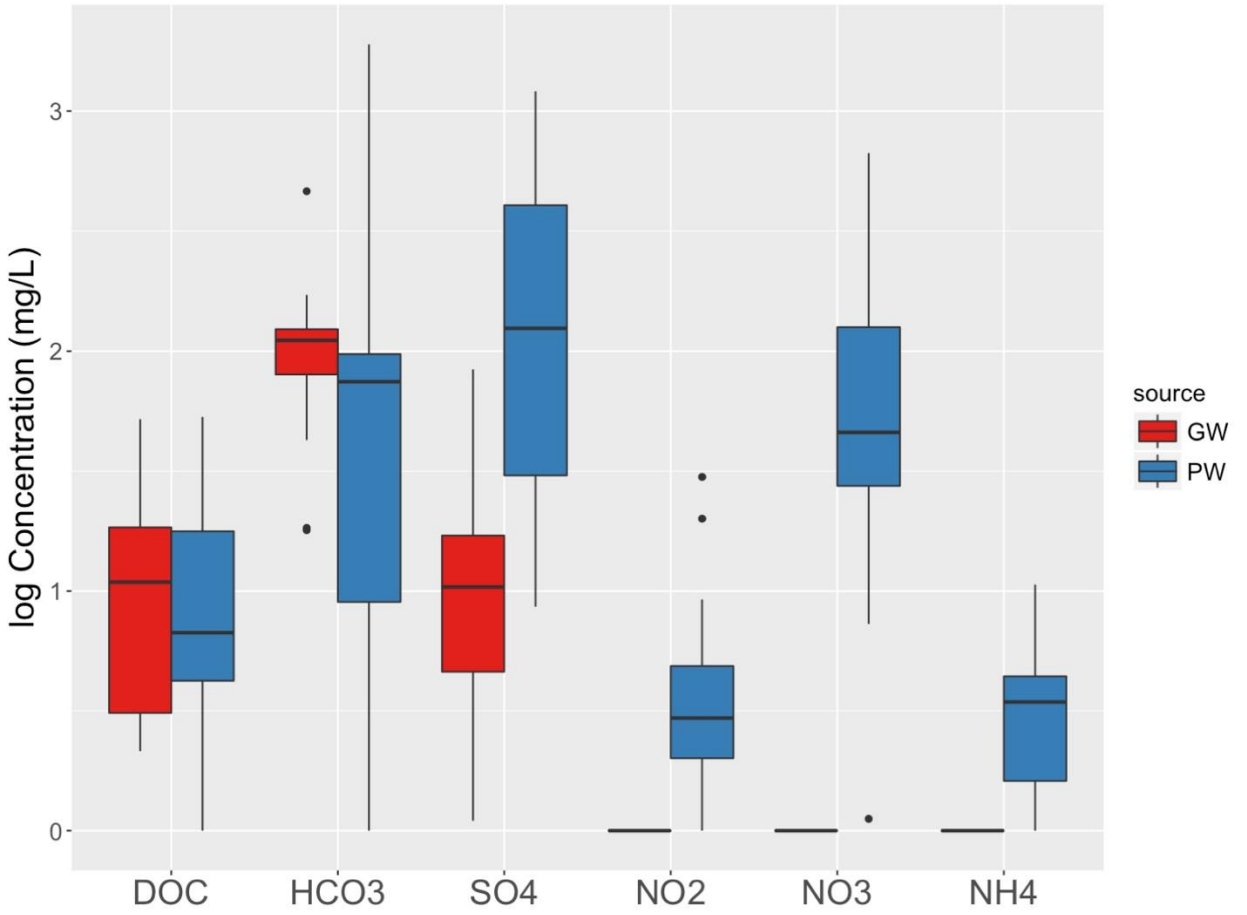


Figure 5-1 Compositional comparison between porewater taken from the drilled rock core samples (blue) and fracture water taken from the drill holes (red) within the crystalline formation at CRL site, Ontario, Canada (King-Sharp et al., 2016; Peterman et al., 2016). The boxplots display the range (minimum and maximum) and median, 25th and 75th percentiles of the compositional data.

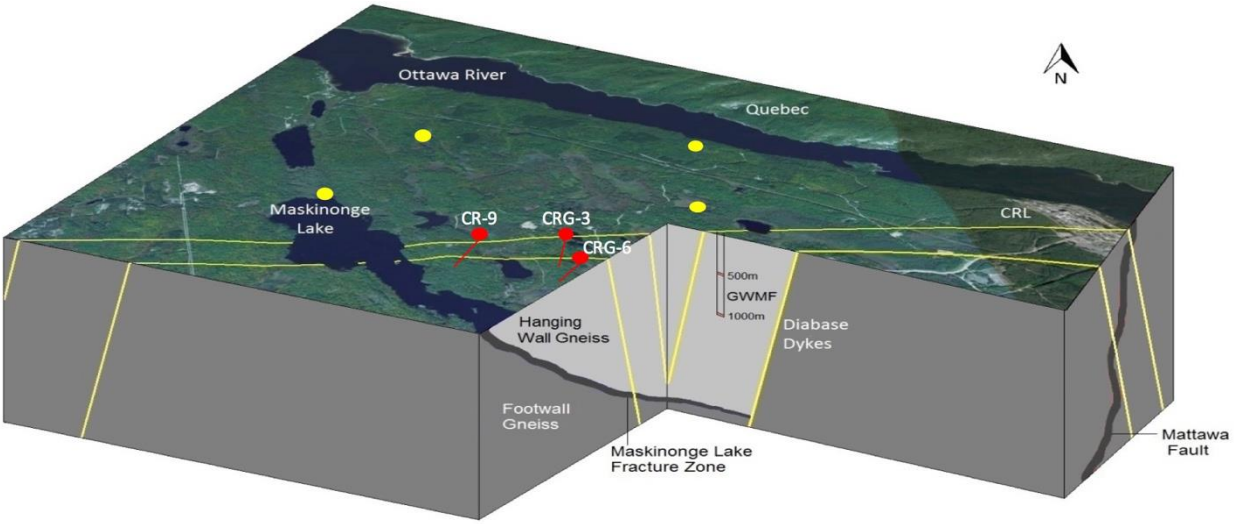


Figure 5-2 Approximate locations of the three sampled boreholes (shown in red; borehole CR-9, CRG-3 and CRG-6) relative to major geologic boundaries -- the Mattawa Fault forming the Ottawa River, the Maskinonge Lake fault and diabase dykes shown as yellow lines traversing the site. The approximate locations of other boreholes are shown as yellow dots. A potential repository location is also shown labeled as GWMF.

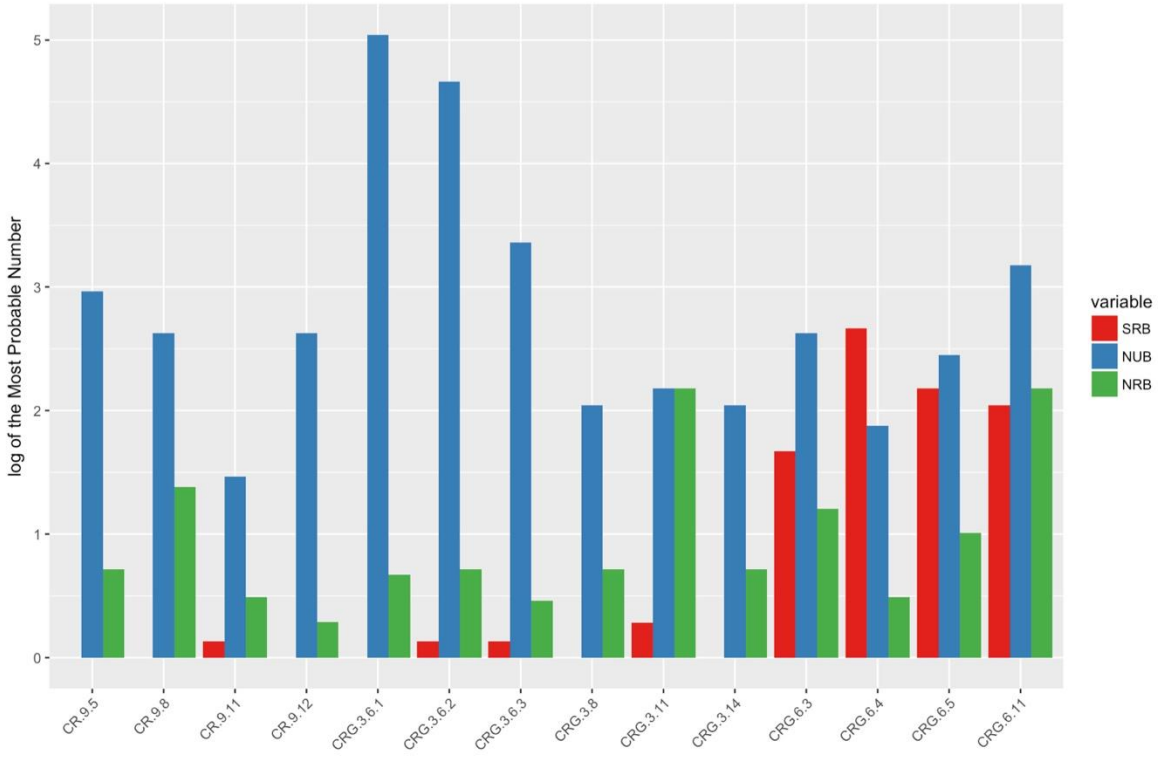


Figure 5-3 Most probable number for nitrogen metabolism as nitrate reduction to nitrite (nitrate utilizing, NUB) and to nitrogen gas (nitrate reducing, NRB) and sulfur metabolism as sulfate reduction (SRB).

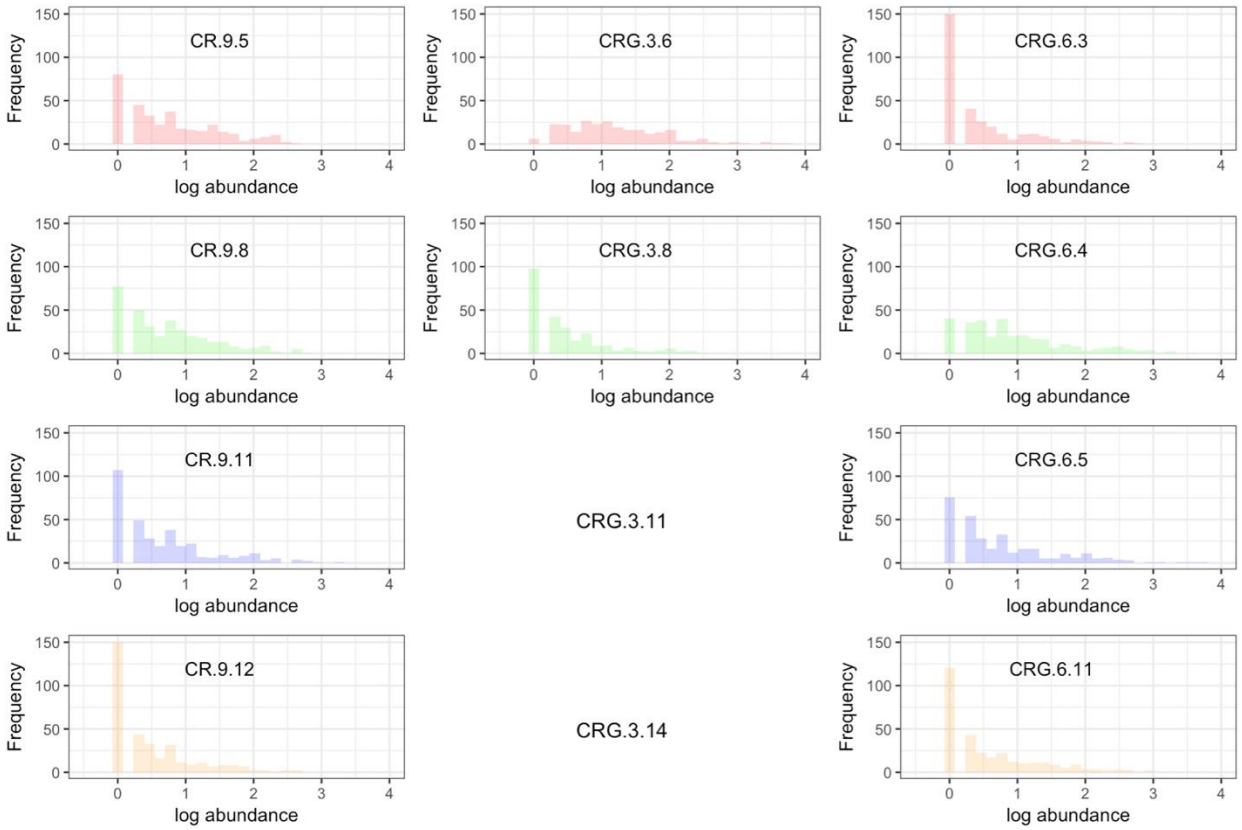


Figure 5-4 Distribution of taxa abundances by sampling location.

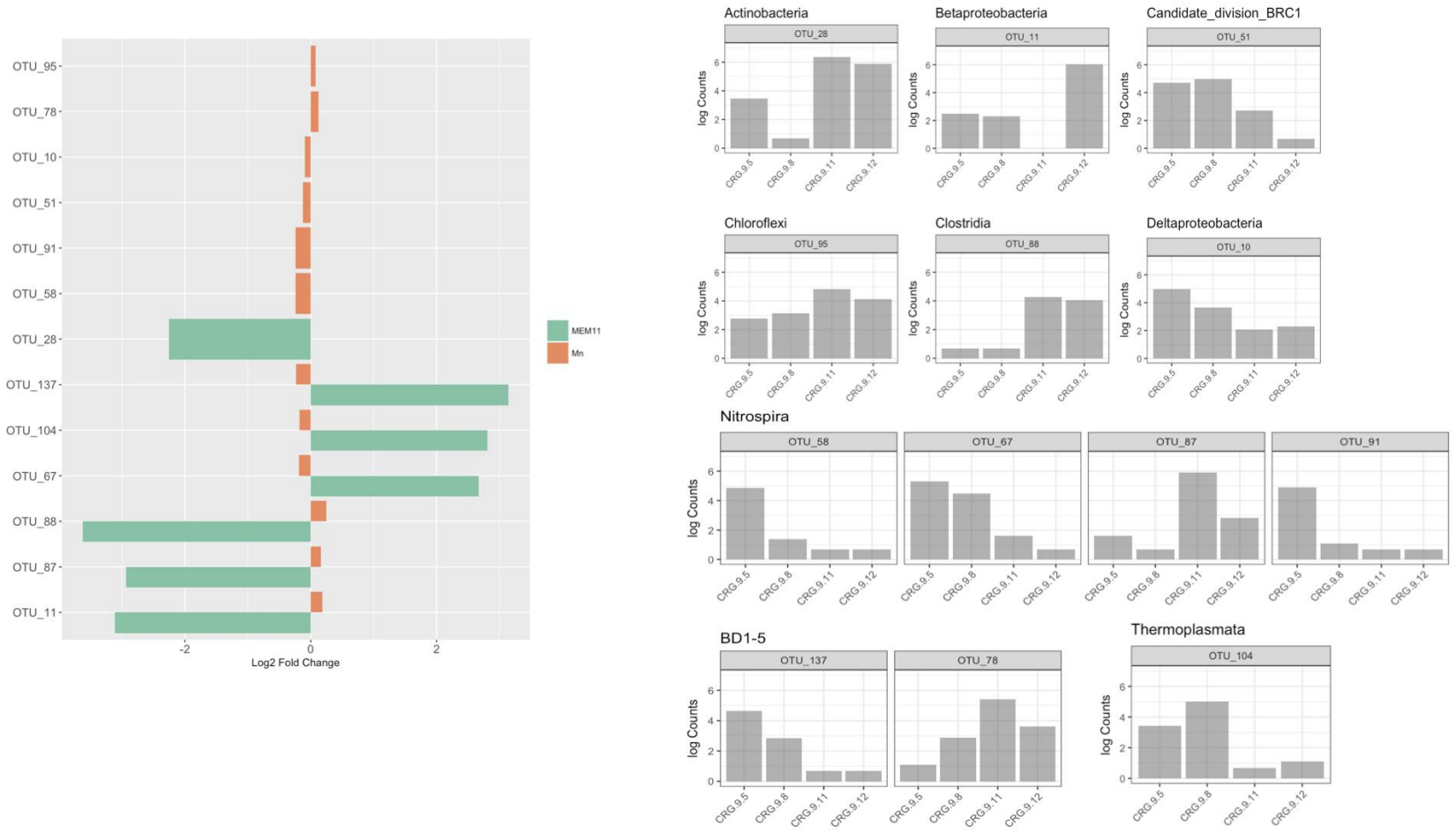


Figure 5-5 Differential abundances within the CR-9 biome. The variables included in this analysis are listed in Table 5-1. The Wald test was used to evaluate the spatial variable, MEM11, or the environmental variable, manganese (Mn). Significance in differential abundances were based on the Benjamini-Hochberg adjusted p-values. The alpha cut-off value was 0.01. As variables, MEM11 and manganese display opposing influences on the differentially abundant taxa. Names of the taxa are listed in Table 5-2.

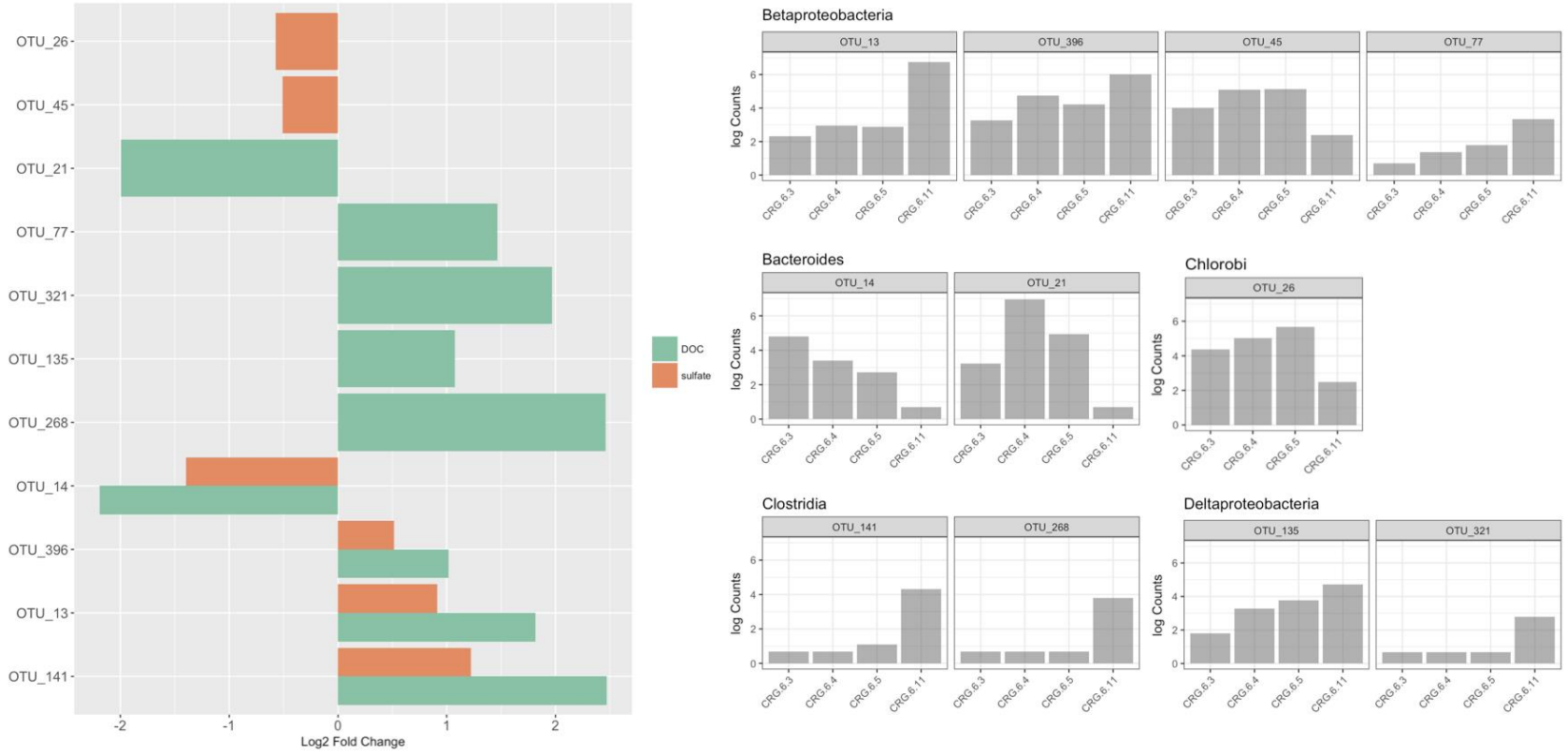


Figure 5-6 Differential abundances within the CRG-6 biome. The variables included in this analysis are listed in Table 5-1. The Wald test was used to evaluate dissolved organic carbon, sulfate and manganese. Significance in differential abundances were based on the Benjamini-Hochberg adjusted p-values. The alpha cut-off value was 0.01. As variables, organic carbon and sulfate displayed complimentary influences on the differentially abundant taxa; manganese was a significant variable. Names of the taxa are listed in Table 5-3.

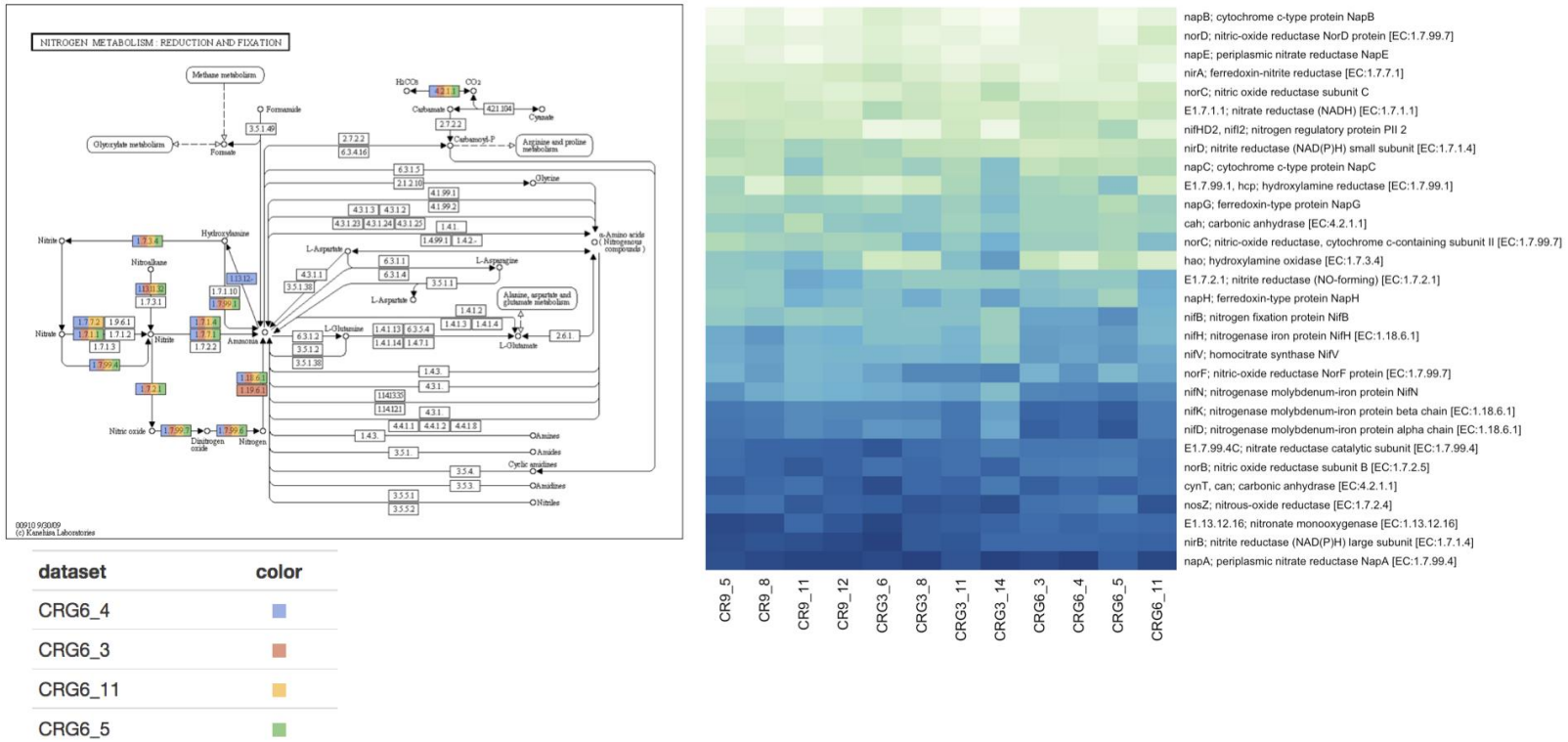


Figure 5-7 KO Annotate genes for nitrogen metabolism. Genes indicated by colors (blue, red, yellow and green) refer to sampling locations within borehole CRG-6.

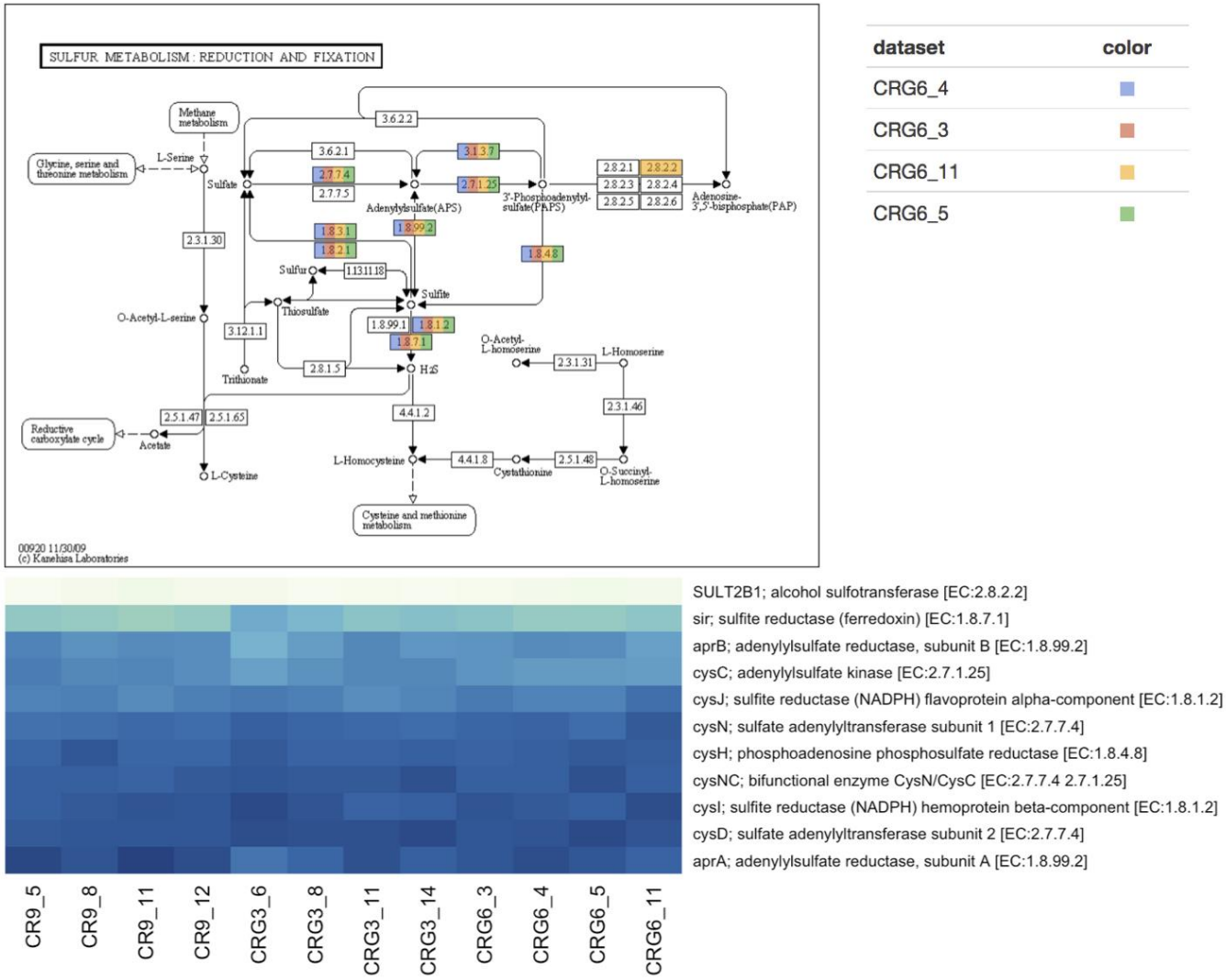


Figure 5-8 KO Annotate genes for sulfur metabolism. Genes indicated by colors (blue, red, yellow and green) refer to sampling locations within borehole CRG-6

5.12 Supplemental Information

TABLE S1. Borehole and interval sampling dates and fracture water chemistry

Borehole	Interval	Sampling Date	pH	HCO3	DOC	SO4	Cl	Fe(II)	Mn(II)
				mg/L#				µg/L	
CRG-3	6	23/9/2013	9.2	90	33	16	12.5	49	3
	8	25/9/2013	9.5	61	2.7	11.3	69	17	3.6
	11	26/9/2013	9.5	74.6	1.4	3	23	89	27
	14	27/9/2013	9.5	60.5	1.2	2.2	23	45	2.6
CRG-6	3	6/5/2013	8.4	110	9.5	3.6	6	47	14
	4	12/5/2013	9.1	91	9.9	5.5	15	68	4
	5	14/5/2013	9.0	74	10.4	4.7	21	430	5
	11	13/5/2013	8.0	116	12.9	10.4	32	178	6
CRG-9	5	5/5/2014	9.4	74.3	1.9	7.3	21.1	13	2.9
	8	6/5/2014	8.7	39.4	1.1	56	206.3	40	12
	11	12/5/2014	8.4	14.2	0.8	87.1	1854	32	40
	12	13/5/2014	8.2	15	1.7	88.3	1836.4	3.9	41

# HCO3 = bicarbonate, DOC = dissolved organic carbon, SO4 = sulfate, Cl = chloride
mdl = minimum detection limit
Ammonia, nitrate and nitrite were not detected in the fracture water

TABLE S2. Borehole and interval descriptions and positive Moran's I spatial coefficients identified by Bio-Env

Borehole	Interval	Elevation, m* collar	interval	Easting**	Northing m	MEM1	MEM2	MEM4	MEM6	MEM11
CRG-3	6	163	-122	313344.3	5102404	0.596	-1.312	1.751	0.517	-1.187
	8		-247	313335.7	5102336	0.753	-1.458	0.623	-0.163	0.159
	11		-417	313318.3	5102241	0.753	-1.458	0.623	-0.163	0.159
	14		-617	313293.2	5102122	0.740	-0.714	-1.584	-0.394	-0.502
CRG-6	3	171	48	313391.2	5102135	1.278	-0.525	-0.614	-1.550	1.227
	4		-4	313373.8	5102116	0.991	1.276	-0.576	0.416	0.344
	5		-68	313351.6	5102092	0.491	1.727	0.927	0.745	-0.480
	11		-390	313234.8	5101980	-0.678	0.982	1.530	-1.296	1.602
CRG-9	5	124.8	-154	312617.4	5102207	-0.625	-0.786	0.273	1.876	1.034
	8		-364	312466.6	5102145	-1.396	-0.278	-1.217	1.095	1.096
	11		-485	312449.8	5102130	-1.462	0.020	-0.208	-0.916	-1.140
	12		-542	312416.1	5102120	-1.462	0.020	-0.208	-0.916	-1.140

* elevation relative to sea level

** Zone 18 of the Universal Transverse Mercator coordinate system

TABLE S3. Total cell counts and most probable number for sulfate and nitrogen metabolism

Borehole	Interval	Total Direct Count (sd)	**SRB** sulfate reducing	**NUB** nitrate reducing to nitrite	**NRB** denitrifying
CRG6	3	4.99E+04 (2.94E+04)	46	420	15
	4	5.49E+04 (3.49E+04)	460	74	2.1
	5	6.88E+04 (2.77E+04)	150	280	9.2
	11	1.69E+05 (6.40E+03)	110	1500	150
CRG3	6 (rep1)	5.75E+05 (2.06E+04)	0	110000	3.7
	6 (rep 2)	1.60E+05 (4.40E+04)	0.36	46000	4.2
	6 (rep 3)	1.38E+05 (4.00E+04)	0.36	2300	1.9
	8	1.55E+05 (6.50E+04)	0	110	4.2
	11	1.13E+05 (6.50E+04)	0.92	150	150
	14	1.39E+05 (5.30E+04)	0	110	4.2
CR9	5	8.10E+04 (3.40E+04)	0	920	4.2
	8	1.30E+05 (7.00E+03)	0	420	23
	11	7.10E+04 (5.00E+04)	0.36	28	2.1
	12	4.70E+04 (2.90E+04)	0	420	0.94

* Standard deviations were calculated from three replicate filters. At least fifteen fields of view and at least 300 cells were counted per filter for a coefficient of variation of 5.8% per filter.

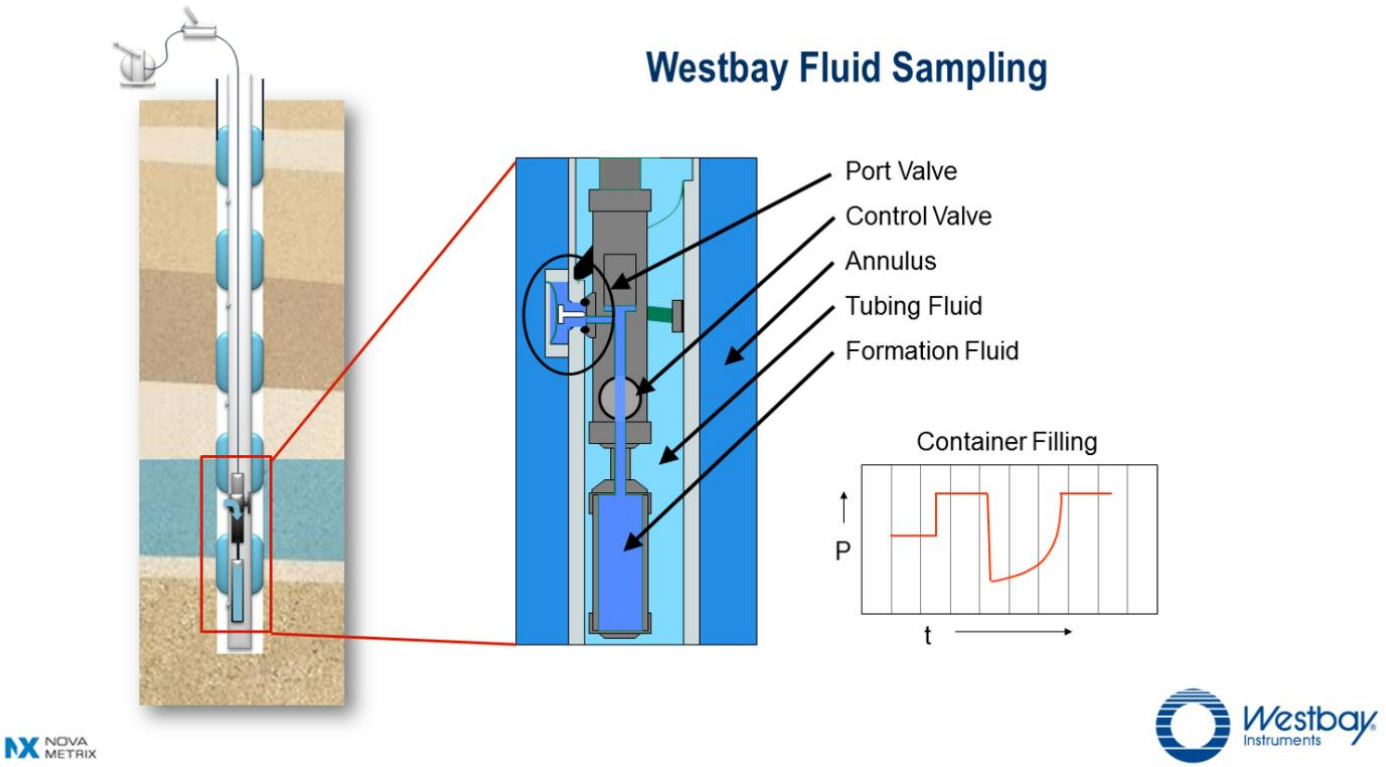
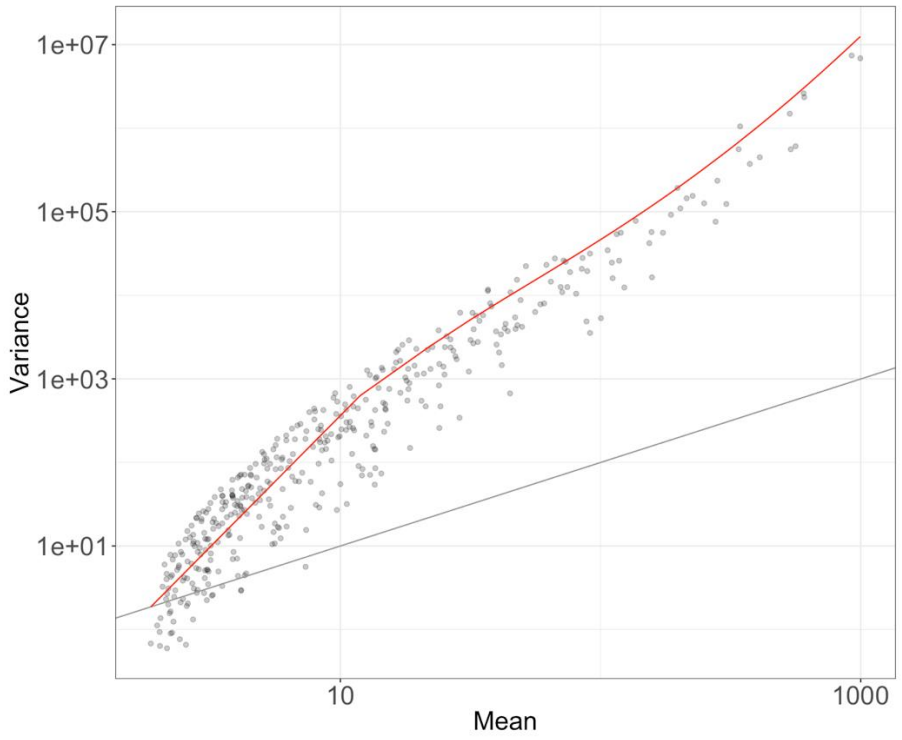


Figure S1. Schematic of the Westbay System

Borehole & Interval	Number of Sequences
CRG.9.5	7807
CRG.9.8	7906
CRG.9.11	10288
CRG.9.12	4190
CRG.3.6	27283
CRG.3.8	2941
CRG.3.11	519
CRG.3.14	248
CRG.6.3	4149
CRG.6.4	35460
CRG.6.5	23062
CRG.6.11	8426
A	



B

Figure S2 A. Total number of sequences. B. The mean-variance relationship of taxa abundances across all sampling locations with total sequence abundances > 600. Each data point represents the mean and variance of a taxon after normalization and variance stabilization (without rarefication) using functions within the package 'DESeq2' (Love at al., 2014). The dark straight line is the mean-variance relationship for a Poisson distribution. The red line is fitted based on the mean-variance relationship of a negative binomial distribution that has an additional variance term that changes with the square of the mean.

TABLE S4. Unique phyla within boreholes CR-9, CRG-3 and CRG-6

	Unique Phyla		
	Within CR-9	**Within CRG-3**	**Within CRG-6**
CR-9	345 of 435 taxa across all sampling locations average cell density: 0.92×10^5 cells/ml (from Table S1)	"Fusobacteria"	"Candidate_division_SR1" "Fusobacteria"
CRG-3	"BD1-5" "WCHB1-60" "Thermotogae" "Candidate_division_WS3" "Thaumarchaeota" "Candidate_division_OD1" "Fibrobacteres" "Candidate_division_BRC1" "Lentisphaerae" "Candidate_division_OP3" "TA06" "Deferribacteres" "GOUTA4"	257 of 435 taxa across all sampling locations average cell density: 4.1×10^5 cells/ml (from Table S1)	"Candidate_division_SR1" "BD1-5" "Thermotogae" "Candidate_division_OP3"
CRG-6	"WCHB1-60" "Candidate_division_WS3" "Thaumarchaeota" "Candidate_division_OD1" "Fibrobacteres" "Candidate_division_BRC1" "Lentisphaerae" "TA06" "Deferribacteres" "GOUTA4"	--	311 of 435 taxa across all sampling locations average cell density: 3.8×10^5 cells/ml (from Table S1)

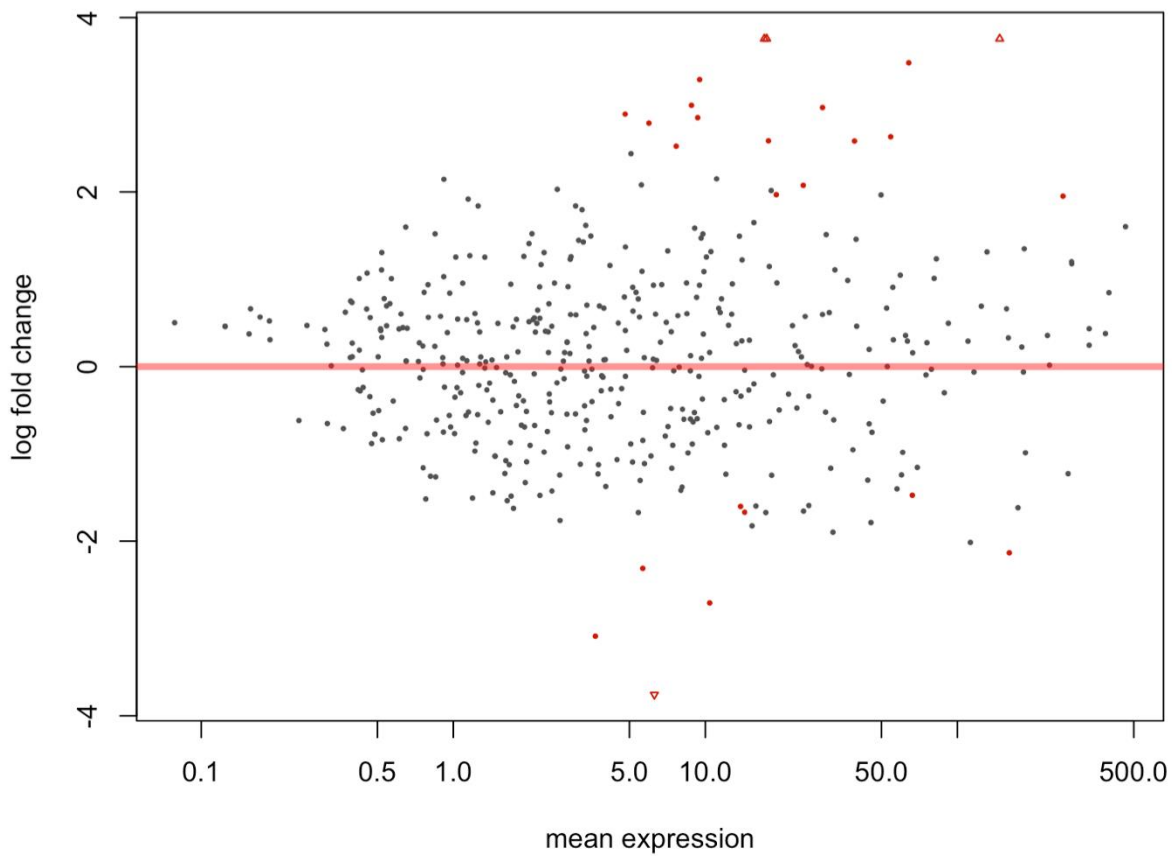


Figure S3 MA plot example of the log2 fold change in abundance (on the y-axis) relative to the mean of normalized counts (on the x-axis). The variable that was tested was the effect of MEM1 across all boreholes, keeping the effect of sampling location fixed. Each point represents one taxon; points in red show the taxa that were significantly differentially abundant, with alpha-value of 0.1 after a Benjamini Hochberg correction.

TABLE S5. Results from Bio-Env analyses -- comparing community compositions with environmental and spatial variables. The abundance data was variance stabilized using functions from the DESeq2 package (Love et al., 2014). Correlations were not assumed a priori. The results from this analysis were used as input to the differential abundance analyses from the DESeq function call-- the results of which are shown in Figure S4.

Sampling Locations	Best model parameters	Correlation
All locations (CRG-9, CRG-3, CRG-6)	Bicarbonate (HCO3), dissolved organic carbon (DOC), MEM1, MEM2, MEM4, MEM6	0.62

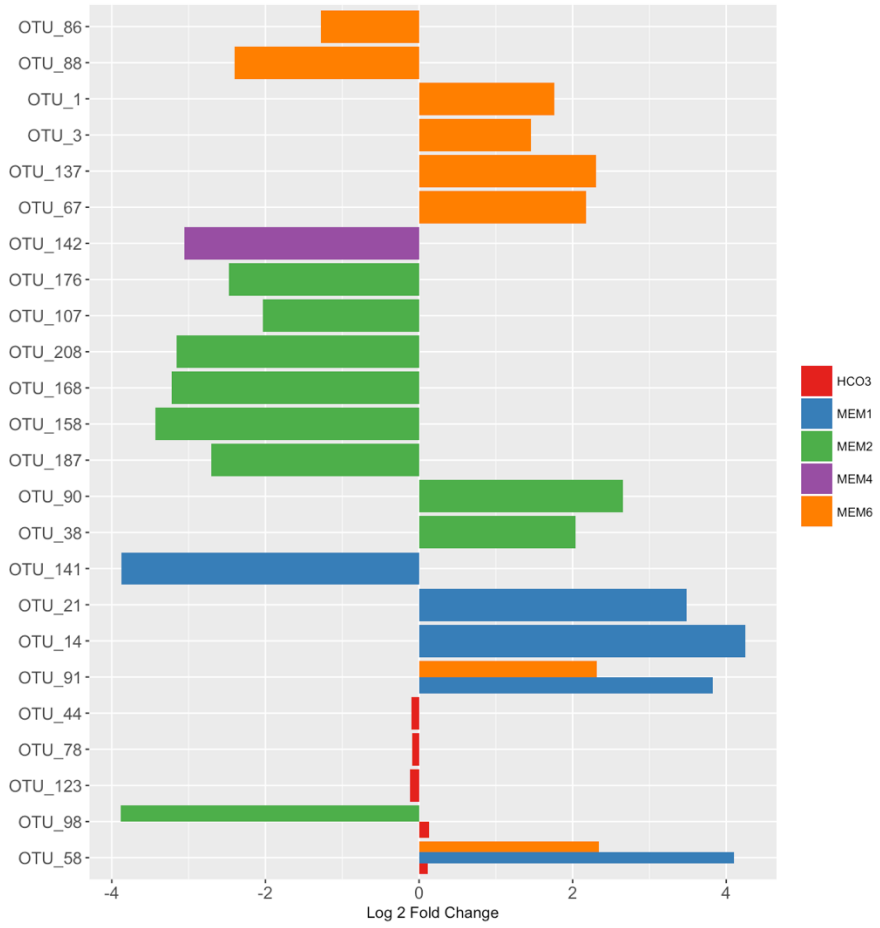


Figure S4. Differential abundances of taxa across all sampling locations. Environmental and spatial variables are listed in Table S5. The Wald test was used to evaluate the spatial variables -- MEM1, MEM2, MEM4 or MEM6 -- and environmental variables -- dissolved organic carbon or bicarbonate -- while controlling for borehole (~bh+variable). Significance in differential abundances were based on the Benjamini-Hochberg adjusted p-values; the alpha cut-off value was 0.01. Only the variables that resulted in significant differential abundance of the 16S rRNA gene are shown. Names of the taxa are listed in Table S6.

TABLE S6. Taxa identifications of the differentially abundant 16S rRNA gene displayed in Figure 8. Each variable associated with significant change in abundance between assemblages from within sampling locations accessed from boreholes CRG-9, CRG-3 and CRG-6 (~bh + variable).

OTU#	Kingdom	Phylum	Class	Order	Family	Genus	Variable Associated with Fold Change
OTU_86	Bacteria	Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae	NA	MEM6
OTU_88	Bacteria	Firmicutes	Clostridia	Clostridiales	Syntrophomonadaceae	g	MEM6
OTU_1	Bacteria	Firmicutes	Clostridia	Clostridiales	Syntrophomonadaceae	Dethiobacter	MEM6
OTU_3	Bacteria	Elusimicrobia	Elusimicrobia	Lineage_IIc	f	g	MEM6
OTU_137	Bacteria	BD1-5	c	o	f	g	MEM6
OTU_67	Bacteria	Nitrospirae	Nitrospira	Nitrospirales	Nitrospiraceae	g	MEM6
OTU_142	Archaea	Euryarchaeota	Halobacteria	Halobacteriales	Deep_Sea_Euryarchaeotic_ Group(DSEG)	g	MEM4
OTU_176	Bacteria	Thermotogae	Thermotogae	Thermotogales	Thermotogaceae	GAL15	MEM2
OTU_107	Bacteria	Chloroflexi	Anaerolineae	Anaerolineales	Anaerolineaceae	g	MEM2
OTU_208	Bacteria	Nitrospirae	Nitrospira	Nitrospirales	Nitrospiraceae	g	MEM2
OTU_168	Archaea	Euryarchaeota	Halobacteria	Halobacteriales	Deep_Sea_Hydrothermal _Vent_Gp_6(DHVEG-6)	Candidatus_ Parvarchaeum	MEM2
OTU_158	Bacteria	Verrucomicrobia	Opitutae	Opitutales	Opitutaceae	Opitutus	MEM2
OTU_187	Bacteria	Firmicutes	Clostridia	NA	NA	NA	MEM2
OTU_90	Bacteria	Firmicutes	Clostridia	Clostridiales	Family_XIII_Incertae_Sedis	g	MEM2
OTU_38	Bacteria	Firmicutes	Clostridia	Clostridiales	Peptococcaceae	Dehalobacterium	MEM2
OTU_141	Bacteria	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	g	MEM1
OTU_21	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	NA	MEM1
OTU_14	Bacteria	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae	Maritimimonas	MEM1
OTU_91	Bacteria	Nitrospirae	Nitrospira	Nitrospirales	Nitrospiraceae	g	MEM1, MEM6
OTU_44	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	NA	HCO3
OTU_78	Bacteria	BD1-5	c	o	f	g	HCO3
OTU_123	Bacteria	Firmicutes	NA	NA	NA	NA	HCO3
OTU_98	Bacteria	Nitrospirae	Nitrospira	Nitrospirales	Nitrospiraceae	g	HCO3, MEM2
OTU_58	Bacteria	Nitrospirae	Nitrospira	Nitrospirales	Nitrospiraceae	g	HCO3, MEM1, MEM6

HCO3 = bicarbonate, MEM = Moran's eigenvector map, NA = not available, -- no effect with the variable, c = class, o = order, f = family, g = genus

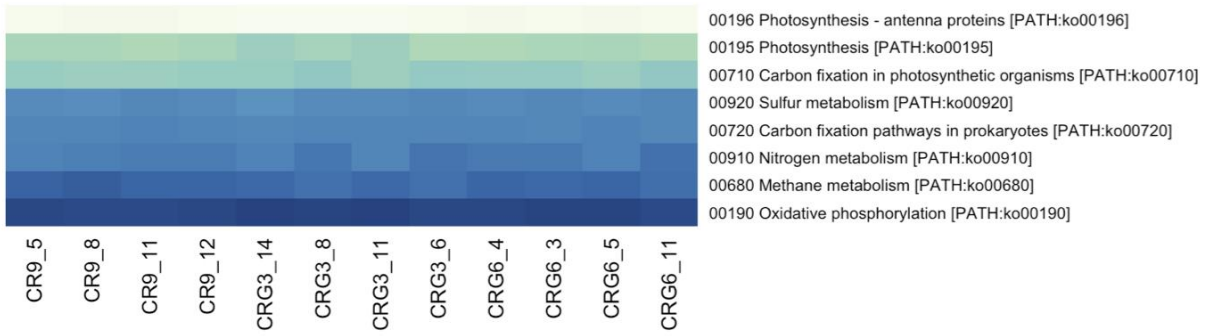


Figure S5: KO Annotate genes for energy metabolism

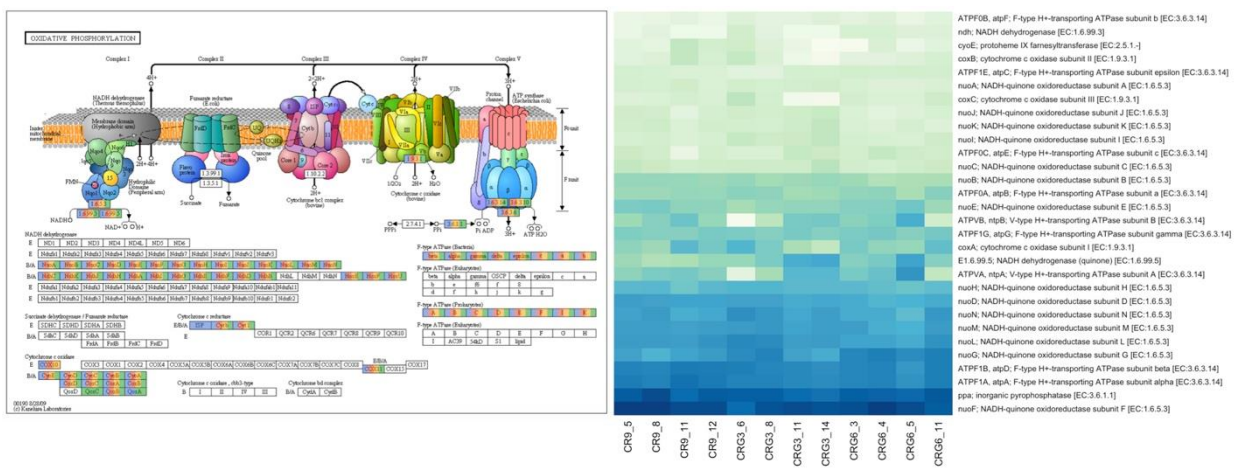


Figure S6 KO Annotate genes for oxidative phosphorylation. Genes indicated by colors (red, green, blue, orange) refer to sampling locations within borehole CRG-6.

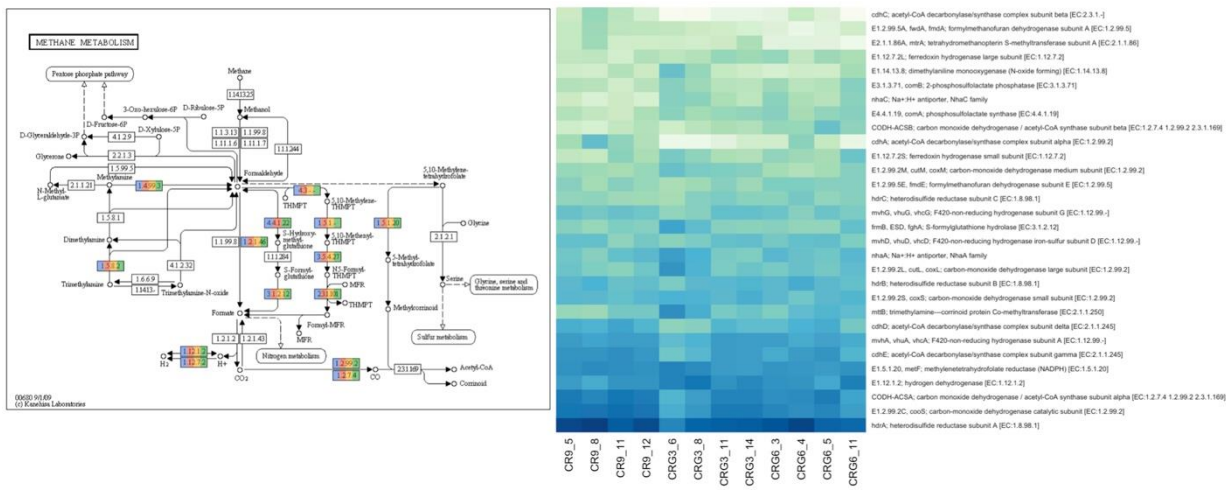


Figure S7 KO Annotate genes for methane metabolism. Genes indicated by colors (red, green, blue, orange) refer to sampling locations within borehole CRG-6.

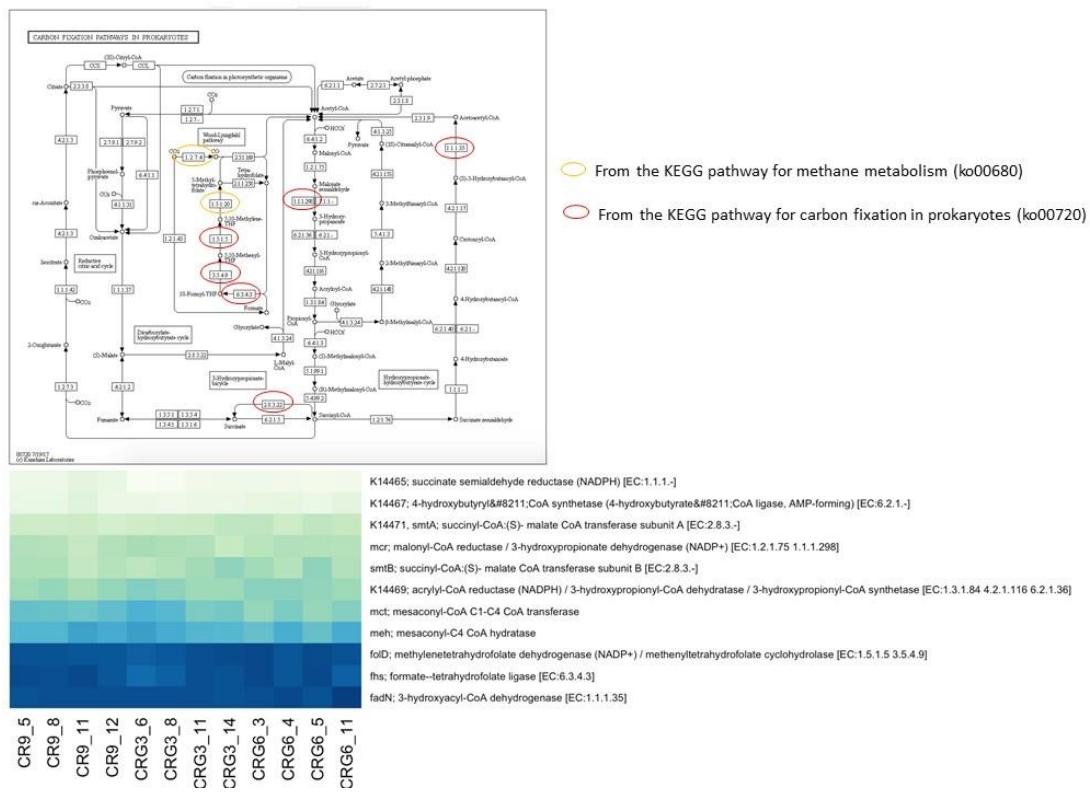


Figure S8 KO Annotate genes for carbon fixation in prokaryotes. Genes indicated by colors (red, green, blue, orange) refer to sampling locations within borehole CRG-6.

6. CONCLUSIONS AND PERSPECTIVES ON FUTURE RESEARCH

6.1 Background

A consequence of the Baas-Becking hypothesis is the long-standing concept in microbial ecology that, by their ubiquity, abiotic and biotic selection are the only drivers of community assembly, and thus environmental variation selects for different microbial species at different locations and along a gradient. Historically, the analysis of microbes in natural environments involved isolation by growth on selective media or, more generally, as counts of colony forming units. With the development of dyes sensitive to a cell's physiology, analysis of microbes also included total and viable counts by microscopy. The differences in the outcomes from these techniques brought to our awareness the so-called "great plate count anomaly" in which 1-5% or fewer of the observed community were cultivable on solid medium. With the advent of the polymerase chain reaction to amplify parts of a microbial genome, and of high throughput sequencing combined with comparisons to known sequences, it is possible to identify the abundant and less abundant microbial taxa without cultivation. These developments also meant that the four high-level ecological processes of Vellend could be used to evaluate community assembly: selection, dispersal, drift and diversification (Vellend, 2010; Vellend, 2016).

6.2 Overall Conclusions

Under an ecological framework of dispersal, drift, diversification and selection, and at the spatial extents between sampling locations, randomness (dispersal, drift and diversification) plays a major role in the subsurface microbial community assembly. Phylogenetic alpha- and beta-diversity metrics (Chapter 3 and Chapter 4) revealed clustering of related taxa and distance decay of similarity; the beta-diversity metric, UniFrac, explained 49.8% of the variance in community composition across all sampling locations. Sampling a smaller spatial extent may increase the resolution for identifying drivers of selection and liely sources of nitrogen and sulfur compounds.

Chapter 2 focused on the distributions of microscope count data, of the measured geochemistry, and of the spatial Moran's Eigenvector map coefficients. A main finding from

this chapter is that the composition of the fracture water displays spatial correlation; inclusion of spatial coefficients improved the residuals distribution in a generalized linear model for cell density. The spatial extent of the sampling locations, therefore, may explain the role of randomness in the datasets. A second main finding from Chapter 2 is that the distributions of the total and viable cell densities align with the distributions of the sulfate and manganese contents of the fracture water, respectively. The co-variance of counts for cell viability and fracture water manganese could mean they share a common cause, such as weathering, or it could mean that manganese is involved in some way in microbial processes related to a cell's ability to reduce tetrazolium salts.

Selection associated with differential abundance of the 16S rRNA gene V4 region was not a major driver of community assembly (Chapter 5); differential abundance of the 16S rRNA gene accounted for ~4% each of the total abundances correlated with sulfate and organic carbon or with manganese and a spatial coefficient, respectively. It may, however, be possible to identify additional environmental factors at spatial scales less than 1 km^3.

Decay of similarity with distance may indicate connections between subsurface sampling locations up to 500 m apart (Chapter 4, covering the smaller spatial extent), and 1.5 km apart (Chapter 3, covering the larger spatial extent). The influences of limited dispersal and drift would be strongest between unconnected sampling locations. Drift, therefore, may account for the unique 16S rRNA phylogenetic diversity represented within each assemblage. After colonization within the subsurface, a loss of diversity would occur by local random extinction of low abundant taxa. The observed phylogenetic clustering and high proportion of sister taxa could reflect the potential influence of diversification (Chapter 4). However, since the 16S rRNA gene copy number varies among the different taxa, the pattern of co-existence of related taxa and the number of sister pairs may reflect, instead, intra-genomic heterogeneity of the V1-V2 and V4 regions of this gene that are greater than the 97% sequence similarity threshold applied when binning sequencing reads into organizational taxonomical units.

Despite these patterns, there was no distinction across sampling locations in the composition and abundances of functional genes for energy metabolism (Chapter 5). Although observed nitrate reduction was common across all sampling locations, the suite of genes within the

metagenome for nitrogen metabolism suggest a potential for complete nitrogen cycling, including nitrous oxide reduction and nitrogen fixation.

Observe metabolism of sulfate reduction was detected within some sampling locations (Chapter 5) and genes for sulfur metabolism were present within all the metagenomes. Based on the metagenomes, there is a potential for carbon metabolism via the Woods-Ljundahl pathway.

Overall, applying the ecological framework of four main drivers of community assembly show that, at the spatial scale of the sampling, up to 50% of the variance among community dynamics reflect randomness. Approximately ~1% of the total abundance was linked to measured metabolism; at smaller spatial scales, ~8% of the total 16S rRNA gene abundance was linked to differential abundances across—potentially connected--sampling locations. At smaller spatial scales than explored, therefore, it may be possible to discern additional metabolic and selective processes. These data will inform models for the performance and long-term safety of geological repositories.

6.3 Future Research

6.3.1 Metatranscriptome Analaysis

Phylogenetic relationships across sampling locations showed that the component taxa were more closely related than expected by chance (Chapter 3 and Chapter 4). This pattern suggests that, at the spatial scale of the analysis, competitive exclusion was not a driver of subsurface community assembly. However, common traits, as suggested by a uniform distribution of genes for energy metabolism (Chapter 5), could mean that traits enabling co-existence within this subsurface are not phylogenetically conserved. A meta-transcriptome assessment would help differentiate between the genes present in this subsurface from the genes expressed. A challenge when sampling fracture water by the Westbay™ system will be preservation the RNA at the time of sampling.

6.3.2 Weathering as a Driver Community Assembly

Demonstrated metabolism included nitrate reduction (common across all sampling locations and at each sampling campaign) and sulfate reduction (observed during a one sampling

campaign). The distributions of total and viable cell counts corresponded with the distributions of sulfate and manganese, respectively (Chapter 2). The fracture water contined both sulfate and manganese but nitrate; a limited analysis of rock porewater identified sulfate and nitrogen compounds (ammonia, nitrite and nitrate) in parts per million concentrations, suggesting that the rock could be a source of these compounds. These observations warrant an assessment of that rock weathering for its influence on the subsurface community.

The compositional differences in nitrate and ammonia content between fracture water and rock porewater (described in Chapter 5) raises an interesting question: what is the source of these nitrogen compounds? Is it nitrogen from an ancient atmosphere, with the driver of microbial assembly being remnant ammonia that is retained within the rock minerals, or is it nitrogen from a modern atmosphere with the driver of microbial assembly being nitrogen fixation? Swanner and Templeton (2011) hypothesized that a source of ammonia in rock-reacted fluids within a granite rock formation in Colorado was from the weathering of ammonia-bearing biotite; spectroscopic analysis of the biotite minerals did not support this hypothesis; however, they could amplify the nifH gene that codes for a subunit of the nitrogenase enzyme. The presence of this gene within this formation would indicate that the source of the ammonia in rock-reacted fluids were from the fixation of nitrogen, rather than from ammonia within the rock. These findings are compelling because they suggest that the microbes are adapted to the low nutrient and low energy crystalline formation by their capability of fixing nitrogen, and that fixed nitrogen may support subsurface biomass via assimilation.

Testing the rock minerals, biotite and garnet, from the CRG-borehole cores for nitrogen content would support rock weathering as a common driver of microbial assembly.

6.4 References

Hallin, S., Philippot, L., Löffler, F. E., Sanford, R. A., Jones, C. M., 2018, “Genomics and ecology of novel N2O-reducing microorganisms”, Trends in Microbiology, 26, 43 – 55, doi: 10.1016/j.tim.2017.07.003.

Swanner, E. D., and Templeton, A. S. (2011). Potential for nitrogen fixation and nitrification in the granite-hosted subsurface at Henderson mine, CO. Front. Microbiol. 2, 254. doi: 10.3389/fmicb.2011.00254.

Vellend, M. (2010). Conceptual synthesis in community ecology, Quarterly Review of Biology, 85, 183–206. doi: 10.1086/652373.

Vellend, M., 2016, “The Theory of Ecological Communities”, Monograph in Population Biology, 57, Princeton University Press.