Reliable bioinformatic prediction of cobamide biosynthesis by core biosynthesis genes and taxonomy

Amanda N. Shelton,[1] Alexa M. Nicolas,[2*] Alekhya Govindaraju,[2] Zachary F. Hallberg,[2] Zoila I. Alvarez-Aponte,[2] Kenny C. Mok,[2] and **Michiko E. Taga**[2]

[*] Alexa_Nicolas@berkeley.edu
[1] *Department of Plant Biology, Carnegie Institution for Science*
[2] *Department of Plant & Microbial Biology, University of California, Berkeley, Berkeley, CA, USA.*

https://tagalab.berkeley.edu

Project Goals: The goal of this project is to gain a comprehensive understanding of microbial metabolic interactions in soil through the study of corrinoids as model metabolites. Corrinoids are a family of structurally diverse cofactors that includes vitamin B_{12}. They are produced by a subset of the bacteria that require them, and thus are shared metabolites. Importantly, bacteria typically can use more than one corrinoid but display distinct preferences for specific structures. For these reasons, we hypothesize that corrinoids are key metabolites that contribute to shaping microbial communities in soil. By using this shared, structurally diverse family of metabolites, we are studying microbial interactions across scales, from the whole community to individual isolates.

The biosynthesis of cobamides, such as example Vitamin B12, is unevenly distributed across bacteria and archaea. However, these essential cofactors are required for diverse cellular reactions in a majority of bacteria that cannot biosynthesize them. Therefore, cobamides, as a category of nutrients, provide an important nutrient to study interactions that underlie complex ecosystems, such as the soil microbiome. Recently, comparative genomic studies suggest that cobamide biosynthesis can be predicted by genome sequence alone. We sought to extend these results to Metagenome-Assembled Genomes (MAGs) and test whether there are phylogenetic trends across taxa that can reliably predict whether an organism can synthesize cobamides *de novo* or is an auxotroph. Previous findings suggest three core cobamide biosynthesis genes may be predictive of biosynthesis capacity. However, how well these predictions can estimate cobamide production *in situ* in MAGs is still unknown. We used CAMISIM to simulate reads for a set of seven complete genomes previously predicted to either encode complete cobamide biosynthesis or not. The use of KOfams for the three core cobamide biosynthesis genes as a proxy for the complete pathway is able to recapitulate cobamide biosynthesis predictions for near-complete MAGs, predicted by the presence of 53 single copy genes.

We sought to determine taxonomic trends in cobamide biosynthesis, as a way to predict potential cobamide sharing networks in environments. Of 11,436 genomes from IMG we examined whether taxonomic patterns existed for their previously categorized cobamide production and dependence. We found that for 59 genera out of 181, 75% or more of the species categorized were not capable of complete cobamide production, suggesting that the inability to produce cobamides may be predicted for a subset of genera. Conversely, for 53 genera 75% or more species studied are predicted to be complete cobamide

producers, suggesting that for this subset of genera cobamide predictions may be extended to a species level. Taken together our findings begin to enable predictions for communities' capacity to produce and use cobamides and generate hypotheses for inter-community interactions centered around cobamides.