HypoNPAtlas: an Atlas of hypothetical natural product for mass spectrometry database search

Yi-Yuan Lee[1], Mustafa Guler[1], Neel Mittal[1], Cameron Miller[1], Benjamin Krummenacher[1], Haodong Liu[1], Liu Cao[1*], Aditya Kannan[1], Keshav Narayan[1], Samuel T Slocum[3], Bryan L Roth[3], Alexey Gurevich[4], Roland Kersten[2], Bahar Behsaz[1], and **Hosein Mohimani**[1]

[1]Computational Biology Department, Carnegie Mellon University, Pittsburgh, USA
[2]College of Pharmacy, University of Michigan, Ann Arbor, USA
[3]Department of Pharmacology, University of North Carolina, Chapel Hill, USA
[4]Center for Algorithmic Biotechnology, Institute of Translational Biomedicine, St. Petersburg State University, St. Petersburg, Russia

http://mohimanilab.cbd.cmu.edu/

Project Goals

Recent analysis of hundreds of thousands of public microbial genomes has resulted in the discovery of over a million biosynthetic gene clusters (BGCs). However, the connection of these BGCs to their molecular products has not kept pace with the speed of microbial genome sequencing. Currently, the molecular products for the majority of BGCs remain unknown. Global natural product social (GNPS) molecular networking infrastructure harbors billions of mass spectra of natural products with unknown structures and BGCs. In order to bridge the gap between large scale genome mining and mass spectral datasets for natural product discovery, we developed hypoNPAtlas, an Atlas of hypothetical natural product structures, which can be readily used for *in silico* database search of tandem mass spectra. HypoNPAtlas is constructed by mining the genomes of 22,671 microbial strains from the RefSeq database using seq2ripp, a novel machine learning tool for prediction of ribosomally synthesized and post-translationally modified peptides (RiPPs). Searching the hypothetical molecules from our Atlas against 46 mass spectral datasets from GNPS resulted in the discovery of numerous RiPPs, including two novel lassopeptides and one lanthipeptide from *Streptomyces* sp. NRRL B-2660, WC-3904 and WC-3560. Moreover, seq2ripp discovered two plant RiPPs from *Oryza sativa* and *Elaeagnus pungens* with a novel post-translational modification (PTM) [1].

References

[1] Yi-Yuan Lee, Mustafa Guler, Neel Mittal, Cameron Miller, Benjamin Krummenacher, Haodong Liu, Liu Cao, Aditya Kannan, Keshav Narayan, Samuel T Slocum, Bryan L Roth, Alexey Gurevich, Roland Kersten, Bahar Behsaz, and Hosein Mohimani “HypoNPAtlas: an Atlas of hypothetical natural product for mass spectrometry database search”. Under review.