

Supporting Information

An updated catalog of diverse type II polyketide synthase biosynthetic gene clusters captured from large-scale nucleotide databases

Christina M. McBride¹, Eric L. Miller², and Louise K. Charkoudian¹

¹ Department of Chemistry, Haverford College, Haverford, PA, USA.

² Department of Biology, Haverford College, Haverford, PA, USA.

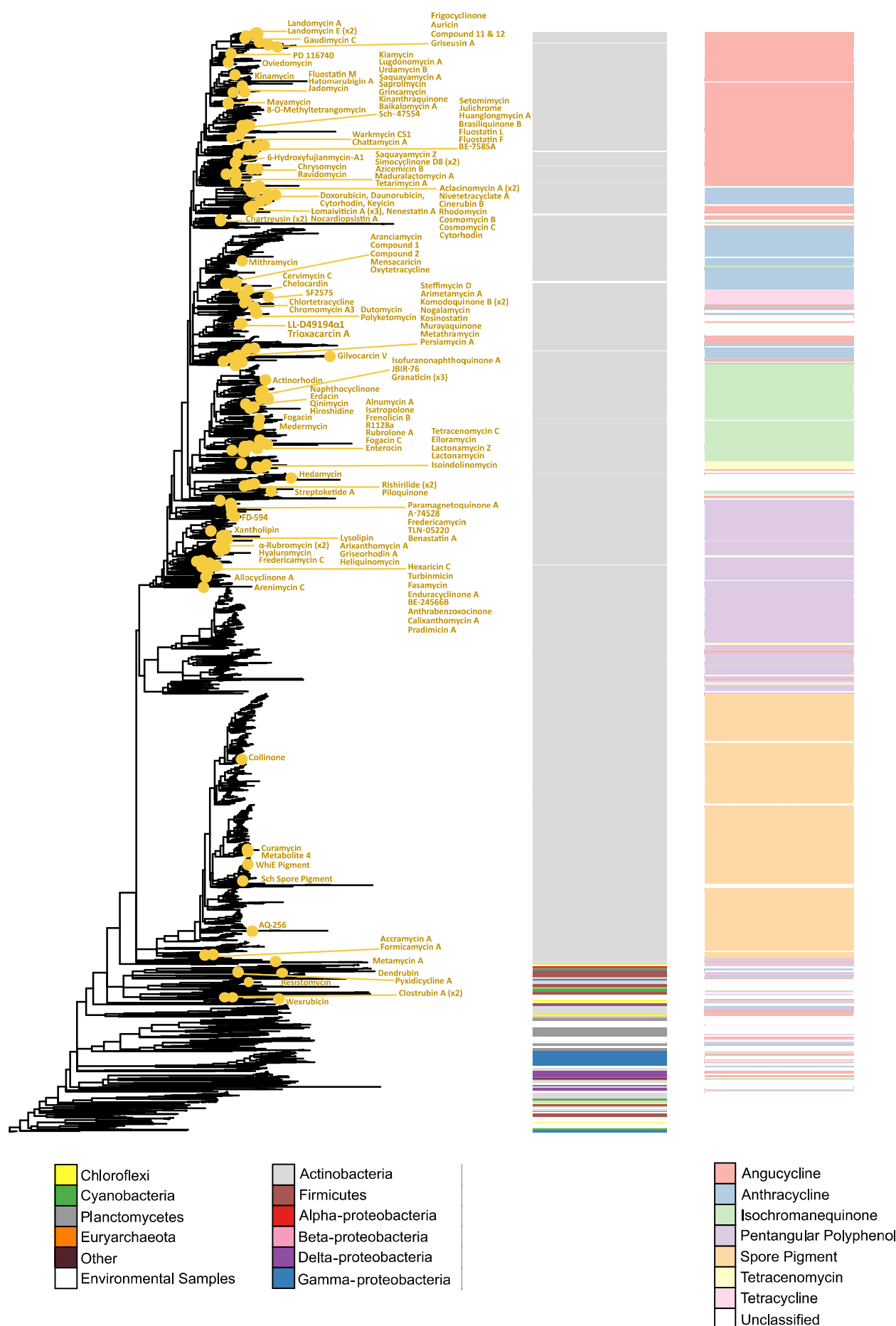


Figure S1. Phylogenetic tree representing 6,322 type II PKS CLF protein sequences. The yellow dots mark the position of CLFs that match a known, characterized CLF sequence as provided by Chen et al (2022). Since nearly all of the 167 characterized CLFs were represented in this tree, we are confident that our workflow can successfully identify CLF diversity across a variety of clades.

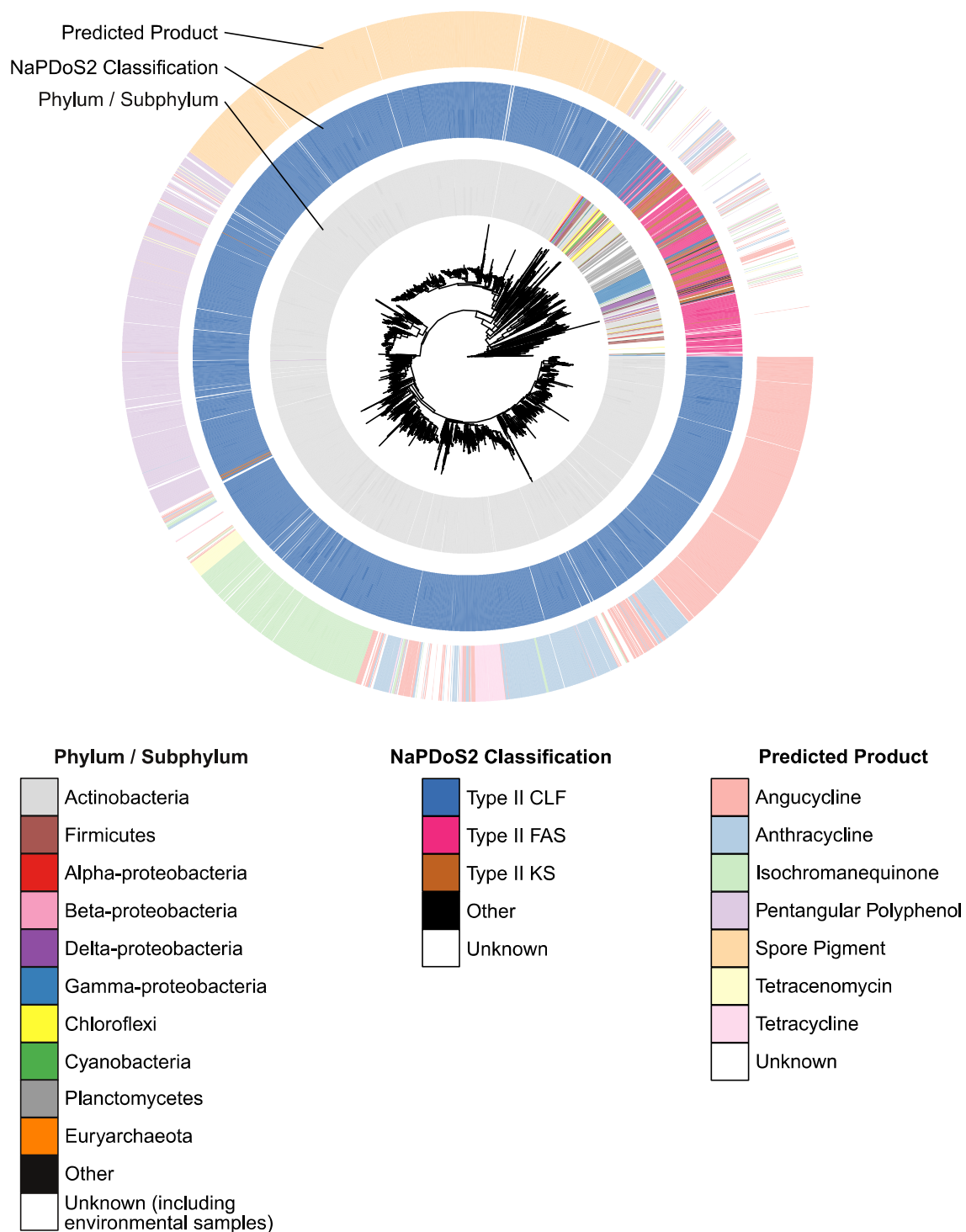


Figure S2. Phylogenetic tree representing 6,322 type II PKS CLF protein sequences. The inner ring represents the taxonomic classification, the middle ring identifies how the sequences were classified by NaPDoS2, and the outer ring represents the predicted molecular product type per NaPDoS2. We observe that the predicted product type is grouped by clade in Actinomycetes, while the non-Actinomycete clade products are largely unexplored.