Integrated metabolome mining and annotation workflow accelerates specialised metabolite discovery

Justin J.J. van der Hooft[1,2], Madeleine Ernst[2], Ricardo da Silva[2], Mingxun Wang[2], Kyo Bin Kang[2,3], Joe Wandy[4], Simon Rogers[5], Marnix H. Medema[1], Pieter C. Dorrestein[2]

[1]Bioinformatics Group, Department of Plant Sciences, Wageningen University, Wageningen, The Netherlands.
[2]Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA, USA.
[3]College of Pharmacy and Research Institute of Pharmaceutical Sciences, Seoul National University, Seoul, Korea.
[4]Glasgow Polyomics, University of Glasgow, Glasgow, United Kingdom.
[5]Department of Computing Science, University of Glasgow, Glasgow, United Kingdom.

E-mail: justin.vanderhooft@wur.nl, mernst@ucsd.edu, ridasilva@ucsd.edu, miw023@ucsd.edu, kbinkang@gmail.com, joe.wandy@glasgow.ac.uk, simon.rogers@glasgow.ac.uk, marnix.medema@wur.nl, pdorrestein@ucsd.edu.

1. Introduction

Microbes and plants produce a gold mine of chemically diverse, high-value molecules like antibiotics. However, chemical structures of many natural products (NPs) remain currently unknown, hampering medicinal applications. A key challenge for natural product discovery is the metabolome complexity in natural extracts, from which mass spectrometry data needs to be coupled to chemical structures. Nevertheless, many NPs share molecular substructures and form structurally related molecular families (MFs), which has inspired metabolome mining tools exploiting these biochemical relationships.

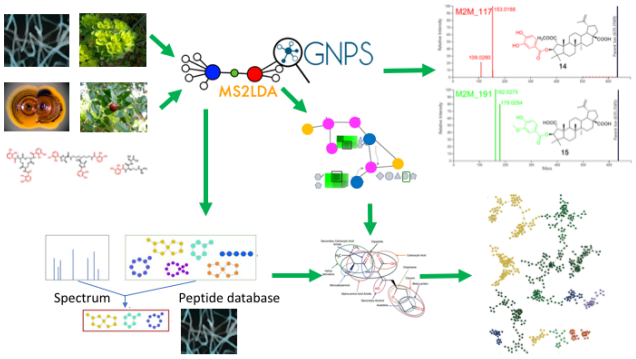

Figure 1. Integrated metabolome miming and annotation leading to enhanced Molecular Networks

2. Approach

Here, we introduce a workflow (Fig. 1) that combines two metabolome mining tools to discover MFs, subfamilies, and subtle structural differences between family members. Where tandem mass spectral Molecular Networking (1) efficiently groups natural products in molecular families, MS2LDA (2) discovers substructures that aid in further recognition of subfamilies and shared modifications.

Furthermore, through the combined use of Network Annotation Propagation (3) and ClassyFire (4), we can automatically perform MF chemical classifications. When unexpected MF classifications are observed, they could represent novel chemical scaffolds, thereby guiding follow-up prioritization efforts towards unknown chemistry.

3. Results

We demonstrate how our integrative workflow discovers dozens of MFs in large-scale metabolomics studies of plant and bacterial extracts. For example, Rhamnaceae plants contained triterpenoid chemistries in which several distinct phenolic acid modifications (e.g., vanillate, protocatechuate) were readily recognized.

Furthermore, a previously not annotated tryptophan-based MF was uncovered in marine Streptomyces extracts. In Photo/Xenorhabdus strains, following leads from peptidic natural products finding software Dereplicator (5), a Xenoamicin-based peptidic MF was deciphered and Mass2Motifs for both the peptidic ring and tail were easily annotated highlighting ring-related modifications.

4. Discussion

Our workflow accelerates NP discovery by MF and substructure annotations and classifications on an unprecedented large scale that will aid in future integration with genome mining workflows. Finally, the workflow applications go beyond the natural products field into nutritional, clinical, and exposome metabolomics.

References

1. Wang, M.. *et al.*, "Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking"., *Nat. Biotech.* 34(8):828-837, 2016.
2. Van der Hooft, J.J.J. *et al.*, "Topic modeling for untargeted substructure exploration in metabolomics". *P.N.A.S.* 113(48):13738-13743, 2016.
3. da Silva, R.R. *et al.*, "Propagating annotations of molecular networks using in silico fragmentation". *PLoS Comp. Biol.* 14(4):e1006089, 2018.
4. Djoumbou Feunang, Y. *et al.* "ClassyFire: automated chemical classification with a comprehensive, computable taxonomy". *J. Cheminformatics* 8(1): 61, 2016.
5. Mohimani, H. *et al.*, "Dereplication of peptidic natural products through database search of mass spectra", *Nat. Chem. Biol.* 13(1):30-37, 2017.