Uncovering the Viral Ecology of Antarctic Soil Microbial Communities

Laidlaw Research Summer Report

John Henry Lotz-McMillen

Research done with Dr. Sarah Stuart Johnson, Dr. Shauna Bennett,

Dr. Gareth Trubl, and Ruby Gilmore

at the Lawrence Livermore National Laboratory, Livermore, CA

GEORGETOWN UNIVERSITY
Center for Research & Fellowships

ABSTRACT:

Studying microbial communities in extreme environments, such as the frigid and arid soils of the Antarctic Dry Valleys, offers a unique opportunity to understand how microbial life can persist under harsh environmental conditions. Evaluating the taxonomic and functional diversity of the viruses present in these communities is critical to developing a comprehensive view of these resilient ecosystems. Viruses play remarkable but obscure roles in microbial ecosystems, managing microbial populations and nutrient cycling while also promoting horizontal gene transfer. Despite viruses' theoretical potential to influence the ecological structure of soil microbial communities, much remains unknown about the full significance of soil viruses due to the technical challenges associated with studying them. To access information about the viral ecology in our microbial communities, we utilized metagenomics to identify the genetic material present in our samples and reconstruct viral populations (vOTUs), groups of clustered viral sequences sharing a specified threshold of genetic similarity that function as taxonomic units. We analyzed the metagenomes from a set of soil samples taken as a depth series from Taylor Valley, and identified 23 dsDNA vOTUs. Further analysis revealed that vOTU richness increases with depth, with no significant viral DNA found in the first two cm of soil. All vOTUs were identified as previously unclassified *Caudoviricetes* dsDNA bacteriophage, predicted to infect three different bacterial phylums (Proteobacteria, Actinobacteriota, and Desulfobacterota B).

INTRODUCTION:

Supported by the Laidlaw Foundation and the Johnson Biosignatures Lab at Georgetown University, Dr. Shauna Bennett, Ruby Gilmore, and I flew out to the Lawrence Livermore

National Laboratory to carry out a bioinformatic investigation of a genetic sequencing dataset extracted from a series of Antarctic soil samples. The trip had several purposes: to strengthen the connection between the Johnson Lab (dedicated to the understanding of how biosignatures prevail within planetary environments) and renowned soil microbial ecologist (and astrobiology enthusiast), Dr. Gareth Trubl; to develop and possibly redirect an unfinished research project in the Johnson Lab by Yu Angela Bai and Julie Grace Bevilacque that involved a diverse sampling of soils from the McMurdo Dry Valleys of Antarctica; to provide an opportunity for the three of us to develop skills in metagenomic analysis that we could take back with us to the Johnson Lab, specifically for the metagenomic study of viruses; and (as an added bonus) to attend the International Soil Virus Conference, hosted at Lawrence Livermore by Trubl and Dr. Joanne Emerson of UC Davis.

Despite its status as the driest and coldest place on Earth, the McMurdo Dry Valleys have nonetheless been found to harbor simple microbial ecosystems. One of the largest ice-free geographic regions of Antarctica, the Dry Valleys are characterized by mean annual surface temperatures of -20 ˚C, with extremely cold winters and summers subject to variable temperatures leading to multiple freeze-saw cycles daily (Adriaenssens et. al, 2017; Bai et. al). In addition, any exposed soils are highly arid, kept below ~2% water content by the katabatic winds and osmotic barriers resulting from the soil's high salinity (Bai et. al). The mineral-based permafrost soils of this hyperarid desert contain very little organic matter, keeping organic life restricted to simple microbial communities (Adriaenssens et. al, 2017). But in the context of astrobiology, the simplicity of these Antarctic soil ecosystems provides an invaluable opportunity to study microbial persistence and cell survival strategies in extreme environments. In addition,

by sampling these sparse but robust soil microbiomes, we can attempt to understand how their ecological structure and processes may contribute to their survival.

With the goal of studying the relationship between taxonomic and functional diversity in Antarctic soil microbial communities, the Johnson Lab led an expedition in 2016 to acquire a comprehensive set of soil samples from the Dry Valleys and several other locations (Bai et. al). Eight years later, with the guidance of Trubl, Bennett, Gilmore and I returned to this dataset with a new research objective: resolve the viral element of these samples and make predictions about the ecological significance of viruses in these extremophile microbial communities. Viruses play remarkable but understudied roles in soil ecosystems generally, managing microbial populations and nutrient cycling while also promoting horizontal gene transfer. However, because of the immense challenges associated with isolating and identifying viruses from soil samples, we have only just begun to realize the viral diversity present in these ecosystems and their significance for microbial life. This is because soil is one of the most complex environments to study due to its highly heterogeneous nature, consisting of a variable mixture of inorganic matter and organic biomass, which makes isolating virions or viral DNA within it very difficult (Trubl et. al, 2020).

RESEARCH PROCESSES:

Identification of Viral Operational Taxonomic Units:

While the majority of the 2016 samples were processed via 16S SSU ribosomal RNA gene analysis (irrelevant to the study of viruses, as viruses lack endogenous ribosomes), we focused on one particular depth series taken from Taylor Valley that had been processed using whole-genome shotgun (WGS) sequencing. Samples in this series were taken at the surface and at depths of 2 cm, 4 cm, 6 cm, 8 cm, and 10 cm. This allowed us to study how viral community

structure varied spatially, as well how sampling depth affected viral taxonomic diversity, functional diversity, and which microbial hosts the viruses might be infecting. But more importantly, the WGS sequencing allowed us to study each sample as a metagenome.

In order to holistically capture the genetic material present in a sample, the DNA is first isolated from the soil and then broken up into smaller fragments and sequenced via WGS sequencing. At this point, all the DNA from the sample has been extracted as a metagenome. While there are different strategies for isolating DNA from the soil, the methods used in this study started by isolating cells, and then subjecting the DNA from lysed cells to WGS sequencing, meaning that primarily intracellular viral DNA (such as proviruses or episomes) was recovered. Metagenomics is extremely useful for studying viruses in soils, as it does not rely on PCR amplification of a universal marker gene like 16S rRNA (viruses have small and highly diverse genomes that lack common marker genes), and does not require culturing of the sample microbiome (which results in the loss of most of the diversity present), as metagenomics directly accesses the DNA of a microbial community as it exists in its environment (Trubl et. al, 2020). Ideally, the full genomes of the viruses and microbes present in the sample are retained within these metagenomes. However, in order to classify and analyze these genomes, the genome sequences must be reconstructed, as contiguous elements of a single genome will be cleaved into shorter sequences during the WGS sequencing process.

The reconstruction and identification of viral genomes is typically an intensive bioinformatic process due to the sheer amount of sequencing data being manipulated, and this step constituted the first phase of our project. We started with six pre-assembled metagenomes, one for each sample, that each consisted of contiguous sequences (contigs), sequences that had been put back together through a process called metagenome assembly. Each project collaborator

was given a pair of metagenomes and tasked with identifying possible viral genomes amongst the contigs present via a computational process called ‘binning,’ where contigs were sorted into bins of similar contigs corresponding to a potential viral genome. To cast as wide a net as possible - while trying to avoid false positives - we each simultaneously ran two separate virus identification pipelines, each involving a different ‘binning’ algorithm. In one pipeline, the dataset was run through VirSorter2 two times, once with wider search parameters that had a minimum contig length of 5 kilobases, and then a second time with stricter parameters involving a minimum contig length of 10 kilobases. VirSorter2 compares contigs from each bin to a database of known viral genomes, hallmark genes, and genomic structures, and then deduces the potential of a contig being a viral genome to a certain degree of confidence. If a contig satisfies the pre-specified confidence threshold and is greater in length than the dictated minimum sequence length, it is saved as a potential viral genome. In addition, viral characteristics are organized into six *a priori* viral ‘groups’ (or virus types): double stranded DNA (dsDNA) phage, RNA phage, single stranded DNA phage, *Lavidaviridae*, or NCLDV (Guo et. al, 2021). In the second pipeline, geNomad was run on the same contig datasets. Similar to VirSorter2, geNomad compares contig gene content with a marker gene database, but it also utilizes a deep neural network that analyzes oligonucleotide frequency patterns to predict viral genomes. geNomad is also able to give reliable taxonomic annotations for identified viral genomes (Camargo et. al, 2023). We ultimately chose a minimum sequence length of 10 kilobases for both binning methods, as most dsDNA bacteriophage we would expect to find are known to have genomes sizes of over 10 kilobases, and these algorithms in their current state are too unreliable at identifying viral genomes shorter than 10 kilobases (Roux et. al, 2019).

Once we obtained the results from both viral identification workflows, the viral genomes identified from each metagenome were compiled into a single dataset. We then clustered this dataset based on 95% average nucleotide identity over 85% alignment fraction, comparing all of the contigs in the dataset against each other and removing any repeated sequences that were similar above the state thresholds (Roux et. al, 2019). This left us with a complete finalized list of unique viral operational taxonomic units (vOTUs; with a vOTU being a possible taxonomically unique viral genome) containing all the viruses we could find across the soil samples using our integrated approach.

At this stage, the 'viruses' (vOTUs) we had identified were the result of taking real genetic information stored in DNA molecules, recording the nucleotide sequences of the molecules using an imperfect sequencing method, and then taking that data and using arbitrary metrics to identify similarities and structural patterns between sequences and between the sequences and external databases in order to recombine the genetic information digitally into longer sequences that resembled viral genomes. Inherent to metagenomics is the reality that biases can be introduced to the data at every stage of the metagenomics workflow, from when the microbes are first isolated from the soil sample, to the accumulation of errors during next-generation sequencing, to how genomes are reassembled, and to how viral genomes are identified. So while metagenomic studies enable progress in our understanding of soil virology, an appropriate and significant amount of uncertainty must be associated with results (Trubl et. al, 2020). That said, the vOTUs identified in this research project adhere to the currently agreed upon standards of the field - the 'minimum information about an uncultivated virus genome' standards - and as such, the vOTUs we found were acceptable approximations of the viral diversity present in our Antarctic soils, given the computational power and information we had

access to (Roux et. al, 2019). Furthermore, finding the consensus vOTU list from the results of the gene-content based and k-mer/oligonucleotide based binning tools gave us higher confidence in our results.

Analysis of vOTU Taxonomic, Functional, and Host Diversity Across the Depth Gradient:

After resolving the final list of vOTUs that existed within our soil samples, we were then able to move to the central research questions of the project and develop ecological meaning from the processed genetic information.

In our first line of inquiry, we used a variety of softwares to probe our vOTU dataset for information on the genes present in the viral genomes, possible microbial hosts the viruses may be infecting, and the taxonomic identity of the viral genomes. To determine the diversity of genes present in our genomes, DRAM-v and Prokka were used to functionally annotate our vOTU dataset. These tools work by predicting the identity of viral genes via comparison to external gene databases, allowing the user to evaluate the metabolic potential of a metagenome (Shaffer et. al, 2020)(Seeman, 2014). One of the most critical roles that viruses play is altering the metabolism of microbial communities by enriching their hosts with viral metabolic genes, which can influence how nutrients flow through a microbiome and how a microbiome may shape its environment. As such, knowledge of the functional diversity present in a viral sample can aid in the understanding of microbial ecosystem dynamics on the whole. In addition to predicting gene function, DRAM-v also is able to identify auxiliary metabolic genes (AMGs), genes that have been identified to not be viral, but microbial. These genes are a result of some sort of contamination of a viral genome with microbial genomes, for example as a result of a splicing error when a provirus is excised from a host chromosome, and can also be significant modifiers

of host metabolism, as they can amplify or expand host metabolic processes (Shaffer et. al, 2020).

iPHoP was used to predict potential host taxonomies that our vOTUs may have been infecting. As a metagenomic study does not involve direct observation of virus-microbial interactions, information regarding host association is not initially obvious. Nonetheless, a virus is only as significant as its virus-host relationship, as viruses are entirely dependent on the molecular infrastructure of their host to reproduce, and viruses create ecological impact through modifying or killing their hosts. In order to generate this critical host association information from a metagenomic workflow, iPHoP integrates a variety of computational approaches to predict host taxonomy, all the way up to the genus level. These approaches include alignment of viral genomes with microbial genomes to test for sequence similarity, and machine learning approaches that use learned virus-host pairs to predict hosts for novel phages (Roux et. al, 2023).

Finally, vConTACT 2.0 was used to predict the taxonomy of our vOTUs. While viral genome databases like the International Committee on Taxonomy of Viruses are rapidly expanding thanks to advances in metagenomics, these databases are still scarcely and sparsely representative of global soil virus diversity. vConTACT 2.0 evaluates the relatedness of newly identified viral genomes at the genus level by computationally analyzing the similarities in viral gene composition between known and unknown viral genomes, accomplished by clustering genomes together using a network-based genome taxonomy framework. If viral genomes only cluster with each other and not significantly with any known viral genomes, their taxa are considered unclassified (Jang et. al, 2019).

Our second line of inquiry involved determining the alpha and beta diversity, community structure, and community membership of the viral communities represented by our data at each

sample depth. In order to have information on how the populations of each vOTU may have been distributed across the sample depths, we had to return to our original unprocessed metagenomes, as this is where the information regarding the relative abundances of each viral population in each sample could be found. Using a program provided by Trubl, we mapped the original datasets of sequenced reads onto our list of all possible vOTUs for each of the samples. Each read in a metagenome has the potential to be complementary to part of a vOTU, and by computationally deriving the read coverage depth (the number of reads mapped to a vOTU) and breadth (the proportion of a vOTU sequence covered by reads), we could predict how common that vOTU was in that sample. By comparing the coverage values of different vOTUs, and then normalizing those values based on a variety of metrics such as vOTU sequence length that biased the values, we could construct a comprehensive normalized relative abundance table to see the abundances of each vOTU at each depth. This table then served as a substrate to determine the alpha diversity (richness, evenness, Shannon diversity) of each sample, and the beta diversity (using Bray-Curtis principal coordinates of analysis (PCoA) to determine the similarity of community compositions between samples).

Throughout this series of bioinformatic transformations of our sequencing data, we transitioned from running the aforementioned software on graphical user interfaces like KBase and NMDC to directly accessing DOE high performance computing via a login terminal and manually running the softwares on a command line.

FINDINGS:

Ultimately, 21 unique vOTUs were identified. Despite being a relatively low number of vOTUs for a soil metagenomics study at this sampling scale, this magnitude is reasonable for a

couple of reasons: overall abundance in the extreme environment of the Dry Valleys is expected to be less; DNA was isolated from the soil samples via methods optimized for preserving microbial cells, not exogenous viral DNA or virions; and WGS sequencing and metagenome assembly performed seven to eight years ago may not have optimal yield. All vOTUs were identified as unclassified dsDNA bacteriophages, which is consistent with the lack of data on Antarctic soil viruses. geNomad suggested the vOTUs belong to the Caudoviricetes of tailed dsDNA bacteriophages.

After evaluating the alpha diversity of the Antarctic soil samples, vOTU richness was suggested to increase with depth, ranging from zero species at 0 cm and 2 cm to fifteen species at 4 cm and 20 species at 10 cm. Evenness and Shannon diversity was found to be highest at 10 cm in depth, decreasing from 4 cm to 6 cm and then increasing with depth. Shannon diversity was suggested to be relatively low across all samples, ranging from 1.5-2.0.

Following Bray-Curtis PCoA, the 6 cm and 8 cm samples were suggested to be similar in vOTU composition, and the 10 cm sample was dissimilar to the other samples in the depth series. While this may suggest that community composition changes with sample depth, statistical analysis has yet to be conducted to confirm significance differences and similarities in the community compositions, and deeper depth samples would have needed to be taken to provide additional evidence for such a pattern. A heat map of the relative abundance table revealed one species to be of the highest abundance at each sample depth by a large margin, which may explain the low evenness results. There are a variety of possible reasons for this pattern that need to be further investigated: for example, the virus could be a provirus prevalent in the most abundant microbe sampled, or its genomic sequence could be heavily represented by the reads in each genome, despite the reads belonging to a diversity of more unique virus genomes.

iPHoP predicted that 9 of the 21 vOTUs were linked to gammaproteobacteria hosts, and only three phylums were represented in host prediction (Proteobacteria/Pseudomonadota, Actinobacteriota, and Desulfobacterota B). This may be consistent with existing literature on Antarctic soil bacterial abundances, as gammaproteobacteria have been found to be dominant in neutral Antarctic soils (Adriaenssens et. al, 2017).

While further analysis of functional annotations is still required, Prokka and DRAM-v identified a variety of viral genes, exhibited in Table 1 (see index). DRAM-v identified 1 AMG in vOTU from the 8cm depth sample, the transporter gene ABC.PE.P1 involved in the peptides/nickel transport system.

REFLECTION ON OUTCOMES AND FUTURE DIRECTIONS:

Future research will apply metagenomic analyses to these depth samples with a focus on microbial communities to identify metagenome assemble genomes (MAGs). Analysis of MAGs may suggest specific virus-host relationships that could allow for observations of the functionality of these interactions in the greater Antarctic soil ecosystem. For example, iPHoP can be rerun using the MAGs we find as a reference microbial dataset, which could give us viral host results that are more fine-tuned to our samples.

We will continue to focus on the suggested relationship between taxonomic diversity, functional diversity, and depth. In particular, we will further unpack our functional diversity results, and explore how functional diversity changes with depth.

Next steps involve a technical replication of the metagenomics workflow to determine whether up-to-date metagenome assembly programs may yield larger metagenomes and ultimately more vOTUs. If different vOTUs are identified in this second run through, further

scrutiny must be applied to the validity of our viral genome identifications, and hopefully an updated, higher quality vOTU list may be developed and analyzed.

ACKNOWLEDGEMENTS:

This experience would not have been possible or as rewarding without the mentorship and support of Dr. Bennett, Dr. Johnson, and Dr. Trubl, and the collaborative work of Gilmore. I would also like to express my deepest gratitude to the Laidlaw Scholars Foundation and the Georgetown University Laidlaw Scholars Programme for your support on my endeavors. Finally, I would like to extend my appreciation to the staff of Lawrence Livermore National Laboratory for your hospitality, and for the welcome shared by the soil virus ecology community at the International Soil Virus Conference.

REFERENCES:

Adriaenssens, Evelien M., Rolf Kramer, Marc W. Van Goethem, Thulani P. Makhalanyane, Ian Hogg & Don A. Cowan. Environmental drivers of viral community composition in Antarctic soils identified by viromics. Microbiome 5, 83 (2017). https://doi.org/10.1186/s40168-017-0301-7

Bai, Yu Angela, Julie Grace Bevilacqua, Aria S. Hahn, Connor Morgan-Lang, Elena Zaikova, Scott Tighe, David Goerlitz, Nicole Wagner, Mia Vanderwilt, Amy McAdam, Christine A. Knudson, Cherie Achilles, Slavka Andrejkovičová, Sarah Stewart Johnson. Selective Pressure of Harsh and Variable Conditions Drives the Decoupling of Taxonomic and Functional Diversity. (unpublished).

Camargo, Antonio Pedro, Simon Roux, Frederik Schulz, Michal Babinski, Yan Xu, Bin Hu, Patrick S. G. Chain, Stephen Nayfach & Nikos C. Kyrpides. Identification of mobile genetic elements with geNomad. Nat Biotechnol (2023). https://doi.org/10.1038/s41587-023-01953-y

Camargo, Antonio Pedro, Stephen Nayfach, I-Min A Chen, Krishnaveni Palaniappan, Anna Ratner, Ken Chu, Stephan J Ritter, T B K Reddy, Supratim Mukherjee, Frederik Schulz, Lee Call, Russell Y Neches, Tanja Woyke, Natalia N Ivanova, Emiley A Eloe-Fadrosh, Nikos C Kyrpides, Simon Roux, IMG/VR v4: an expanded database of uncultivated virus genomes within a framework of extensive functional, taxonomic, and ecological metadata, Nucleic Acids Research, Volume 51, Issue D1, 6 January 2023, Pages D733–D743, https://doi.org/10.1093/nar/gkac1037

Guo, Jiarong, Ben Bolduc, Ahmed A. Zayed, Arvind Varsani, Guillermo Dominguez-Huerta, Tom O. Delmont, Akbar Adjie Pratama, M. Consuelo Gazitúa, Dean Vik, Matthew B. Sullivan & Simon Roux. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome 9, 37 (2021). https://doi.org/10.1186/s40168-020-00990-y

Jang, Ho Bin, Benjamin Bolduc, Olivier Zablocki, Jens H. Kuhn, Simon Roux, Evelien M. Adriaenssens, J. Rodney Brister, Andrew M Kropinski, Mart Krupovic, Rob Lavigne, Dann Turner & Matthew B. Sullivan. Taxonomic assignment of uncultivated prokaryotic

virus genomes is enabled by gene-sharing networks. Nat Biotechnol 37, 632–639 (2019). https://doi.org/10.1038/s41587-019-0100-8

Roux, S., Adriaenssens, E., Dutilh, B. et al. Minimum Information about an Uncultivated Virus Genome (MIUViG). Nat Biotechnol 37, 29–37 (2019). https://doi.org/10.1038/nbt.4306

Roux, S., Camargo, A. P., Coutinho, F. H., Dabdoub, S. M., Dutilh, B. E., Nayfach, S., & Tritt, A. (2023). iPHoP: An integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. PLoS biology, 21(4), e3002083. https://doi.org/10.1371/journal.pbio.3002083

Seemann, Torsten, Prokka: rapid prokaryotic genome annotation, Bioinformatics, Volume 30, Issue 14, July 2014, Pages 2068–2069, https://doi.org/10.1093/bioinformatics/btu153

Shaffer, Michael, Mikayla A Borton, Bridget B McGivern, Ahmed A Zayed, Sabina Leanti La Rosa, Lindsey M Solden, Pengfei Liu, Adrienne B Narrowe, Josué Rodríguez-Ramos, Benjamin Bolduc, M Consuelo Gazitúa, Rebecca A Daly, Garrett J Smith, Dean R Vik, Phil B Pope, Matthew B Sullivan, Simon Roux, Kelly C Wrighton. DRAM for distilling microbial metabolism to automate the curation of microbiome function, Nucleic Acids Research, Volume 48, Issue 16, 18 September 2020, Pages 8883–8900, https://doi.org/10.1093/nar/gkaa621

Trubl, G., Hyman, P., Roux, S., & Abedon, S. T. (2020). Coming-of-age characterization of soil viruses: a user's guide to virus isolation, detection within metagenomes, and viromics. Soil Systems, 4(2), 23.

INDEX:

Table 1. vOTU contig length, taxonomic classifications, and gene annotations generated by DRAM-v, IMG/VR BLAST9, and Prokka.

vOTU (contig ID)	Contig Length, bp	Virus taxonomy (via GeNomad or IMG/VR BLAST)	BLAST hits	Annotation
4_k151_101218	10,520	Caudoviricetes	112	[Prokka] DNA gyrase subunit A; DNA gyrase subunit B; DNA replication and repair protein RecF; Chromosomal replication initiator protein DnaA; Beta sliding clamp
4_k151_119893	23,322	Caudoviricetes	163	[Prokka] Tyrosine recombinase XerC
4_k151_129772	10,028	Caudoviricetes	16	[Prokka] tRNA uridine 5-carboxymethylaminomethyl modification enzyme MnmG; Ribosomal RNA small subunit methyltransferase G; Leucine--tRNA ligase; Chromosome-partitioning protein ParB; Chromosome partitioning protein ParA; IS1595 family transposase ISMtsp22
4_k151_151330	16,344	Caudoviricetes	13	[Prokka] ATP-dependent RecD-like DNA helicase
4_k151_163166	11,977	Caudoviricetes	7	[Prokka] acetyltransferase; Myo-inositol 2-dehydrogenase; transaminase; UDP-N-acetylglucosamine 4-epimerase; UDP-4-amino-4-deoxy-L-arabinose--oxoglutarate aminotransferase; Transcription antitermination protein RfaH; Lipid III flippase
4_k151_49688	11,864	Caudoviricetes	0	[Prokka] DNA-invertase hin
4_k151_51123	18,019	Caudoviricetes	7	[Prokka] Lon protease 2; Sodium-dependent dicarboxylate transporter SdcS; Sodium-dependent dicarboxylate transporter SdcS;
4_k151_55420	16,097	Caudoviricetes	2	[Prokka] carboxylate synthase; Poly(3-hydroxyalkanoate) polymerase; putative diacyglycerol O-acyltransferase tgs1
4_k151_58061	13,603	Caudoviricetes	12	[Prokka] RNA polymerase-associated protein RapA; FMN reductase (NADH) RutF; Protein YrdA; Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex; Dihydrolipoyl dehydrogenase; GDP-perosamine synthase
6_k151_6007	12,135	Caudoviricetes	5	[Prokka] Error-prone DNA polymerase
6_k151_	11,554	Caudoviricetes	7	[Prokka] Aerobactin synthase; L-lysine N6-monooxygenase; IS1182

7818				family transposase ISAzch1; Shikimate dehydrogenase (NADP(+))
8_k151_12891	23,115	Caudoviricetes	1	[Prokka] dehydrooxytetracycline reductase; ATP-dependent zinc metalloprotease FtsH; putative sensor histidine kinase TcrY; Transcriptional regulatory protein TcrA; NAD(P)H-quinone oxidoreductase, chloroplastic; NADH-quinone oxidoreductase; Na(+)/H(+) antiporter;
8_k151_26165	12,191	Caudoviricetes	521	[DRAM-v] tRNA-Met(cat); [Prokka] DNA primase; RNA polymerase sigma factor RpoD; Multifunctional non-homologous end joining protein LigD;
8_k151_34356	15,219	Caudoviricetes	4	[Prokka] ATP-dependent RecD-like DNA helicase
8_k151_38164	13,266	Caudoviricetes	1	none identified
8_k151_43	11,891	Caudoviricetes	13	[Prokka] Trehalase; Glucokinase; Transaldolase; Glucose-6-phosphate 1-dehydrogenase 1; 6-phosphogluconate dehydrogenase, NAD(+)-dependent, decarboxylating; Ribulose-phosphate 3-epimerase; Transketolase 2; Ubiquinone biosynthesis O-methyltransferase, mitochondrial
8_k151_60064	23,197	Caudoviricetes	3	[DRAM-v AMG] ABC.PE.P1; peptide/nickel transport system permease protein [Prokka] Glutathione transport system permease protein GsiD; Oligopeptide transport ATP-binding protein OppF; Oligopeptide transport ATP-binding protein OppD; Oligopeptide transport system permease protein OppC; Glutathione transport system permease protein GsiC; Oligopeptide-binding protein OppA; Protease HtpX
8_k151_81666	11,545	Caudoviricetes	58	[Prokka] D-inositol-3-phosphate glycosyltransferase; GDP-mannose 4,6-dehydratase
10_k151_39328	24,398	Caudoviricetes	156	[DRAM-v AMG] GlycosylTransferases [DRAM-v] integrase; putative DNA primase/helicase; adenine-specific DNA-methyltransferase; type III restriction enzyme
10_k151_56258	26,944	Caudoviricetes	8	[DRAM-v] polysaccharide biosynthesis protein PslH; exopolysaccharide production protein ExoZ; adenylate cyclase
10_k151_67091	12,249	Caudoviricetes	32	none identified