Metagenomic investigation of the microbial diversity in a chrysotile asbestos mine pond, Lowell, Vermont, USA

Heather E. Driscoll a,e, James J. Vincent b, Erika L. English c,1, Elizabeth D. Dolcic c

a Vermont Genetics Network, Department of Biology, Norwich University, 158 Harmon Drive, Northfield, VT 05663, USA
b Vermont Genetics Network, Department of Biology, University of Vermont, 337 Marsh Life Science, Burlington, VT 05405, USA
c Department of Environmental and Health Sciences, Johnson State College, 337 College Road, Johnson, VT 05656, USA

Keywords: microbial diversity, metagenomic analysis, alkaline, asbestos, serpentine, Illumina, whole metagenome shotgun sequencing

Here we report on a metagenomics investigation of the microbial diversity in a serpentine-hosted aquatic habitat created by chrysotile asbestos mining activity at the Vermont Asbestos Group (VAG) Mine in northern Vermont, USA. The now-abandoned VAG Mine on Belvidere Mountain in the towns of Eden and Lowell includes three open-pit quarries, a flooded pit, mill buildings, roads, and ~26 million metric tons of eroding mine waste that contribute alkaline mine drainage to the surrounding watershed. Metagenomes and water chemistry originated from aquatic samples taken at three depths (0.5 m, 3.5 m, and 25 m) along the water column at three distinct, offshore sites within the mine’s flooded pit (near 44°46′00.7673″; −72°31′36.2699″; UTM NAD 83 Zone 18 T 0655720 E, 4960030 N). Whole metagenome shotgun Illumina paired-end sequences were quality trimmed and analyzed based on a translated nucleotide search of NCBI-NR protein database and lowest common ancestor taxonomic assignments. Our results show strata within the pit pond water column can be distinguished by taxonomic composition and distribution, pH, temperature, conductivity, light intensity, and concentrations of dissolved oxygen. At the phylum level, metagenomes from 0.5 m and 3.5 m contained a similar distribution of taxa and were dominated by Actinobacteria (46% and 53% of reads, respectively), Proteobacteria (45% and 38%, respectively), and Bacteroidetes (7% in both). The metagenomes from 25 m showed a greater diversity of phyla and a different distribution of reads than the two upper strata: Proteobacteria (60%), Actinobacteria (18%), Planctomycetes, (10%), Bacteroidetes (5%) and Cyanobacteria (2.5%). Armatimonadetes (<1%), Verrucomicrobia (<1%), Firmicutes (<1%), and Nitrospirae (<1%). Raw metagenome sequence data from each sample reside in NCBI’s Short Read Archive (SRA ID: SRP056095) and are accessible through NCBI BioProject PRJNA277916.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1 Direct link to deposited data

NCBI BioProject PRJNA277916 includes raw data and sample metadata, including water chemistry: http://www.ncbi.nlm.nih.gov/bioproject/PRJNA277916.

2 Experimental Design, Materials and Methods

2.1 Study site description

The Vermont Asbestos Group (VAG) Mine, situated on a chrysotile asbestos-bearing ultramafic rock outcrop located within the Lamoille and Missisquoi River watersheds in northern Vermont, USA, is an abandoned asbestos mine and mill that ceased operation in 1993. High levels of chrysotile asbestos fibers and elevated concentrations of magnesium, nickel, chromium, and arsenic have been recorded at several aquatic and terrestrial sites within the VAG Mine area [1]. The site of sample collection within the VAG Mine area was the Lowell mine pit pond, a flooded open pit quarry adjacent to massive tailings piles (Fig. 1). The
Fig. 1. Vermont Asbestos Group Mine area: A) Aerial map with sample collection sites and Burgess Branch stream colored blue (map source: http://anrmaps.vermont.gov/websites/anra5/). B) Lowell tailings piles adjacent to the pit pond and sample collection site (photo source: Steve Schlipf, steveschlipfphotography.com).
mineralogical character of the mine waste was identified previously (Mg, Si, Fe, and Al, major constituents; Ni and Cr, minor constituents) [2]. The water column of the VAG mine’s pit pond was sampled at three sites (site (1) UTM NAD 83 Zone 18 T 0695810 E 4960010 N (44°46’00.0311″, -72°31’32.2070″); site (2) UTM NAD 83 Zone 18 T 0695720 E 4960030 N (44°46’00.7673″, -72°31’36.2699″); site (3) UTM NAD 83 Zone 18 T 0695650 E 4960080 N (44°46’02.4551″, -72°31’39.3823″)) on August 6 and 7, 2012. Elevation of the three sites was recorded at 369, 367, and 365 m, respectively.

2.2. Sample collection and processing

A portable pumping system was designed for a 16 ft Jon boat allowing us to collect 15–20 L of water aseptically from the epilimnion (0.5 m), metalimnion (3.5 m) and hypolimnion (25 m) of the pond at sites 1–3. The following notation was used for naming samples: epilimnion (1-S, 2-S, 3-S), metalimnion (1-M, 2-M, 3-M), and hypolimnion (1-B, 2-B, 3-B). Water samples were collected using two methods. On August 6, 2012, 15 L of water were pumped from each depth at site 1 and directly filtered through 0.45 μm polycarbonate filters (MilliporeSigma, Billerica, MA, USA). Filters were stored on ice and upon return to the laboratory maintained at -80°C until processing for DNA extraction. On August 7, 2012, sites 2 and 3 within the pond pit were sampled. Water was pumped from each depth and collected aseptically in Whirl-pak bags (Nasco, Fort Atkinson, WI, USA) for the metalimnion and hypolimnion samples and sterile 5 gal polycarbonate containers for the epilimnion samples. Whirl-pak bags were stored on ice until return to the laboratory, and the sealed, polycarbonate containers remained submerged until returning to shore. These samples stored in Whirl-pak bags were stored at 4°C until return to the laboratory. Within 12–24 h of collection, microorganisms from 10 to 15 L of water per site (2-S, 2-M, 2-B, 3-S, 3-M and 3-B) were concentrated onto 0.45 μm polycarbonate filters and stored at -80°C for subsequent DNA extraction and Illumina library preparation at the Advanced Genome Technologies Core (AGTC) at the University of Vermont (UVM).

Physical and chemical parameters (temperature, pH, dissolved oxygen, specific conductivity, and light intensity) of the water column were measured using a Hydrolab DS5 multiprobe surveyor (Hach Hydromet, Loveland, Colorado, USA) and LI-COR LI-250A light meter with LI-192 connected. Water samples were filtered through 0.45 μm polycarbonate filters (MilliporeSigma, Billerica, MA, USA) and assessed using the Nanodrop spectrophotometer (Thermo Scientific, Madison, WI, USA), and the resulting DNA was quantified by qPCR quantification using the Qubit fluorometer and qPCR quantification (kit #4824, KAPA Biosciences, Woburn, MA, USA). Library quality and size distribution was assessed using the Agilent Bioanalyzer 2100.

2.3. DNA extraction

Total genomic DNA was extracted from all nine samples using standard techniques at the Advanced Genomic Technologies Core at UVM. Briefly, 5 mL of sterile molecular grade phosphate-buffered saline (PBS) was added to the membrane filters in 50 mL conical tubes and shredded using a FastPrep-24 (MP Biomedical, Inc., Santa Ana, CA, USA) and a scalpel. The PBS was transferred to a new tube, centrifuged, and volume adjusted to ~0.5 mL. Homogenization was performed using a bead beater approach with the FastPrep-24 system and a 3 mm ceramic ball and Al2O3 abrasive added to the samples. A cocktail of enzymes (10 μL of 10 μg/μL lysozyme, 4 μL of 400 μg/μL achromopeptidase and 2 μL of 1 μg/μL mutanolysin) was added to the sample and allowed to incubate overnight at 37 °C. DNA was then extracted using the standard method outlined in the E.Z.N.A.® Mollusc DNA isolation kit (OmegaBiotek, Inc., Norcross, GA, USA), and the resulting DNA was quantified and assessed using the Nanodrop spectrophotometer (Thermo Scientific, Madison WI, USA), and Qubit Spectrofluorometer (Life Technologies, Carlsbad, CA, USA). Fragmentation of 10–100 ng of the resulting DNA was performed using a Covaris S2 AFA sonicator (Covaris Corp., Woburn, MA, USA) equipped with MicroVails to yield a size range of 200–500 base pairs as confirmed through a high sensitivity DNA chip on the Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). No genomic DNA was recovered from sample 1-B, and it was not included in further preparations.

2.4. Illumina library preparation

Library preparation was performed using 10–100 ng of DNA per sample in accordance with the Illumina TruSeq (Illumina Inc., San Diego, CA, USA) DNA Sample Prep LT version 2 SOP (Part #15026486 Rev. C July 2012) with the indicated reagents (DNA kit #FC-121-2001) at the AGTC at UVM. Each sample was subject to end repair, adenylation, and ligation of Illumina adapters for indexing purposes. PCR amplification was performed using Illumina reagents (Part #15012995) followed by quantification using the Qubit spectrophotometer and qPCR quantification (kit #4824, KAPA Biosciences, Woburn, MA, USA). Library quality and size distribution was assessed using the Agilent Bioanalyzer 2100.

2.5. Whole metagenome shotgun sequencing

Cluster generation and paired-end sequencing were performed using an eight lane high-capacity v3 flow cell on the Illumina cBOT and HiSeq 2000 sequencer, respectively, at the DNA Sequencing & Genotyping Center located in the Delaware Biotechnology Institute, an interdisciplinary research unit at the University of Delaware. Illumina PhiX Control v3 was used as a low-concentration spike-in during sequencing. Eight FASTQ files with raw sequence data were delivered as one uncompressed tar file download.

2.6. Sequence processing

Bioinformatics infrastructure and analysis were provided by the Vermont Genetics Network (VGN) Bioinformatics Core. Raw sequences were examined for quality with FastQC v0.11.2 [3]. Adapters were removed and low-quality base calls and reads were filtered using Trimmomatic v0.33 [4]. A custom adapter file containing TruSeq universal and index primers, as well as the reverse complement of each, was used for removing adapters. Leading and trailing bases below quality 20 were removed. Additionally, reads were scanned using a 5-base wide sliding window and cut when average quality per base dropped below 20. Reads <75 bases in length were also removed (Trimmomatic parameters: PE -phred33 ILLUMINACLIP:TruSeq3_25.fa:2:30:7 LEADING:20 TRAILING:20 SLIDINGWINDOW:5:20 MINLEN:75). Quality-trimmed FASTQ files for each sample were aligned to the PhiX genome (NCBI RefSeq NC_001422.1) using Bowtie2 2.2.3 [5] and all aligned reads were removed. Quality-trimmed and filtered reads were verified with FastQC prior to characterizing the taxonomic composition of the mine microbial community.

2.7. Taxonomic classification

Translated trimmed paired-end reads treated as single-end reads served as input for a protein-level homology search against NCBI-NR, a comprehensive non-redundant protein database (downloaded March 25, 2015), using the BLAST-like tool RAPSearch2 v2.16 [6]. All reads with alignments to the NR protein database, maximum 50 alignments per read, were imported into META Genome AAnalyzer (MEGAN) v5.7.10 [7] and parsed using a lowest common ancestor (LCA) algorithm. The following MEGAN parameters were used: maxMatches = 100 minScore = 50.0 maxExpected = 1.0 topPercent = 10.0 minSupportPercent = 0.5 minSupport = 50 minComplexity = 0.44 useMinimalCoverageHeuristic = false useSeed = true useCOG = true useKegg = true paired = false useIdentityFilter = false useTextStoragePolicy = InRMAZ blastFormat = BlastTAB mapping = Taxonomy:Built_IN = true, Taxonomy:Gl_MAP = true,
SEED:GL_MAP = true, KEGG:GL_MAP = true, COG:GL_MAP = true. If sequences were unambiguously assigned to a taxon and passed the filter defined by these parameters, input reads were given a taxonomic assignment by LCA based on GI accession numbers and the complete NCBI taxonomy (ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/taxdump.zip downloaded Mar. 24, 2015). A single MEGAN combined sample file was generated from all individual sample MEGAN files with read counts normalized to the sample with the fewest input reads.

Two positive controls were used to validate our bioinformatics pipeline and to establish a minimum support threshold for taxonomic profiling. One control data set was comprised of single-end Illumina reads from a synthetic microbial sample prepared by CosmosID, which simulates organisms found in the Delaware River. A description of this constructed freshwater sample can be found here: http://www.cosmosid.net/constructed-freshwater.php. The second control dataset was a set of single-end Illumina reads from the Human Microbiome Project (HMP) mock community even sample (SRA accession SRR172902). Reads from both positive control samples can be downloaded from BaseSpace with a free account: https://basespace.illumina.com/projects/20039022/samples.

2.8. Statistical analysis

Statistical Analysis of Metagenomic Profiles (STAMP) software v2.1.340 [8] was used to test statistical significance of differentially abundant taxonomic groups and functional categories for each pit pond depth. LCA taxonomic profiles at all ranks including abundances, were imported to STAMP and a one-way ANOVA was used to compare strata with an effect size (ETA-squared) and multiple test correction using the Benjamini-Hochberg FDR method. Tukey-Kramer post-hoc test (0,95) was used to determine which means were significantly different when an ANOVA produced a significant p-value.

3. Results

3.1. Water chemistry

Metagenomes were generated from samples taken at three depths along the water column at three distinct sites within the mine’s pit pond. Environmental parameters of the epilimnion and metalimnion were notably different from the parameters of the hypolimnion. Marked differences were observed in pH, temperature, conductivity, light intensity, and concentrations of D.O. (Fig. 2, Supplementary Table ST1). The pit pond is alkaline, and in contrast to most freshwater lakes, pH increased with depth steadily rising from an average of 8.76 at the surface to 9.18 in the hypolimnion. Conductivity was high throughout the water table and increased with depth at all three sites. Average conductivity at the surface measured 271.3 μS increasing to 351.6 μS at 25 m indicating significantly higher ionic concentrations than expected in Vermont waters at this elevation (Vermont Department of Conservation, personal communication). The hypolimnion also presented greater extremes in temperature and light. Water temperature was nearly fourfold higher at the surface (23.53 °C) than in the hypolimnion (6.03 °C), and light was absent below 10 m. The steep thermocline revealed a narrow metalimnion. Oxygen levels remained near saturation throughout the water column and supersaturated in the thermocline. Heavy metal analysis of the water column revealed significant concentrations of Ni (26 μg/L), Fe (563 μg/L), and Mn (950 μg/L) in sample 1-B only. Levels of Ca, Mg, K, Na, Al, Fe, Mn, Cu, Zn and As were below 5 μg/L in all water samples (not shown).

3.2. Bacterial community structure

WMS sequencing of eight samples generated over 132 million paired-end reads, 101 bp in length, with an average depth of 16.5 million reads per sample (6–33 million) (Table 1). Seventy-one percent of raw reads (93,988,834) were retained after quality-trimming and aligned to the NCBI-NR protein database. Of the 48,646,975 reads with at least one hit to NR proteins, approximately 73% were assigned taxonomy by the LCA algorithm in MEGAN. Almost half of quality-trimmed reads in our samples (48.2%) had no protein hits in NCBI-NR and 27.2% of reads with protein hits could not be classified by the LCA algorithm. As a result, these reads were designated “Not Assigned” in MEGAN. There were 55,787 reads with repetitive sequence and were assigned to a “Low Complexity” node. The minimum-support percent threshold in MEGAN was set to 0.5% based on our bioinformatics workflow results from the HMP and Delaware River positive controls (Supplementary Figures SF1-SF2).

The epilimnion and metalimnion exhibited a similar distribution of taxa at the phylum level (Fig. 3; Supplementary Table ST2), and were dominated by Actinobacteria (46% and 53% of reads, respectively), Proteobacteria (45% and 36%, respectively), and Bacteroidetes (7% in both). The phyla Cyanobacteria and Planctomycetes contained <1% of the reads and each were present at only one site. The hypolimnion exhibited greater diversity of phyla and the distribution of reads differed from the surface and metalimnion. Within the two samples of the hypolimnion, nine distinct phyla were identified. These phyla largely paralleled the epilimnion and metalimnion, but a distinct distribution was evident. Proteobacteria was the predominant phylum containing 60% of the reads with markedly lower proportion of Actinobacteria (18%), Planctomycetes, (10%), Bacteroidetes (5%) and Cyanobacteria (2.5%). Armamonadetes, Verrucomicrobia, Firmicutes, Nitrospirae, and unnamed phyla contributed <1% each of the total reads, and with
the exception of Verrucomicrobia, were exclusive to one sample. Read
counts for phyla and other taxonomic ranks down to species are avail-
able for all samples in Supplementary Material (Supplementary Tables
ST2-ST7).

Based on the LCA of each lineage in the NCBI taxonomy, we observed
a set of taxa that were shared across the water column and others that
were unique to each stratum (Figs. 4-5; Supplementary Table ST8).

The more highly alkaline, psychrophilic hypolimnion exhibited the
greatest diversity with 15 LCA taxa representing nine phyla, including
Verrucomicrobia, Firmicutes, and Nitrospirae. Taxa unique to the epi-
limnion and metalimnion belong to Proteobacteria and Actinobacteria.
No taxon was shared between only the epilimnion and hypolimnion,
but six occurred in all three strata, including three unclassi-

fied Actinobacteria. The surface and metalimnion shared 11 taxa, including
the planktonic genus Limnohabitans, and the chlorophyll a containing
genus Sandarakinorhabdus.

### 3.3. Statistical results

Results from STAMP showed significant differences with strong ef-
fect sizes at all taxonomic ranks (Supplementary Tables ST9-ST14). At
the rank of phylum, the relative abundance of reads from Planctomycetes, Verrucomicrobia, and Cyanobacteria were significantly
greater in hypolimnion samples than in epilimnion and metalimnion samples (with FDR corrected \( p \)-values of \( 7.98 \times 10^{-6}, 0.0002 \), and
0.0010, respectively). The relative abundance of reads from Actinobacteria were significantly lower in hypolimnion samples than
in epilimnion and metalimnion samples (FDR corrected \( p \)-value of
0.007).

### 4. Conclusions

The distribution of microbial diversity is reflective of the VAG
Mine pit pond's physical and chemical stratification. The epilimnion
and metalimnion share a greater number of taxa (11 LCA taxa) than
do the metalimnion and hypolimnion (two LCA taxa). The physio-
chemical properties of the upper two limnetic layers more closely
paralleled each other and were in contrast to the hypolimnion,
where inhabitants are required to metabolize at a higher pH, lower tem-
peratures, and in the absence of sunlight. The genera Limnohabitans, Illumatobacter, Spingomonas, and Sandarakinorhabdus, present in
both the epi- and metalimnion, are ubiquitous freshwater plankton.

Sandarakinorhabdus is a novel, bacteriochlorophyll a containing, aero-
bic anoxygenic photoheterotrophic genus and Spingomonas taxa are
capable of producing the yellow carotenoid, nostoxanthin. Neither of
these two genera were identified in the light-deprived hypolimnion.

The contrasting environments of the metalimnion and hypolimnion
are reflected by the paucity of shared taxa. Only two LCA taxa,
Acinetobacter junii and Flavobacteriales were shared between these
strata. No taxon was common to only the epilimnion and hypolimnion,
but six LCA taxa were prevalent throughout the water column, includ-
ing three unclassified Actinobacteria. It is noteworthy that over half of
quality-trimmed reads in our samples (48.2%) had no protein hits.
This is 30% higher than would be expected based on the proportion
of non protein-coding regions in microbial genomes (approximately

![Fig. 3. Percent relative abundance of reads assigned to each stratum at the phylum level by MEGAN (Strata: epilimnion = 0.5 m, metalimnion = 3.5 m, hypolimnion = 25 m).](image-url)
Fig. 4. Taxonomic comparison of eight pit pond samples within three strata (blue: epilimnion, yellow: metalimnion, red: hypolimnion). Each node in the NCBI taxonomy is shown as a bar chart indicating the number of normalized reads from each sample that have been assigned to the node. Terminal taxa are the lowest common ancestor for each lineage shown, i.e. species, genera, or other ranks.
20%), which should not align to proteins in the NCBI-NR database. Additionally, 27.2% of reads with protein hits were not given a taxonomic assignment by LCA and were designated “Not Assigned” in MEGAN. These results suggest the presence of novel proteins and/or organisms in this microbial community and that additional research and alternative bioinformatics approaches may help further characterize the metagenomes of the VAG Mine pond.

4.1. Nucleotide sequence accession numbers

Raw WMS Illumina sequence data from this study were submitted to the NCBI SRA under accession numbers SRS872561 (1-S: site 1, 0.5 m), SRS962537 (2-S: site 2, 0.5 m), SRS963313 (3-S: site 3, 0.5 m), SRS963552 (1-M: site 1, 3.5 m), SRS963574 (2-M: site 2, 3.5 m), SRS963594 (3-M: site 3, 3.5 m), SRS963611 (2-B: site 2, 25 m), and SRS963627 (3-B: site 3, 25 m).

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

Funding for this study was made possible through a Vermont Genetics Network (VGN) Project Award for Elizabeth Dolci. VGN is funded through the IDeA Networks of Biomedical Research Excellence Program (grant number 8P20GM103449) of the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH. We acknowledge and thank Timothy Hunter and Scott Tighe from the Advanced Genome Technologies Core at the University of Vermont for their technical support in the development and preparation of genomic DNA extractions as well as Illumina TrueSeq library preparations. Brewster Kingham from the DNA Sequencing & Genotyping Center, Delaware Biotechnology Institute, University of Delaware completed the WMS Illumina sequencing. Additionally, we acknowledge and thank Steve Schlipf and Johnson State College undergraduate scholar Jaclyn Baker for their technical assistance on this project, Jaclyn Baker’s work was supported by the Johnson State College President’s Fund For Excellence in Teaching and Learning.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2016.11.004.

References

[1] R. Levey, Memorandum: summary of biological and chemical assessments conducted in 2005 & 2007 within the Lamoille River and Missisquoi River watersheds in Lowell and Eden, Vermont Agency of Natural Resources Department of Conservation, Water Quality Division. 2008.
[2] D.M. Levitan, J.M. Hammarstrom, M.E. Gunter, R.R. Seal, I.-M. Chou, N.M. Piatak, Mineralogy of mine waste at the Vermont Asbestos Group mine, Belvidere Mountain, Vermont. 2009. Am. Mineral. 94 1063–1066, http://dx.doi.org/10.2138/am.2009.3258.
[3] S. Andrews, F. Krueger, A. Seconds-Pichon, F. Biggins, S. Wingett, FastQC. A quality control tool for high throughput sequence data. Babraham Bioinformatics. 2014. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
[4] A.M. Bolger, M. Lohse, B. Usadel, Trimmmomatic: a flexible trimmer for Illumina sequence data. 2014. Bioinformatics 30 2114–2120, http://dx.doi.org/10.1093/bioinformatics/btu170.
[5] B. Langmead, C. Trapnell, M. Pop, L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. 2009. Genome Biol. 10 R25, http://dx.doi.org/10.1186/gb-2009-10-3-25.
[6] Y. Zhao, H. Tang, Y. Ye, RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. 2012. Bioinformatics 28 125–126, http://dx.doi.org/10.1093/bioinformatics/bts595.
[7] D.H. Huson, S. Mitra, H.-J. Ruscheweyh, N. Weber, S.C. Schuster, Integrative analysis of environmental sequences using MEGAN4. 2011. Genome Res. 21 1552–1560, http://dx.doi.org/10.1101/gr.120618.111.
[8] D.H. Parks, G.W. Tyson, P. Hugenholtz, R.C. Beiko, STAMP: statistical analysis of taxonomic and functional profiles. 2014. Bioinformatics 30 3123–3124, http://dx.doi.org/10.1093/bioinformatics/btu494.