Investigation of antibiotic and arsenic resistance gene co-selection over a long-term stressor

Personnel: Ashley Shade, MSU (PI); Tammy Tobin, Susquehanna University (PI); Taylor Dunivin, MSU (Graduate student)

Background: The dissemination of antibiotic resistance genes (ARGs) is a pressing public health concern. By 2050, 10 million people are expected to die annually of infections from antibiotic resistant bacteria.[1] The One Health initiative recognizes the intrinsic link between evolution of bacterial resistance in clinical and environmental settings,[2] but environmental reservoirs of ARGs remain under-characterized.[3] Thus, predictions of the dissemination of ARGs in the environment are impeded by a lack of understanding of their diversification, maintenance, and dissemination.[4] Studies that measure environmental evolution of resistance are necessary to predict and prevent the dissemination of antibiotic resistance.

ARGs are often associated with genes conferring resistance to other microbial stressors, such as heavy metals or pollutants.[5–7] Insights into mechanisms of co-selection, dissemination, and diversification can be gained by examining resistance genes that co-occur, have the potential for horizontal gene transfer (HGT), and/or have lengthy evolutionary histories. For example, co-occurrences between resistance genes along an environmental gradient can highlight (co)selection pressures within an environment. Additionally, the presence of resistance genes on plasmids suggests a mechanism for their rapid dissemination and shortens timescales necessary to record evolutionary change. Finally, evolutionarily old resistance genes offer expansive known diversity, so understanding contemporary patterns of resistance gene diversification offers insights into modern selection pressures.

One example of resistance genes that co-occur, exhibit HGT, and are evolutionary old include ARGs and arsenic resistance genes (AsRGs). Like ARGs, AsRGs have important implications for public health, as they impact the biogeochemical cycling of the toxic metalloid arsenic;[8] however, investigations of the forces driving their co-selection are obscured by their ancient evolutionary origins and vertical inheritances:[9–11] some ARGs are expected to have evolved >2 billion years ago,[9] and several AsRGs likely evolved before bacterial and archaeal divergence.[12,13,10] Thus, community members may harbor these resistance genes without selective pressure from antibiotics or arsenic, and changes in community structure can impact the abundance of these resistance genes. ARGs and AsRGs have been shown to be correlated through physical linkage on both chromosomes and plasmids,[7] and HGT is known to be widespread.[4,8] Further knowledge of the contemporary diversity improves the known evolutionary history of these resistance genes, which is important because they are known to be ancient, and diversification of these resistance genes can therefore offer insights into modern selection pressures.

Long-term environmental disturbances impact multiple generations and provide opportunities to investigate the eco-evolutionary dynamics of resistance genes. Disturbances such as elevated temperature and pollution have been shown to increase diversification,[14] population-level diversity,[15] and rates of horizontal gene transfer (HGT)[16,17] Our model ecosystem is the Centralia, PA, coal mine fire which ignited in 1962 and has advanced along the coal seam ever since, creating a chronosequence (space-for-time proxy) of contemporary and historical fire impact. This multigenerational disturbance exposes soil microbial communities to increased temperatures (21-80°C)[18] as well as coal combustion pollutants like arsenic.[19] The surface soil microbial communities overlying the underground coal mine fire thus experience a multitude of stressors which are expected to increase the incidence of HGT.[16,17,20,21] We will examine evolutionary dynamics of resistance genes along this disturbance gradient by examining their co-occurrence, potential HGT, and diversification. *We hypothesize that AsRG and ARG co-occur along the Centralia chronosequence, that AsRG and ARG co-occur on plasmids and that these plasmids are more abundant in fire-affected soils, and that the genetic variation of resistance genes is greater in disturbed soils.*

Preliminary data: Clusters of orthologous groups (COGs)[22] were analyzed from 12 existing Centralia metagenomic datasets that were analyzed through the Joint Genome Institute's Integrated Microbial Genomes (IMG) portal.[23] The abundance of COGs for arsenite efflux pumps (ArsB and ACR3 family) and for aminoglycoside phosphotransferase (kanamycin resistance) were normalized to RNA polymerase β subunit (RpoB). Relative abundance of antibiotic and arsenic resistance COGs had positive Pearson's R correlations with each other (P value = 0.006) and soil temperature (P value = 0.012, 0.013 respectively). This suggests that antibiotic and arsenic resistance genes co-occur in Centralia and are more abundant in fire-affected soils. Additionally, our study of arsenic resistant isolates from Centralia fire-affected soil provides evidence suggesting HGT of three separate AsRGs (Dunivin, *et al*., *in review*), highlighting HGT as a potentially important evolutionary mechanism for AsRGs in Centralia. These AsRG sequences were identical to others in NCBI on an amino acid level but had divergent nucleotide sequences, showing unique sequence variants in Centralia.

Investigation of antibiotic and arsenic resistance gene co-selection over a long-term stressor

Personnel: Ashley Shade, MSU (PI); Tammy Tobin, Susquehanna University (PI); Taylor Dunivin, MSU (Graduate student)

Research approach: We will use the Centralia chronosequence as a model disturbance gradient to investigate the co-selection of antibiotic and arsenic resistance genes. We have >400 Gbases of existing shotgun metagenomic datasets (13 total) from surface soils in Centralia spanning fire-affected, recovered, and reference sites. We will assemble antibiotic and arsenic resistance genes from these datasets using a gene-targeted metagenome assembler, Xander.[24] This advanced assembly method offers multiple advantages for evolutionarily focused genetic studies by improving chances of finding diverse gene sequences, improving assembly of full length genes of interest, and revealing strain level variation.[24] Functional gene (FunGene) databases[25] will be used to guide the assembly and exist for >50 ARGs, several AsRGs, and single copy genes *gyrB* and *rpoB* which will be used for abundance normalization.[26] FunGene databases for remaining AsRG will be constructed. Thus, we will query all metagenomic datasets for each gene ARG and AsRG of interest, to output amino acid and nucleotide sequence, taxonomy, and abundance information for each assembled contig. We will then analyze the co-occurrence, diversity, and potential HGT of these genes.

Co-occurrence: The total gene abundance for all ARGs and AsRGs assembled using Xander will be normalized for each site and compared across the chronosequence. We will employ network analysis, a tool for assessing co-occurrence patterns in complex datasets,[27] to test for correlations between the relative abundance of ARGs and AsRGs in Centralia. This work can be further enhanced by repeating the analysis on individual sequence variants. While many co-occurrence patterns are possible, we will target co-selection by emphasizing positive correlations between these resistance genes and temperature. Correlations between specific sequence variants could be due to multiple evolutionary mechanisms: selection for these sequence variants, selection for organism(s) harboring these sequences, or HGT of physically linked variants.

HGT: The presence of resistance genes on plasmids or phages highlights their potential to be transferred in the environment. To test for HGT as a mechanism of co-occurrence, we will examine the potential transfer of co-occurring ARGs and AsRGs. We will use Recycler,[28] a recently developed tool to extract circular contigs from metagenomic datasets, to assemble plasmids and phages from along the Centralia chronosequence. These small circular elements will then be tested for presence of ARGs and AsRGs. We hypothesize that plasmids with resistance genes are more abundant in fire affected soils. We will estimate the abundance of plasmids or phages containing ARGs and AsRGs across the chronosequence by measuring both the number of contigs assembled and their coverage.

Diversity: The diversity of each resistance gene will be compared both within and across soils using phylogenetic diversity metrics. Phylogenetic trees will be constructed for all assembled genes. Using Xander, we can detect strain level variation; this genetic diversity in assembly data along with extensively characterized genes of interest allows for calculation of rates of synonymous to nonsynonymous substitutions. Comparing these rates along the Centralia chronosequence will offer insights into selection pressures on these resistance genes. Furthermore, Xander's assembly in amino acid space (rather than nucleotide) allows for assembly of diverse (and potentially novel) resistance gene sequences. Thus, this analysis will contribute to the known diversity of ARGs and AsRGs and therefore to evolutionarily ancient genes.

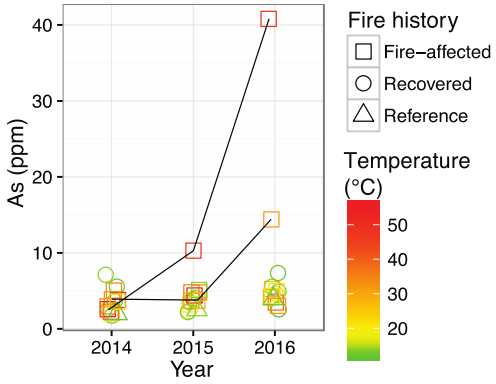

Figure 1. Soil arsenic concentrations in Centralia. Total arsenic was measured in 18 Centralia soil samples with ICP-MS over three sampling years. Black lines connect samples with continued arsenic increase.

Co-selection: While this approach offers a space-for-time proxy of disturbance, the analysis pipeline can be subsequently applied to future Centralia datasets to permit true time-series analysis of the evolution of resistance genes in the face of a multigenerational disturbance. We have continued collecting Centralia soils and see changes in geochemical data, including a gradual increase in arsenic in two fire-affected soils (**Figure 1**). This initial assessment of resistance genes in Centralia would provide preliminary data for future funding to sequence Centralia metagenomes over time, which would allow more specific characterization of co-selection of resistance genes. Additionally, the final analysis pipeline as well as the manually curated reference databases will be publicly available so that others may examine evolution of resistance genes in different environments, broadening the scope of this work.

Investigation of antibiotic and arsenic resistance gene co-selection over a long-term stressor

Personnel: Ashley Shade, MSU (PI); Tammy Tobin, Susquehanna University (PI); Taylor Dunivin, MSU (Graduate student)

References:

1. O'Neill, J. *Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations*. (2014).
2. Kahn, L. H. *One Health Initiative: Antimicrobial Resistance in the Environment*. (2016).
3. Naujokas, M. F. *et al.* The broad scope of health effects from chronic arsenic exposure: Update on a worldwide public health problem. *Environ. Health Perspect.* **121,** 295–302 (2013).
4. Hiltunen, T., Virta, M. & Laine, A.-L. Antibiotic resistance in the wild : an eco- evolutionary perspective. *Philos. Trans. R. Soc. B* **372,** (2017).
5. Jechalke, S. *et al.* Effects of 100 years wastewater irrigation on resistance genes, class 1 integrons and IncP-1 plasmids in Mexican soil. *Front. Microbiol.* **6,** 1–10 (2015).
6. Seiler, C. & Berendonk, T. U. Heavy metal driven co-selection of antibiotic resistance in soil and water bodies impacted by agriculture and aquaculture. *Front. Microbiol.* **3,** (2012).
7. Pal, C., Bengtsson-palme, J., Kristiansson, E. & Larsson, D. G. J. Co-occurrence of resistance genes to antibiotics , biocides and metals reveals novel insights into their co-selection potential. *BMC Genomics* **16,** (2015).
8. Andres, J. & Bertin, P. N. The microbial genomics of arsenic. *FEMS Microbiol. Rev.* **40,** 299–322 (2016).
9. Aminov, R. I. & Mackie, R. I. Evolution and ecology of antibiotic resistance genes. *FEMS Microbiology Letters* **271,** 147–161 (2007).
10. Jackson, C. R. & Dugas, S. L. Phylogenetic analysis of bacterial and archaeal arsC gene sequences suggests an ancient, common origin for arsenate reductase. *BMC Evol. Biol.* **3,** 18 (2003).
11. van Lis, R., Nitschke, W., Duval, S. & Schoepp-cothenet, B. Arsenic as bioenergetic substrates. *Biochim. Biophys. Acta* **1827,** 176–188 (2013).
12. Lebrun, E. *et al.* Arsenite Oxidase , an Ancient Bioenergetic Enzyme. *Mol. Biol. Evol.* **20,** 686–693 (2003).
13. Zhu, Y., Yoshinaga, M., Zhao, F. & Rosen, B. P. Earth Abides Arsenic Biotransformations. *Annu. Rev. Earth Planet. Sci.* **42,** 443–467 (2014).
14. Gomez, P. & Buckling, A. Real-time microbial adaptive diversification in soil. *Ecol. Lett.* **16,** 650–655 (2013).
15. Aertsen, A. & Michiels, C. W. Diversify or die: generation of diversity in response to stress. *Crit. Rev. Microbiol.* **31,** 69–78 (2005).
16. Aminov, R. I. Horizontal gene exchange in environmental microbiota. *Front. Microbiol.* 1–19 (2011). doi:10.3389/fmicb.2011.00158
17. Williams, H. G., Day, M. J., Fry, J. C. & Stewart, G. J. Natural transformation in river epilithion. *Appl. Environ. Microbiol.* **62,** 2994–2998 (1996).
18. Elick, J. M. Mapping the coal fire at Centralia, Pa using thermal infrared imagery. *Int. J. Coal Geol.* **87,** 197–203 (2011).
19. Pone, J. D. N. *et al.* The spontaneous combustion of coal and its by-products in the Witbank and Sasolburg coalfields of South Africa. *Int. J. Coal Geol.* **72,** 124–140 (2007).
20. Walsh, T. R., Weeks, J., Livermore, D. M. & Toleman, M. A. Dissemination of NDM-1 positive bacteria in the New Delhi environment and its implications for human health: An environmental point prevalence study. *Lancet Infect. Dis.* **11,** 355–362 (2011).
21. Smalla, Kornelia, Jechalke, Sven, Top, E. M. Plasmid detection, characterization and ecology. *Cancer* **121,** 1265–1272 (2015).
22. Tatusov, R. L. *et al.* The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* **29,** 22–28 (2001).
23. Markowitz, V. M. *et al.* IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res.* **42,** 560–567 (2014).
24. Wang, Q. *et al.* Xander : employing a novel method for efficient gene-targeted metagenomic assembly. *Microbiome* **3,** (2015).
25. Fish, J. A. *et al.* FunGene : the functional gene pipeline and repository. *Front. Microbiol.* **4,** 1–14 (2013).
26. Roux, S., Enault, F., Bronner, G. & Debroas, D. Comparison of 16S rRNA and protein-coding genes as molecular markers for assessing microbial diversity (Bacteria and Archaea) in ecosystems. *FEMS Microbiol Ecol* **78,** 617–628 (2011).
27. Barberán, A., Bates, S. T., Casamayor, E. O. & Fierer, N. Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* **6,** 343–351 (2012).
28. Rozov, R. *et al.* Genome analysis Recycler: an algorithm for detecting plasmids from de novo assembly graphs. *Bioinformatics* 1–8 (2016). doi:10.1093/bioinformatics/btw651