Metatranscriptomic expression patterns of biosynthetic genes in soil communities

Alex Crits-Christoph[1,2], Spencer Diamond[3], Cristina Butterfield[3], and Jillian Banfield[2,3].

[1]Department of Plant and Microbial Biology, [2]The Innovative Genomics Institute, [3]Department of Earth and Planetary Science, University of California, Berkeley.

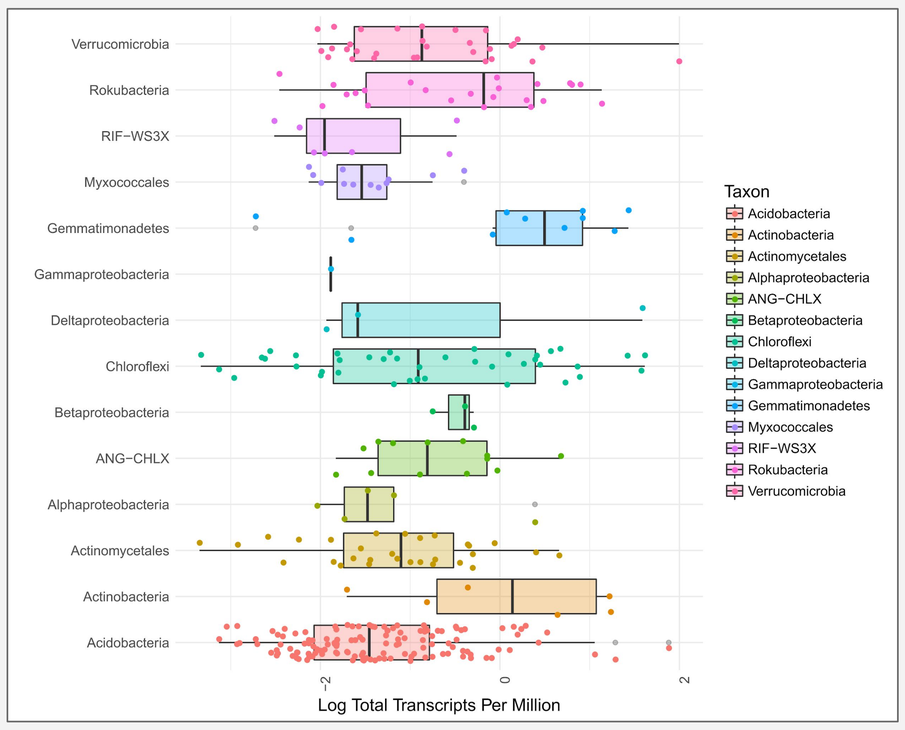
Background: Genomes from soil metagenomes

In a large soil microbiome project, hundreds of high quality draft genomes were previously assembled from soil metagenomes from a Californian meadow. In these genomes, we previously identified biosynthetic gene clusters (BGCs) predicted to synthesize secondary metabolites with antiSMASH. Surprisingly, novel members of the Acidobacteria were found to encode for a large repertoire of secondary metabolites[1].

To investigate secondary metabolite expression, we extracted RNA from 120 soil microcosms subject to glucose, water, and methanol supplementation over 24 hours. **Reads were mapped back to transcripts predicted directly from the genomes obtained from this environment**.

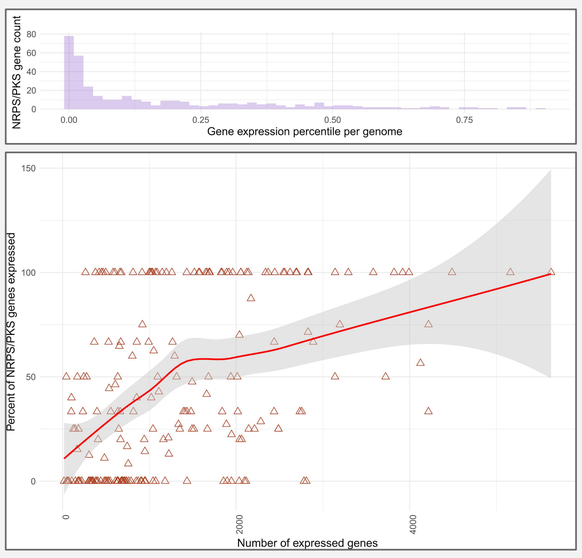
Tracking secondary metabolite gene expression

Right: total expression across 120 soil microcosms observed for NRPS and PKS genes found within BGCs in 896 genomes. Members of the Acidobacteria possessed a plurality of NRPS/PKS genes found in the metagenome and dominated total NRPS/PKS gene expression. Expression of NRPS/PKS proteins was also observed for four organisms with >10 BGCs: **Angelobacter ANG1** (a novel Acidobacteria), **Myxococcus ANG2**, **Chloroflexi ANG3**, and **Micromonospora ANG4** genera.



How highly expressed are NRPS/PKS genes in the environment?

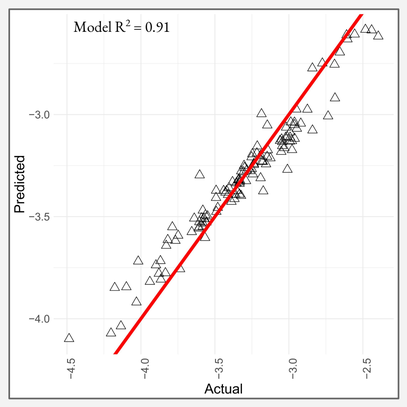
When normalized by gene expression rank within each genome, most NRPS/PKS genes score in the bottom 25% of genes, indicating that they are expressed at relatively low levels. A small proportion of NRPS/PKS genes were expressed in the >50% percentile of genes per genome. As the number of transcripts observed for a genome increased, the likelihood of observing expression for all NRPS/PKS genes approached 1.

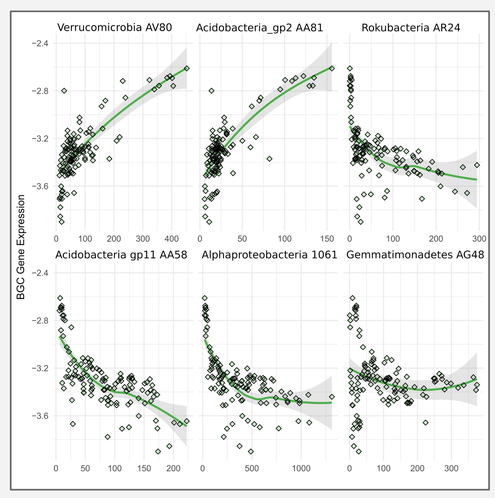


Predicting BGC expression from community structure with machine learning

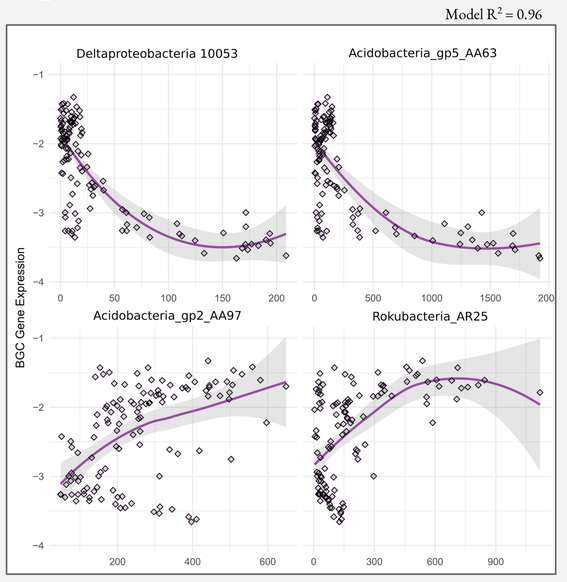
It has been shown in laboratory co-culture experiments that secondary metabolites are often upregulated in response to the presence of another organism through competition sensing[2] (or ecological sensing). If in the natural environment BGCs are upregulated in response to competitors, it would be possible to predict BGC expression using community dynamics.

Random Forest regression models accurately predict the normalized sum expression level of BGCs from 330 features: total activity levels for every other organism in the dataset.





Top: RF Predicted vs Actual values for biosynthetic gene expression from ANG4



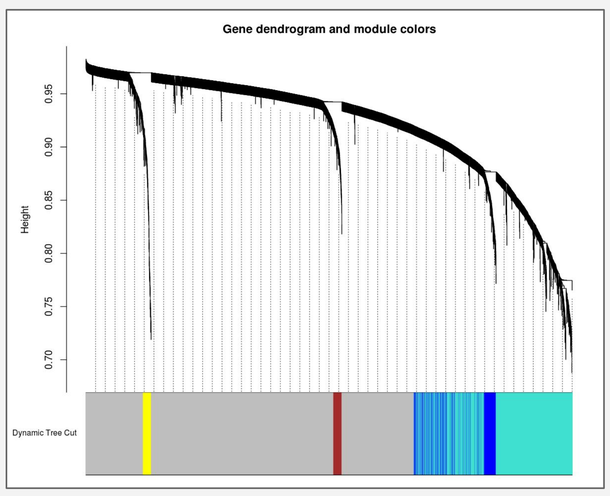
Above and left: Top important features in the RF models for the prediction of BGC expression- these are organisms with general transcriptomic activity that best predicted the expression of biosynthetic genes for **Micromonospora ANG4** (top), **Angelobacter ANG1** (bottom left), and **Chloroflexi ANG3 (**top left).

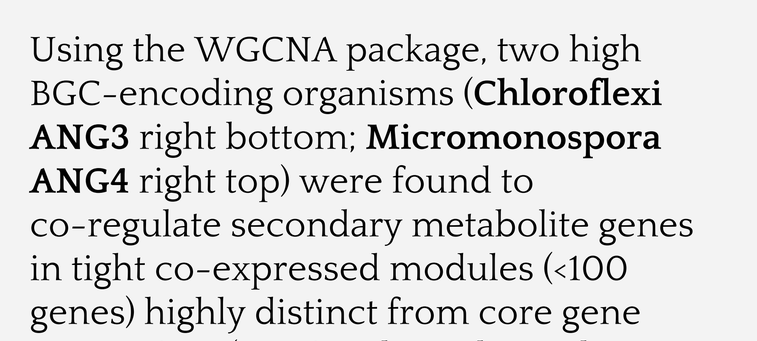
Secondary metabolism gene co-expression networks

Using weighted gene coexpression networks it is possible to dissect correlations in expression data to identify patterns of gene co-regulation.

Using the WGCNA package, two high BGC-encoding organisms (**Chloroflexi ANG3** right bottom; **Micromonospora ANG4** right top) were found to co-regulate secondary metabolite genes in tight co-expressed modules (<100 genes) highly distinct from core gene expression. (Previously we have shown this for **Angelobacter ANG1**).

For these organisms, BGC expression was one of the most distinct signals of co-expression in the genome.





Left: the module of co-expressed genes significantly enriched in NRPS/PKS biosynthetic genes (circled) from **Micromonospora ANG4.** This module included transporters (mmr; MFS transporter), regulators (LuxR; MarR; PadR; TetR), 2-component systems, and biosynthetic genes, including the strB gene in streptomycin biosynthesis.

References and Acknowledgements

1. Crits-Christoph, Alexander, et al. "Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis." Nature (2018) **2**. Cornforth, Daniel M., and Kevin R. Foster. "Competition sensing: the social side of bacterial stress responses." Nature Reviews Microbiology 11.4 (2013): 285.

We thank S. Spaulding for assistance with fieldwork, and M. Traxler and W. Zhang for helpful discussions. Sequencing was carried out under a Community Sequencing Project at the Joint Genome Institute. Funding was provided by the US Department of Energy Grant DOE-SC10010566, the Paul G. Allen Family Foundation and the Innovative Genomics Institute of the University of California, Berkeley.