Data Article

Seasonal Ely Copper Mine Superfund site shotgun metagenomic and metatranscriptomic data analysis

Lesley-Ann Giddings\textsuperscript{a,b,*}, George Chlipala\textsuperscript{d}, Heather Driscoll\textsuperscript{c}, Kieran Bhave\textsuperscript{a}, Kevin Kunstman\textsuperscript{d}, Stefan Green\textsuperscript{d}, Katherine Morillo\textsuperscript{a}, Holly Peterson\textsuperscript{e}, Mark Maienschein-Cline\textsuperscript{d}

\textsuperscript{a} Department of Chemistry & Biochemistry, Middlebury College, Middlebury, VT 05753, USA
\textsuperscript{b} Department of Chemistry, Smith College, Northampton, MA 01063, USA
\textsuperscript{c} Vermont Genetics Network, Department of Biology, Norwich University, Northfield, VT 05663, USA
\textsuperscript{d} Research Resources Center, University of Illinois at Chicago, Chicago, IL 60612, USA
\textsuperscript{e} Department of Geology, Guilford College, Greensboro, NC 27403, USA

\textbf{Article history:}
Received 6 July 2020
Revised 28 August 2020
Accepted 31 August 2020
Available online 8 September 2020

\textbf{Keywords:}
Acid rock drainage
Metagenome
Metatranscriptome
Illumina NextSeq
Differential analysis

\textbf{Abstract}
High throughput sequencing data collected from acid rock drainage (ARD) communities can reveal the active taxonomic and functional diversity of these extreme environments, which can be exploited for bioremediation, pharmaceutical, and industrial applications. Here, we report a seasonal comparison of a microbiome and transcriptome in Ely Brook (EB-90M), a confluence of clean water and upstream tributaries that drains the Ely Copper Mine Superfund site in Vershire, VT, USA. Nucleic acids were extracted from EB-90M water and sediment followed by shotgun sequencing using the Illumina NextSeq platform. Approximately 575,933 contigs with a total length of 1.54 Gbp were generated. Contigs of at least a size of 3264 (N50) or greater represented 50% of the sequences and the longest contig was 488,568 bp in length. Using Centrifuge against the NCBI "nt" database 141 phyla, including candidate phyla, were detected. Roughly 380,000 contigs were assembled and \sim{}1,000,000 DNA and \sim{}550,000 cDNA sequences were identified and function-
ally annotated using the Prokka pipeline. Most expressed KEGG-annotated microbial genes were involved in amino acid metabolism and several KEGG pathways were differentially expressed between seasons. Biosynthetic gene clusters involved in secondary metabolism as well as metal- and antibiotic-resistance genes were annotated, some of which were differentially expressed, colocalized, and coexpressed. These data can be used to show how ecological stimuli, such as seasonal variations and metal concentrations, affect the ARD microbiome and select taxa to produce novel natural products. The data reported herein is supporting information for the research article “Characterization of an acid rock drainage microbiome and transcriptome at the Ely Copper Mine Superfund site” by Giddings et al. [1].

© 2020 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

### Specifications Table

| Subject | Microbial Ecology, Genomics and Molecular Biology |
|---------|--------------------------------------------------|
| Specific subject area | Metagenomics |
| Type of data | Tables, figures, raw data |
| How data were acquired | Shotgun metagenomic and metatranscriptomic sequence data were acquired using an Illumina NextSeq500 instrument. Centrifuge was used to perform a read-based taxonomic analysis of metagenomic data. Prokka was used to detect and functionally annotated open reading frames. The predicted amino acid sequence was searched against Swiss-Prot database using DIAMOND. KEGG orthology annotations were predicted for open reading frames. All differential and statistical analyses on taxonomic summaries were performed in edgeR [2], BacMet [3], antiSMASH 5.0 [4], ARTS version 2.0 [5] databases were used to annotate genes. |
| Data format | Annotated data, Bray-Curtis dissimilarity matrices, Non-metric multidimensional scaling (NMDS) plots, principal component analysis (PCA) plots, heat map and hierarchal clustering, raw count data, and gradient plots. |
| Parameters for data collection | Seasonal environmental water and sediment samples were collected and sequenced. Five water and three sediment samples from summer as well as three sediment samples from winter. |
| Description of data collection | Shotgun metagenomic and metatranscriptomic sequencing was performed using an Illumina NextSeq500 instrument. |
| Data source location | Sediment (July 28th, 2017 and January 14th, 2018) and water (July 14th, 2017 and July 28th, 2017) samples were collected 90 m upstream from the mouth of Ely Brook (EB-90M) at Ely Copper Mine, Vershire, VT, USA (43°55′9″ N, 72°17′11″ W). |
| Data accessibility | Data are shown in this article. Raw metagenomic and metatranscriptomic data have been deposited in the Sequence Read Archive of the National Center for Biotechnology Information (BioProject identifier, PRJNA540505). Taxonomic and functional annotations as well as normalized count data used for all analyses are available in a public repository: Repository name: FigShare Data identification number: 10.6084/m9.figshare.c.4864863 Direct URL to data: https://doi.org/10.6084/m9.figshare.c.4864863 |
| Related research article | L.-A. Giddings, G. Chipala, K. Kunstman, S. Greene, K. Morillo, K. Bhave, H. Peterson, H. Driscoll, M. Maienschein-Cline, Characterization of an acid rock drainage microbiome and transcriptome at Ely Copper Mine Superfund site, PLoS One, 15(8) (2020) e0237599. https://doi.org/10.1371/journal.pone.0237599 |
Value of the Data

- This is the first characterization of an acid rock drainage (ARD) metagenome and transcriptome within the Vermont copper belt region, USA, which is comprised of Ely Copper Mine, Elizabeth Mine, and Pike Hill Copper Mine.
- The metagenomic data provide seasonal taxonomic profiles of the microbial diversity in the sediment and water of EB-90M.
- Active taxa in ARD environments are understudied and the metagenomic and metatranscriptomic data provide insight into their seasonal functional roles within these acidic, metal-rich environments.
- These data can be used to perform comparative taxonomic and functional analyses with other ARD metagenomes.
- These data can be used to bioprospect enzymes that can be exploited for the bioremediation of metal polluted environments.
- These data can be used to identify novel genes encoding proteins involved in the production of bioactive secondary metabolites, which can be used for pharmaceutical and industrial applications.

2. Data Description

Ten water and six sediment samples at Ely Brook (EB-90M) (Fig. 1), Ely Copper Mine Superfund site were collected in July 2017 and January 2018. Shotgun metagenomic sequencing of nucleic acids extracted from water and sediment samples generated \( \sim 31,545,991 \) reads with an average length of 147 bp and a total length of 1.54 Gb for 11 samples. Samples of the same sample type (i.e., water or sediment) or season (i.e., summer or winter) were treated as biological replicates. Summer water samples were denoted as July_Water1, July_Water2, July_Water3, July_Water4, July_Water5. Summer sediment samples were denoted as July_Sed1, July_Sed2, and July_Sed3. Winter sediment samples were denoted as Jan_Sed1, Jan_Sed2, and Jan_Sed3. All winter water samples (five samples) did not yield viable sequencing data. Of the remaining 11 samples, \( \sim 12 \) Gb of data (50M clusters) were produced per sample with an average of 25,181,359 reads per sample over a range of 8,657,966 and 44,323,783 reads for both metagenomic and metatranscriptomic data. Contigs of \( \geq 3264 \) bp (N50) represented 50% of data and the longest contig was 488,568 bp in length. Using Centrifuge [6] to perform read-based taxonomic annotation, 141 distinct phyla were annotated, including candidate phyla (Table 1). Taxonomic differences across season and sample type were observed by NMDS and PCA analyses of normalized count data (i.e., counts per million) between the bacteria, archaea, and fungi in samples as well as molecule types (Figs. 2–8). Differences between molecule type (i.e., DNA or RNA) across sample type and season were assessed by multivariate principal component analyses (PCA) (Fig. 9). Using Prokka-annotated open reading frames [7], Kyoto Encyclopedia of Genes and Genomes (KEGG) reference pathways [8] were annotated and quantified (Table 2). Significantly differentially expressed KEGG pathways and genes in winter versus summer were defined as having winter/summer RNA \( p \)-values \( \leq 0.05 \) for the interaction of season and molecule type followed by false discovery rate (FDR) corrections [9] (\( q \)-values) \( \leq 0.05 \) (Figs. 10–12). Secondary metabolite gene clusters (Table 3), metal resistance genes (Table 4), and antibiotic resistance genes were identified (Table 5). Approximately 288 metal resistance genes were differentially expressed between winter and summer seasons (Fig. 13). Furthermore, some of these genes were colocalized and coexpressed with genes involved in secondary metabolism (Table 6; Figs. 14–18).
Fig. 1. Vermont copper belt. A) Map of Vermont copper belt (highlighted in yellow), which includes Ely Copper Mine (sampling site), Pike Hill Mine, and Elizabeth Mine. B) Map of Ely Brook sample site, which drains Ely Copper Mine. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Table 1
Taxonomic annotation. List of 141 unique phyla across water and sediment metagenomic samples at EB-90 M. sk, superkingdom; k, kingdom; p, phylum. Incertae sedis represents kingdoms that have not been assigned.

| Unique phyla across water and sediment metagenomic samples |
|-----------------------------------------------------------|
| sk_Archaea;k__Archaea incertae sedis;p__Archaea Parvarchaeota | sk_Archaea;k__Bacteria incertae sedis;p__Planctomycetes |
| sk_Archaea;k__Archaea incertae sedis;p__Candidatus Korarchaeota | sk_Archaea;k__Bacteria incertae sedis;p__Proteobacteria |
| sk_Archaea;k__Archaea incertae sedis;p__Candidatus Micrarchaeota | sk_Archaea;k__Bacteria incertae sedis;p__Spirochaetes |
| sk_Archaea;k__Archaea incertae sedis;p__Candidatus Nanohaloarchaeota | sk_Archaea;k__Bacteria incertae sedis;p__Synergistetes |
| sk_Archaea;k__Archaea incertae sedis;p__Candidatus Parvarchaeota | sk_Archaea;k__Bacteria incertae sedis;p__Tenericutes |
| sk_Archaea;k__Archaea incertae sedis;p__Crenarchaeota | sk_Bacteria;k__Bacteria incertae sedis;p__Thermodesulfobacteria |
| sk_Archaea;k__Archaea incertae sedis;p__Euryarchaeota | sk_Bacteria;k__Bacteria incertae sedis;p__Thermotogae |
| sk_Archaea;k__Archaea incertae sedis;p__Nanoarchaeota | sk_Bacteria;k__Bacteria incertae sedis;p__Verrucomicrobia |
| sk_Archaea;k__Archaea incertae sedis;p__Thaumarchaeota | sk_Bacteria;k__Bacteria incertae sedis;p__candidate division CPR2 |
| sk_Bacteria;k__Bacteria incertae sedis;p__Actinobacteria | sk_Bacteria;k__Bacteria incertae sedis;p__candidate division CPR3 |
| sk_Bacteria;k__Bacteria incertae sedis;p__Aquificae | sk_Bacteria;k__Bacteria incertae sedis;p__candidate division NC10 |
| sk_Bacteria;k__Bacteria incertae sedis;p__Armatimonadetes | sk_Bacteria;k__Bacteria incertae sedis;p__candidate division WPS-2 |
| sk_Bacteria;k__Bacteria incertae sedis;p__Bacteroidetes | sk_Eukarya;k__Eukarya incertae sedis;p__candidate division WW3 |
| sk_Bacteria;k__Bacteria incertae sedis;p__Balneolaeota | sk_Eukarya;k__Eukarya incertae sedis;p__Armatimonadetes |
| sk_Bacteria;k__Bacteria incertae sedis;p__Caldiserica | sk_Eukarya;k__Eukarya incertae sedis;p__Candidate division |
| sk_Bacteria;k__Bacteria incertae sedis;p__Caldiviricarchaeota | sk_Eukarya;k__Eukarya incertae sedis;p__Candidate division |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidateus Acidothermia | sk_Eukarya;k__Eukarya incertae sedis;p__Colponemidia |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidateus Adlerbacteria | sk_Eukarya;k__Eukarya incertae sedis;p__Euglenida |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidateus Amesbacteria | sk_Eukarya;k__Eukarya incertae sedis;p__Eukaryota |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidateus Attribacteria | sk_Eukarya;k__Eukarya incertae sedis;p__Fagamophycocae |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidateus Azambacteria | sk_Eukarya;k__Eukarya incertae sedis;p__Haplosporidia |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidateus Beekwithbacteria | sk_Eukarya;k__Eukarya incertae sedis;p__Phaeophycaceae |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidateus Berkelbacteria | sk_Eukarya;k__Eukarya incertae sedis;p__Picozoa |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidateus Campbellbacteria | sk_Eukarya;k__Eukarya incertae sedis;p__Siphonophycaceae |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidateus Cloacimonetes | sk_Eukarya;k__Fungi;p__Ascomycota |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidateus Collierbacteria | sk_Eukarya;k__Fungi;p__Basidiomycota |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidateus Curtissbacteria | sk_Eukarya;k__Fungi;p__Blastocladiomycota |

(continued on next page)
Unique phyla across water and sediment metagenomic samples

| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Daviesbacteria | sk_Eukaryota;k__Fungi;p__Chytridiomycota |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Falkowbacteria | sk_Eukaryota;k__Fungi;p__Cryptomycota |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Giovannonibacteria | sk_Eukaryota;k__Fungi;p__Entorrhizomycota |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Gottesmanbacteria | sk_Eukaryota;k__Fungi;p__Fungi incertae sedis |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Gracilibacteria | sk_Eukaryota;k__Fungi;p__Microsporidia |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Jorgensenbacteria | sk_Eukaryota;k__Fungi;p__Zooopagomyctota |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Kaiserbacteria | sk_Eukaryota;k__Metazoa;p__Acanthocephala |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Kuenenbacteria | sk_Eukaryota;k__Metazoa;p__Annelida |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Levybacteria | sk_Eukaryota;k__Metazoa;p__Arthropoda |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Magasanikbacteria | sk_Eukaryota;k__Metazoa;p__Brachiopoda |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Melainabacteria | sk_Eukaryota;k__Metazoa;p__Bryozoa |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Moranbacteria | sk_Eukaryota;k__Metazoa;p__Chaetognatha |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Nomurabacteria | sk_Eukaryota;k__Metazoa;p__Chelata |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Omnitrophica | sk_Eukaryota;k__Metazoa;p__Chondaria |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Pacebacteria | sk_Eukaryota;k__Metazoa;p__Ctenophorah |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Parcubacteria | sk_Eukaryota;k__Metazoa;p__Ctenophora |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Peregrinibacteria | sk_Eukaryota;k__Metazoa;p__Ctenophora |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Roizmanbacteria | sk_Eukaryota;k__Metazoa;p__Ctenophora |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Saccharibacteria | sk_Eukaryota;k__Metazoa;p__Echinodermata |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Shapirobacteria | sk_Eukaryota;k__Metazoa;p__Enteroprotea |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Tectomicrobia | sk_Eukaryota;k__Metazoa;p__Entoprocta |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Uhrbacteria | sk_Eukaryota;k__Metazoa;p__Gastrotricha |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Woesebacteria | sk_Eukaryota;k__Metazoa;p__Gnathostomulida |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Wolfbacteria | sk_Eukaryota;k__Metazoa;p__Hemichordata |
| sk_Bacteria;k__Bacteria incertae sedis;p__Candidatus Yanofskybacteria | sk_Eukaryota;k__Metazoa;p__Kinorhyncha |
| sk_Bacteria;k__Bacteria incertae sedis;p__Chlamydiae | sk_Eukaryota;k__Metazoa;p__Mollusca |
| sk_Bacteria;k__Bacteria incertae sedis;p__Chlorobi | sk_Eukaryota;k__Metazoa;p__Nematoda |
| sk_Bacteria;k__Bacteria incertae sedis;p__Chloroflexi | sk_Eukaryota;k__Metazoa;p__Nematomorpha |
| sk_Bacteria;k__Bacteria incertae sedis;p__Chloroarchaeota | sk_Eukaryota;k__Metazoa;p__Nemertea |
| sk_Bacteria;k__Bacteria incertae sedis;p__Cyanobacteria | sk_Eukaryota;k__Metazoa;p__Onychophora |
| sk_Bacteria;k__Bacteria incertae sedis;p__Coprothermobacterota | sk_Eukaryota;k__Metazoa;p__Placozoa |
| sk_Bacteria;k__Bacteria incertae sedis;p__Cyanobacteria | sk_Eukaryota;k__Metazoa;p__Platyhelminthes |

(continued on next page)
### Table 1 (continued)

Unique phyla across water and sediment metagenomic samples

| sk_Bacteria;k__Bacteria incertae sedis;p__Deferribacteres | sk_Eukaryota;k__Metazoa;p__Porifera |
| sk_Bacteria;k__Bacteria incertae sedis;p__Deinococcus-Thermus | sk_Eukaryota;k__Metazoa;p__Priapulida |
| sk_Bacteria;k__Bacteria incertae sedis;p__Dictyoglomi | sk_Eukaryota;k__Metazoa;p__Rhombozoa |
| sk_Bacteria;k__Bacteria incertae sedis;p__Elusimicrobia | sk_Eukaryota;k__Metazoa;p__Rotifera |
| sk_Bacteria;k__Bacteria incertae sedis;p__Fibrobacteres | sk_Eukaryota;k__Metazoa;p__Tardigrada |
| sk_Bacteria;k__Bacteria incertae sedis;p__Firmicutes | sk_Eukaryota;k__Metazoa;p__Xenacoelomorpha |
| sk_Bacteria;k__Bacteria incertae sedis;p__Gemmatimonadetes | sk_Eukaryota;k__Viridiplantae;p__Chlorophyta |
| sk_Bacteria;k__Bacteria incertae sedis;p__Ignavibacteriae | sk_Eukaryota;k__Viridiplantae;p__Streptophyta |
| sk_Bacteria;k__Bacteria incertae sedis;p__Kiritimatiellaeota | sk_Viroids;k__Viroids incertae sedis;p__Viroids incertae sedis |
| sk_Bacteria;k__Bacteria incertae sedis;p__Nitrospirae | sk_Viroids;k__Viroids incertae sedis;p__Viruses incertae sedis;p__Viruses incertae sedis |

### Table 2

**BRITE level 1 annotation statistics.** Average percentages of normalized counts that were annotated at BRITE level 1 using the KEGG database.

| Average BRITE level 1 Observations Across All Sediment Samples | Percentage, % |
|---------------------------------------------------------------|----------------|
| 09100 Metabolism                                             | 0.1726149 |
| 09120 Genetic Information Processing                        | 0.036885 |
| 09130 Environmental Information Processing                   | 0.0244681 |
| 09140 Cellular Processes                                     | 0.0197084 |
| 09150 Organismal Systems                                     | 0.017033 |
| 09160 Human Diseases                                         | 0.001983 |
| 09180 BRITE Hierarchies                                      | 0.1912439 |
| 09190 Not Included in Pathway or BRITE                       | 0.0202365 |
| Unassigned                                                   | 0.5039421 |

### Table 3

**antiSMASH annotation.** Summary of the number of genes and gene clusters annotated by antiSMASH 5.0 as well as those that match the Prokka-annotated data.

| Total count of contigs | 575,933 |
|------------------------|---------|
| Total number of contigs annotated by antiSMASH               | 1589    |
| Total number of contigs not annotated by antiSMASH           | 574,344 |
| antiSMASH annotated genes                                     | 10,579  |
| antiSMASH annotated genes that aligned with PROKKA analyzed data | 4977    |
| antiSMASH annotated genes that did not align with PROKKA analyzed data | 5602    |
| antiSMASH annotated gene clusters that align with PROKKA analyzed data | 1349    |
| antiSMASH annotated gene clusters that did not align with PROKKA analyzed data | 240     |
| antiSMASH annotated gene clusters that aligned with PROKKA analyzed data and met the criteria of a sum of at least 100 counts across all samples and 10 counts in three samples | 449     |
| Annotated gene clusters that remain after filtering by p-interaction value | 176     |
| Annotated gene clusters that remain after subsequent filtering by q-winter/summer RNA value | 65      |
Table 4
Metal resistance gene annotation. Statistics on metal resistance genes identified using the BacMet database. A gene identifier (i.e., gene ID) is defined as a gene symbol plus a number, for example, copR_X, where X is a number. The eight missing gene IDs that were not expressed, include copR_13, corC_121, cusR_32, czcA_647, nikE_38, pstC_144, ruvB_54, Int_122. Differentially expressed features were defined based on 1) the interaction term p-value (Type:Season) of 0.05 or less in combination with 2) the pairwise seasonal comparison of RNA expression ('Winter.rna/Summer.rna') FDR-adjusted p-value (q-value) of 0.05 or less.

| DNA and RNA | 296,476 |
| DNA and RNA with gene IDs | 161,984 |
| Number of gene symbols found in DNA and RNA | 5579 |
| Number of gene symbols found in DNA and RNA in BacMet database | 133 |
| Number of gene IDs from DNA and RNA found in BacMet | 7021 |
| Number of gene IDs from DNA in BacMet database that are not found in RNA | 8 (copR_13, corC_121, cusR_32, czcA_647, nikE_38, pstC_144, ruvB_54, Int_122) |
| Number of gene IDs that are differentially expressed | 947 |
| Ref. Set: Actinobacteria | Contigs | 1–3712 | 3,713– | 4,242– | 10,000– | 15,485– | 25,000– | 35,574– | 45,001– | 66,478– | 85,001– | 110,410– | 135,001– | 169,689– | 240,000– | 330,000– | 440,000– | 501,400– |
|-------------------------|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Core Essential Genes    |        | 395    | 303    | 387    | 387    | 387    | 383    | 381    | 392    | 377    | 384    | 379    | 382    | 399    | 391    | 370    | 331    | 328    |
| BGC Hits                | 136     | 9      | 98     | 81     | 90     | 74     | 70     | 119    | 74     | 95     | 75     | 80     | 134    | 142    | 139    | 92     | 81     | 1588   |
| Known Resistance        | 944     | 71     | 595    | 411    | 526    | 420    | 331    | 580    | 406    | 473    | 332    | 474    | 683    | 744    | 742    | 346    | 423    | 8501   |
| Phylogeny/ HGT          | N/A     | N/A    | N/A    | N/A    | N/A    | N/A    | N/A    | N/A    | N/A    | N/A    | N/A    | N/A    | N/A    | N/A    | N/A    | N/A    | N/A    | N/A    |
| Nodes                   | 1–3712  | 3,713– | 4,242– | 9999   | 15,484 | 25,000 | 35,573 | 45,001 | 66,478 | 85,001 | 110,409 | 135,001 | 169,689 | 240,000 | 330,000 | 440,000 | 501,400 |
| Core Essential Genes    | 516     | 359    | 506    | 495    | 506    | 505    | 504    | 517    | 492    | 509    | 488    | 502    | 510    | 504    | 488    | 444    | 444    | 8289   |
| BGC Hits                | 136     | 9      | 98     | 81     | 90     | 74     | 70     | 119    | 74     | 95     | 75     | 80     | 134    | 142    | 139    | 92     | 81     | 1589   |
| Known Resistance        | 944     | 71     | 595    | 411    | 526    | 420    | 331    | 580    | 406    | 473    | 332    | 474    | 683    | 744    | 742    | 346    | 423    | 8501   |

**Table 5**

ARTS annotated contigs. ARTS (https://arts3.ziemertlab.com) annotated contigs using Actinobacteria and Alphaproteobacteria reference sets. Phylogeny is not applicable (N/A) to this metagenomic dataset. These data are also located on Figsquare; DOI: 10.6084/m9.figshare.c.11879226. URL – https://doi.org/10.6084/m9.figshare.c.11879226.
### Table 6
Colocalized and/or coexpressed genes. Colocalized and/or coexpressed BacMet genes with BGCs. Differentially expressed features were defined based on 1) the interaction term p-value (Type:Season) (p-interaction) of 0.05 or less in combination with 2) the pairwise seasonal comparison of RNA expression (‘Winter.rna/Summer.rna’) FDR-adjusted p-value (q-value) of 0.05 or less.

| Cluster | Contig # | Genus | Percent match to genus | Gene ID | Gene | Function | p-interaction value | p-winter RNA/summer RNA value | q-winter RNA/summer RNA value | Winter RNA/summer RNA Log_2 fold change |
|---------|---------|-------|------------------------|--------|------|----------|--------------------|-------------------------------|---------------------------------|-----------------------------------|
| Cluster 1 Metal resistance | 4689 | 4689 | Metal resistance | 4689 | FHBHJPKI_167716 | psiA_11 | L-methionine sulfoximine/L-methionine sulfone acetyltransferase | 4.37E-05 | 0.000733 | 0.00468 | −2.94 |
| Cluster 1 Secondary metabolite | 4689 | 4689 | Metal resistance | 4689 | FHBHJPKI_167725 | Involved in synthesis of homoserine lactone-nonribosomal peptide | 1.93078E-11 | 2.85272E-09 | 5.90633E-08 | −3.50087665 |
| Cluster 2 Metal resistance | 80 | 80 | Candidatus Solibacter uromitus Ellin66976 | FHBHJPKI_12377 | mgtA_4 | Magnesium-transporting ATPase-2C P-type 1 | 0.00701 | 0.0182 | 0.0167 | −2.01 |
| Cluster 2 Secondary metabolite | 80 | 80 | Candidatus Solibacter uromitus Ellin66976 | FHBHJPKI_12365 | lgrD_9 | Linear gramicidin synthase subunit D | 0.841648661 | 0.004247211 | 0.021296771 | −2.46382034 |
| Cluster 3 Metal resistance | 12,335 | 12,335 | Acidobacterium capsulatum ATCC SL196 | FHBHJPKI_283288 | mdtA_189 | Multidrug resistance protein MdtA | 3.68E-10 | 1.23E-16 | 8.92E-15 | −4.71 |
| Cluster 3 Secondary metabolite | 12,335 | 12,335 | Acidobacterium capsulatum ATCC SL196 | FHBHJPKI_283295 | crtB_77 | All-trans-phytoene synthase | 0.15602993 | 0.00446359 | 0.03026887 | −2.05628682 |
| Cluster 4 Metal resistance | 214 | 214 | Ralstonia solanacearum CMR15 | FHBHJPKI_24632 | smtB_5 | Succinyl-CoA-l-malate CoA-transferase beta subunit | 0.042 | 0.00392 | 0.0027 | −4.18 |
| Cluster 4 Secondary metabolite | 214 | 214 | Ralstonia solanacearum CMR15 | FHBHJPKI_24627 | shc_2 | Squalene–hopene cyclase | 0.067593728 | 0.00537894 | 0.00359324 | −2.48913791 |
| Cluster 5 Metal resistance | 185 | 185 | Candidatus Koribacter versatilis Ellin345 | FHBHJPKI_22308 | czcA_9 | Cobalt-zinc-cadmium resistance protein CzcA | 0.0321 | 0.000717 | 0.00968 | −2.31 |
| Cluster 5 Secondary metabolite | 185 | 185 | Candidatus Koribacter versatilis Ellin345 | FHBHJPKI_22329 | Putative ligase/MSMEl_5285 | Putative ligase/MSMEI_5285 | 0.270379299 | 0.013944273 | 0.111550994 | −1.87907266 |
| Cluster 6 Metal resistance | 4698 | 4698 | Metal resistance | 4698 | FHBHJPKI_166937 | mdtA_99 | Multidrug resistance protein MdtA | 0.0292 | 0.0299 | 0.106 | −3.16 |
| Cluster 6 Secondary metabolite | 4698 | 4698 | Metal resistance | 4698 | FHBHJPKI_167934 | ppsE_5 | Involved in synthesis of Phthiocerol/phenolphthiocerol polyketide | 0.01973596 | 0.00158153 | 0.009156685 | −1.99280005 |
Fig. 3. Bray-Curtis dissimilarity indices for bacteria in sediment. A) Matrix of dissimilarity indices calculated for genera of bacteria in sediment samples using the Bray-Curtis method. ‘Sed’ = sediment. B) NMDS plot to visualize the dissimilarity between genera of bacteria in summer (July_Sed1, July_Sed2, and July_Sed3 in orange) and winter (Jan_Sed1, Jan_Sed2, and Jan_Sed3 in blue) sediment collected at EB-90M. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3. Experimental Design, Materials and Methods

3.1. Sample collection

On July 28th, 2017 and January 14th, 2018, Ely Brook (43°55′9″ N, 72°17′11″ W), 90 m upstream from the mouth of the brook (EB-90M), was sampled along with unsaturated sediment (10 cm deep). The physicochemical properties, nucleic acid extraction, library preparation, and metatranscriptomic and metatranscriptomic sequencing, taxonomic annotation of raw reads, metagenomic assembly, and functional annotations of these samples were reported by Giddings et al. [1].
A.

|            | Jan_Sed3  | Jan_Sed1  | Jan_Sed2  | July_Sed1 | July_Sed3 | July_Sed2 |
|------------|-----------|-----------|-----------|-----------|-----------|-----------|
| Jan_Sed3   | 0.02604939|           |           |           |           |           |
| Jan_Sed1   | 0.02281105| 0.02776086|           |           |           |           |
| Jan_Sed2   | 0.08264324| 0.07758015| 0.08953024|           |           |           |
| July_Sed1  | 0.0761283 | 0.06914731| 0.07990833| 0.0259954 |           |           |
| July_Sed2  | 0.07445164| 0.06903036| 0.07944745| 0.03169012| 0.0253259 |

B.

Fig. 4. Bray-Curtis dissimilarity indices for eukaryota in sediment. A) Matrix of dissimilarity indices calculated for genera of eukaryota in sediment samples using the Bray-Curtis method. ‘Sed’ = sediment. B) NMDS plot to visualize the dissimilarity between genera of eukaryota in summer (July_Sed1, July_Sed2, and July_Sed3 in orange) and winter sediment (Jan_Sed1, Jan_Sed2, and Jan_Sed3 in blue) collected at EB-90M. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2. Statistical comparison of microbial community, DNA, and RNA

EB-90M samples of the same sample type or season were treated as biological replicates. Subsets (i.e., season or sample type) of data were compared to each other in statistical analyses. Beta diversity was evaluated via Bray-Curtis measure of dissimilarity [10] using default parameters in R in the vegan library [11]. Prior to analysis, data were \( \log_{10}(x + 1) \) transformed and the resulting dissimilarity indices were used to generate NMDS in R using the metaMDS functions in vegan and ggplot2 library [11, 12]. Multivariate PCAs were performed in Partek Flow software v8.0 to assess sample group variation based on genera using normalized read counts from read-based taxonomic annotations and quantification. Feature counts (e.g., taxon) were standardized prior to the PCA so that the contribution of each feature did not depend on its variance. PCA
|                | July_Sed1 | July_Sed3 | July_Sed2 | July_Water4 | July_Water2 | July_Water1 | July_Water5 | July_Water3 |
|----------------|-----------|-----------|-----------|-------------|-------------|-------------|-------------|-------------|
| July_Sed1      | 0.14453049|           |           |              |              |              |              |              |
| July_Sed3      |           | 0.09605933| 0.05430961|              |              |              |              |              |
| July_Sed2      | 0.24711851| 0.19535912| 0.20074453|              |              |              |              |              |
| July_Water4    | 0.29968121| 0.16805777| 0.20470678|              |              |              |              |              |
| July_Water1    | 0.28777118| 0.17660392| 0.19807451| 0.09659091   | 0.04812817  |              |              |              |
| July_Water2    | 0.29968121| 0.16805777| 0.20470678|              |              |              |              |              |
| July_Water3    | 0.2453889 | 0.18719198| 0.19203378| 0.03242655   | 0.11782878  | 0.08741529  | 0.04802209  |

**Fig. 5.** Bray-Curtis dissimilarity indices for archaea in summer. A) Matrix of dissimilarity indices calculated for genera of archaea in summer samples using the Bray-Curtis method. 'Sed' = sediment. B) NMDS plot to visualize the dissimilarity between genera of archaea in summer sediment (July_Sed1, July_Sed2, and July_Sed3 in orange) and water (July_Water1, July_Water2, July_Water3, July_Water4, and July_Water5 in blue) collected at EB-90M. The ellipse indicates a clustering of more than 3 samples. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Plots were generated for DNA and RNA using 1) normalized read counts (i.e., fractions for relative abundance) from the metagenomic assembly and 2) normalized read counts from the metatranscriptome, respectively. Heat maps and hierarchal clusters were generated in Partek Flow v8.0 using the following, respectively: 1) normalized counts of taxa from the metagenome and predicted open reading frames (ORFs) across samples and 2) the Euclidean dissimilarity index and average linkage method to cluster similar expression patterns and taxon abundances. The normalized data were standardized to a mean of zero and a standard deviation of 1 prior to hierarchal clustering.

### 3.3. Differential expression and visualization of KEGG pathways

Differentially expressed KEGG pathways were represented by color gradation maps (Figs. S14–S15). Log2-fold-changes from gene expression analysis results were converted to a color gradation using KEGG Mapper – Color Pathway tool ([https://www.genome.jp/kegg/tool/map_pathway3.html](https://www.genome.jp/kegg/tool/map_pathway3.html)), where blue denotes decreased expression in the winter (RGB color code #6363F7) and red denotes increased expression in the winter (RGB color code #FF000). Genes with no change in expression are shaded in light gray (RGB color code #D3D3D3). Genes shaded
A.

| Sample       | Bray-Curtis Dissimilarity Indices |
|--------------|----------------------------------|
| July_Sed1    | 0.03925216                       |
| July_Sed3    | 0.03172766                       |
| July_Sed2    | 0.03172766                       |
| July_Water4  | 0.11343901                       |
| July_Water2  | 0.16102392                       |
| July_Water1  | 0.15869585                       |
| July_Water5  | 0.11630563                       |
| July_Water3  | 0.11187043                       |

B.

Fig. 6. Bray-Curtis dissimilarity indices for bacteria in summer. A) Matrix of dissimilarity indices calculated for genera of bacteria in summer samples using the Bray-Curtis method. 'Sed' = sediment. B) NMDS plot to visualize the dissimilarity between genera of bacteria in summer sediment (July_Sed1, July_Sed2, and July_Sed3 in orange) and water (July_Water1, July_Water2, July_Water3, July_Water4, and July_Water5 in blue) collected at EB-90M. The ellipse indicates a clustering of more than 3 samples. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in white indicates that the gene was undetected in the dataset. The numbers in boxes refer to enzyme nomenclature from the KEGG database. Expression data (i.e., normalized counts) for sediment were fit to a linear model, assuming a negative binomial distribution, that included season (i.e., winter versus summer), molecule type (i.e., RNA versus DNA), as well as the interaction of season and molecule type (p-interaction). Pairwise comparison tests of season were performed within and between each data type and p-values were FDR-corrected [9]. Significant differentially expressed genes met the following criteria: a molecule type-season interaction term p-value of 0.05 or less in combination with an FDR-adjusted p-value (q-value) of 0.05 or less for the pairwise comparison of winter RNA versus summer RNA. Significant data were indicated by an orange star; however, the overall expression of a node may include other genes.

3.4. Analysis of genes involved in natural product biosynthesis, metal resistance, and antibiotic resistance

Contigs were mined for secondary metabolite biosynthetic gene clusters (BGCs) in the bacterial and fungal antiSMASH 5.0 [4] database using default parameters. The BacMet database was used to mine DNA and RNA for experimentally validated metal resistance genes [3]. After filtering annotated-BGCs and BacMet genes that had ≥ 100 raw counts in each sample and at least
Fig. 7. Bray-Curtis dissimilarity indices for eukaryota in summer. A) Matrix of dissimilarity indices calculated for genera of eukaryota in summer samples using the Bray-Curtis method. "Sed" = sediment. B) NMDS plot to visualize the dissimilarity between genera of eukaryota in summer sediment (July_Sed1, July_Sed2, and July_Sed3 in orange) and water (July_Water1, July_Water2, July_Water3, July_Water4, and July_Water5 in blue) collected at EB-90M. The ellipse indicates a clustering of more than 3 samples. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

10 counts in three or more samples, relative BGC and BacMet gene expression was assessed by comparing the counts of Prokka-annotated transcripts to those of DNA using the criteria described by Giddings et al. [1]. Gradient plots were generated in Partek Flow v8.0 for differentially expressed BGCs and those co-expressed with metal resistance genes. Contigs were also mined for antibiotic resistance genes that were within close proximity or colocalized with BGCs using the Antibiotic Resistant Target Seeker (ARTS) version 2 [5] using default parameters. Duplication and BGC proximity, resistance model screens, and genomes that mapped to the following reference phyla were selected: Actinobacteria and Alphaproteobacteria.
Fig. 8. Taxonomic differences. PCA plot demonstrating the differences between genera in summer water and sediment as well as summer (orange) and winter (blue) sediment. Plot is based on normalized read counts at the genus level from the taxonomic annotation and quantification of paired-end reads. The sample name notation is based on the month the sample was collected, the sample type (i.e., sediment or water), and individual sample number. ‘Sed’ = sediment. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Fig. 9. Differences in DNA and RNA. PCA plots of A) DNA in water and sediment and B) RNA present in summer and winter sediment based on normalized counts of all functionally annotated genes from the metagenomic assembly, demonstrating differences between sample type. Each gene’s normalized read count contributes equally to the PCA. The sample name notation is based on the month the sample was collected, the sample type (i.e., sediment or water), and individual sample number. ‘Sed’ = sediment.
Select differentially expressed KEGG Pathways in winter sediment samples

| q-value       | log2 fold change |
|---------------|------------------|
| 9.81E-26      | -2.17            |
| 2.81E-22      | -1.98            |
| 7.38E-11      | -1.13            |
| 2.09E-09      | -1.2             |
| 4.07E-09      | -0.82            |
| 1.43E-06      | -0.81            |
| 2.36E-06      | -0.79            |
| 2.76E-06      | -0.78            |
| 3.37E-06      | -0.78            |
| 3.98E-06      | -0.72            |
| 1.09E-05      | -0.72            |
| 2.23E-05      | 0.66             |
| 3.05E-05      | 0.81             |
| 1.37E-04      | 1.35             |
| 1.42E-04      | 0.98             |
| 1.49E-04      | 0.58             |
| 1.41E-04      | 0.81             |
| 4.62E-04      | 0.58             |
| 7.06E-04      | 0.81             |
| 8.50E-04      | 0.58             |
| 9.58E-04      | 0.81             |
| 2.66E-03      | -1.03            |
| 2.79E-03      | -0.8             |
| 2.94E-03      | -0.56            |
| 3.66E-03      | -0.71            |
| 3.82E-03      | -0.52            |
| 3.88E-03      | -0.61            |
| 4.16E-03      | 0.65             |
| 6.39E-03      | 0.57             |
| 6.57E-03      | 0.97             |
| 6.65E-03      | -0.61            |
| 6.67E-03      | -0.61            |
| 6.70E-03      | -0.69            |
| 8.37E-03      | -0.53            |
| 8.72E-03      | 0.61             |
| 9.33E-03      | -0.46            |
| 1.09E-02      | -0.46            |
| 1.47E-02      | -0.05            |
| 1.74E-02      | -0.43            |
| 2.60E-02      | -0.39            |
| 3.05E-02      | -0.39            |
| 3.50E-02      | -0.54            |
| 4.12E-02      | 0.56             |
| 4.70E-02      | 0.55             |

Fig. 10. Significantly differentially expressed KEGG pathways. Bar graph of select significantly differentially expressed KEGG pathways in winter versus summer. Differentially expressed pathways were defined based on an unadjusted p-value ≤ 0.05 for the interaction term (molecule type-season) in combination with a q-winter/summer RNA value ≤ 0.05, respectively. Red and blue represent increased and decreased expression in winter, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
**Fig. 11.** Carbon fixation in photosynthetic organisms. Carbon metabolism KEGG reference pathway map ([https://www.kegg.jp/pathway/map00710](https://www.kegg.jp/pathway/map00710)) with color gradation highlighting KEGG genes that change significantly between seasons. Log₂-fold-changes from gene expression analyses were converted to a color gradation using the KEGG Mapper - Color Pathway tool, where blue denotes decreased expression in the winter (RGB color code #6363F7) and red denotes increased expression in the winter (RGB color code #FF0000). The Log₂-fold-changes range from −2.33 (blue) to +1.88 (red). Genes with no change in expression are shaded in light gray (RGB color code #D3D3D3) and genes shaded white were undetected in the dataset. Significantly differentially expressed genes are indicated by a star and met the following criteria: p-interaction value ≤ 0.05 in combination with a q-winter/summer RNA value ≤ 0.05, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Fig. 12. Nitrogen metabolism gene expression. Nitrogen metabolism KEGG reference pathway map diagram (https://www.kegg.jp/pathway/map00910) with color gradation highlighting KEGG genes that change significantly between seasons. Log₂-fold-changes from gene expression analyses were converted to a color gradation using the KEGG Mapper – Color Pathway tool, where blue denotes decreased expression in the winter (RGB color code #6363F7) and red denotes increased expression in the winter (RGB color code #FF0000). The Log₂-fold-changes range from −3.92 (blue) to +1.91 (red). Genes with no change in expression are shaded in light gray (RGB color code #D3D3D3) and genes shaded white were undetected in the dataset. Significantly differentially expressed genes are indicated by a star and met the following criteria: $p$-interaction value $\leq 0.05$ in combination with a $q$-winter/summer RNA value $\leq 0.05$, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Fig. 13. **Metal resistance gene expression.** Hierarchical clustering and heat map of differentially expressed select (288) genes (e.g., dnaK, copA, copB, copD, ps65, cusA, cusB, mdtA, mdtB, actP, mco, ycnJ, corA, csoR, and copZ) from the BacMet database across sediment samples. Increases or decreases in gene expression range from −2.04 (blue) to +2.04 (red). All data met the following criteria: p-interaction value ≤ 0.05 in combination with a q-winter/summer RNA value ≤ 0.05, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 14. **Colocalization and coexpression of metal resistance and secondary metabolite genes.** Gradient plot demonstrating the differential coexpression of mdtA, a metal resistance gene encoding a multidrug resistance protein, with a gene (ppsE) annotated to be involved in phthiocerol/phenolphthiocerol polyketide biosynthesis in contig 4698 (20,390 nucleotides long) in summer (orange) and winter (blue). The lines on the y-axis represent the maximum, minimum, and mean of the standardized expression values (i.e., counts per million). All data met the following criteria: p-interaction ≤ 0.05 in combination with a p-winter/summer RNA ≤ 0.05, respectively. Nucleotide positions in contig are shown below gene IDs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Fig. 15. Colocalization and coexpression of metal resistance and secondary metabolite genes. Gradient plot demonstrating the differential coexpression of mgtA, a metal resistance gene encoding a cation transport ATPase that mediates magnesium influx into the cytosol, with genes (igrd) annotated to be involved in gramicidin biosynthesis in contig 80 (113,676 nucleotides long) in summer (orange) and winter (blue). The lines on the y-axis represent the maximum, minimum, and mean of the standardized expression values (i.e., counts per million). Only mgtA met the following criteria: p-interaction ≤ 0.05 in combination with a q-winter/summer RNA ≤ 0.05, respectively. Nucleotide positions in contig are shown below gene IDs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Fig. 16. Colocalization and coexpression of metal resistance and secondary metabolite genes. Gradient plot demonstrating the differential coexpression of czcA, a metal resistance gene encoding a cobalt-zinc-cadmium resistance protein, with a ligase/MSMEI_5285 gene annotated to be involved in the biosynthesis of a polyketide in contig 185 (85,942 nucleotides long) in summer (orange) and winter (blue). The lines on the y-axis represent the maximum, minimum, and mean of the standardized expression values (i.e., counts per million). Only czcA met the following criteria: $p$-interaction $\leq 0.05$ in combination with a $q$-winter/summer RNA $\leq 0.05$, respectively. Nucleotide positions in contig are shown below gene IDs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
**Fig. 17.** Colocalization and coexpression of metal resistance and secondary metabolite genes. Gradient plot demonstrating the differential coexpression of *smtB*, a zinc-resistance gene encoding a repressor protein of the metallothionein gene *smtA*, with a gene annotated to be involved in the biosynthesis of a terpene in contig 214 (80,995 nucleotides long) in summer (orange) and winter (blue). The lines on the y-axis represent the maximum, minimum, and mean of the standardized expression values (i.e., counts per million). Only SmtB met the following criteria: \( p \)-interaction \( \leq 0.05 \) in combination with a \( q \)-winter/summer RNA \( \leq 0.05 \), respectively. Nucleotide positions in contig are shown below gene IDs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Fig. 18.** Colocalization and coexpression of metal resistance and secondary metabolite genes. Gradient plot demonstrating the differential coexpression of *mdtA*, a metal resistance gene encoding multidrug resistance protein, with genes annotated to be involved in the biosynthesis of a terpene in contig 12,335 (11,958 nucleotides long) in summer (orange) and winter (blue). The lines on the y-axis represent the maximum, minimum, and mean of the standardized expression values (i.e., counts per million). Only *mdtA* met the following criteria: \( p \)-interaction \( \leq 0.05 \) in combination with a \( q \)-winter/summer RNA \( \leq 0.05 \), respectively. Nucleotide positions in contig are shown below gene IDs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

Acknowledgments

We thank Ed Hathaway (Environmental Protection Agency) & Linda Elliot (Vermont Department of Environmental Conservation) for escorting us to EB-90M at Ely Mine. This work was funded by the United States Geological Survey, Vermont Water Resources and Lake Studies Center 104b grant (subaward number G16AP00087), Middlebury College, and the Institutional Development Award from the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) under grant number P20GM103449. The content of this research is solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH. The bioinformatics analyses performed by the UIC Research Informatics Core were supported in part by the National Center for Advancing Translational Sciences under grant number UL1TR002003.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi: 10.1016/j.dib.2020.106282.

References

[1] L.-A. Giddings, G. Chipala, K. Kunstman, S. Green, K. Morillo, K. Hhave, H. Peterson, H. Driscoll, M. Mainschein-Cline, Characterization of an acid rock drainage microbiome and transcriptome at the Ely Copper Mine Superfund site, PLoS ONE 15 (8) (2020) e0237599.
[2] D.J. McCarthy, Y. Chen, G.K. Smyth, Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation, Nucl. Acids Res. 40 (10) (2012) 4288–4297.
[3] C. Pal, J. Bengtsson-Palme, C. Rensing, E. Kristiansson, D.G.J. Larsson, BacMet: antibacterial biocide and metal resistance genes database, Nucl. Acids Res. 42 (D1) (2013) D737–D743.
[4] K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S.Y. Lee, M.H. Medema, T. Weber, AntiSMASH 5.0: updates to the secondary metabolite genome mining pipeline, Nucl. Acids Res. 47 (W1) (2019) W81–W87.
[5] M. Alanjary, B. Kronmiller, M. Adamek, K. Blin, T. Weber, D. Huson, B. Philmus, N. Ziemert, The Antibiotic resistant target seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery, Nucl. Acids Res. 45 (W1) (2017) W42–W48.
[6] D. Kim, L. Song, F.P. Breitwieser, S.L. Salzberg, Centrifuge: rapid and sensitive classification of metagenomic sequences, Genome Res. 26 (12) (2016) 1721–1729.
[7] T. Seemann, Prokka: rapid prokaryotic genome annotation, Bioinformatics 30 (14) (2014) 2068–2069.
[8] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, KEGG as a reference resource for gene and protein annotation, Nucl. Acids Res. 44 (D1) (2016) D457–D462.
[9] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc. Ser. B. 57 (1) (1995) 289–300.
[10] J.R. Bray, J.T. Curtis, An ordination of the upland forest communities of Southern Wisconsin, Ecol. Monogr. 27 (4) (1957) 325–349.
[11] J. Oksanen, Vegan: An introduction to ordination, University of Oulu, Oulu, 2018.
[12] H. Wickham, Ggplot2, 1st ed., Springer, New York, NY, USA, 2009.