USEDAT: USA-Europe Data Analysis Training School
MOL2NET, International Conference Series on Multidisciplinary Sciences

Dereplication strategy of HRMS data from *Coccoloba cowellii* extract

Daniel Méndez[a], *Julio C. Escalona-Arranz*[b], *Ann Cuypers*[c], *Paul Cos*[d], *Luc Pieters*[e].

[a] *Chemistry Department, Faculty of Applied Sciences, University of Camagüey, Carretera de Circunvalación Km 5 ½, Camagüey 74650, Cuba*

[b] *Pharmacy Department, Faculty of Natural and Exact Sciences, Oriente University, Avenida Patricio Lumumba s/n, Santiago de Cuba 90500, Cuba*

[c] *Centre for Environmental Sciences, Campus Diepenbeek, Hasselt University, Agoralaan Building D, BE-3590, Diepenbeek, Belgium*

[d] *Laboratory of Microbiology, Parasitology and Hygiene (LMPH), Faculty of Pharmaceutical, Biomedical and Veterinary Sciences, University of Antwerp, Universiteitsplein 1, BE-2610, Antwerp, Belgium*

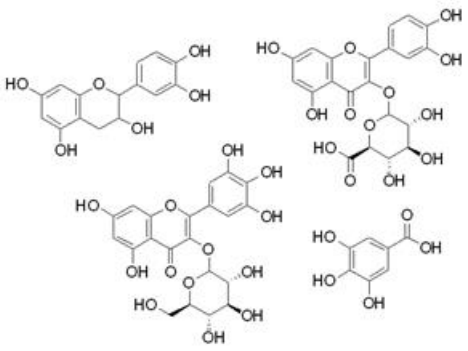
[e] *Natural Products & Food Research and Analysis (NatuRA), Department of Pharmaceutical Sciences, University of Antwerp, Universiteitsplein 1, BE-2610, Antwerp, Belgium.*

Graphical Abstract

Coccoloba cowellii

Dereplication strategy

GNPS



Abstract.

A dereplication strategy of the UHPLC-HRMS data by means of Feature-Based Molecular Networking (FBMN) analysis in the Global Natural Products Social Molecular (GNPS) Network, together with the results obtained from the MS-DIAL and MS-FINDER softwares is described. The strategy allowed the identification of 13 secondary metabolites from *Coccoloba cowellii*, an endemic endangered plant from Cuba, using as little plant material as possible. The followed strategy highlights the capabilities of the integration between several HRMS data processing softwares and GNPS online platform with the goal of a rapid characterization of plant crude extracts.

Abstrac

Introduction

Coccoloba cowellii Britton (Polygonaceae) is an endemic and critically endangered plant that only grows in Camagüey, province of Cuba. A total of 13 compounds were identified in a methanolic leaf extract of this plant, employing a dereplication of the UHPLC-HRMS data. The major constituents

were glucuronides and glycosides of myricetin and quercetin, as well as epichatechin-3-O-gallate, catechin, epicatechin and gallic acid, all of them being reported for the first time in *C. cowellii* leaves [1].

The dereplication strategy used to analyze the raw data employed the capabilities of Feature-Based Molecular Networking (FBMN) analysis in the Global Natural Products Social Molecular (GNPS) Network, together with the results obtained from the MS-DIAL and MS-FINDER softwares. Dereplication provides fast identification of known metabolites in complex biological mixtures using small quantities of material, speeding up the discovery of novel natural products [2]. Feature-based molecular networking (FBMN), available on the Global Natural Products Social (GNPS) Molecular Networking web platform at https://gnps.ucsd.edu, is ideally suited for advanced molecular networking analysis, enabling the characterization of isomers, relative quantification, and the integration of ion mobility data. That is why FBMN is considered the recommended way to analyze single LC-MS^2 metabolomics data [3].

Materials and Methods

UHPLC-HRMS analysis

The analysis of the *Coccoloba cowellii* extract was carried out using a LC-HRMS method according to Bijttebier et al., 2016 [4] and Baldé et al., 2020 [5]. For the HPLC-DAD-QTOF analyses, accurate mass measurements were done using a Xevo G2-XS QTOF spectrometer (Waters, Milford, MA, USA) coupled with an ACQUITY LC system equipped with MassLynx version 4.1 software. Data were recorded using MS^E in the positive and negative ionization modes (two analyses per mode), and a ramp collision energy from 10 to 30 V was applied to obtain additional structural information. Leucine-encephalin was used as the lock mass. UV detection was performed at 360 nm.

Dereplication strategy

The HPLC-MS raw data was converted to abf files (ReifycsAbf Converter) and processed with MS-DIAL version 4.24 [6] for mass signal extraction between 50 and 1500 Da from 0 to 36 min. MS1 and MS2 tolerance was set at 0.01 in centroid mode. The optimized detection threshold was set to 8000 for MS1 and 5000 for MS2. The alignment results were exported using the GNPS export function of MS-DIAL. MS-FINDER software (http://prime.psc.riken.jp/compms/msfinder/main.html) [7] was used for *in silico* fragmentation predictions. Compounds were tentatively identified according to their similarity score, which was based on comparison between experimental MS/MS fragments and *in silico* spectra. To find the potential candidates, only compounds consisting essentially of C, H, and O have been considered (considering no nitrogen containing compounds are reported for the genus *Coccoloba*). The MS1 and MS2 tolerances were set at 0.01 and 0.25 Da, respectively. The natural product databases integrated in MSFINDER (PlantCyc, UNPD, KNApSAcK, and NANPDB) have been used for compound identification. The molecular formulas determined by using MS-FINDER software were then queried in different databases (PubChem (https://pubchem.ncbi.nlm.nih.gov/), ChemSpider (https://www.chemspider.com/), etc.) in order to obtain the molecular structures of the compounds. The following public databases were also consulted: MassBank of North America (MoNA) (http://mona.fiehnlab.ucdavis.edu/) and NIST Mass Spectrometry Data Center (http://chemdata.nist.gov/).

A molecular network was then created with the Feature-Based Molecular Networking (FBMN) workflow [3] on the Global Natural Products Social (GNPS) molecular networking web-platform (https://gnps.ucsd.edu) [8]. The precursor ion mass tolerance was set to 0.05 Da and the MS/MS fragment ion tolerance to 0.05 Da. The network was then created where edges were filtered to have a cosine score above 0.70 and more than 6 matched peaks. Further, edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. The spectra in the network were then searched against GNPS spectral libraries [8,9]. The library spectra were filtered in the same manner as the input data. The data were imported into Cytoscape v3.7.2 (The Cytoscape Consortium, New York, NY, USA) for visualization. The molecular networking job obtained as a result can be publicly accessed at: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=a2f9e6e25ca64043a36a3d2fb09270c5.

References

1. Méndez, D.; Escalona, J.C.; Cuypers, A.; Pieters, L.; Cos, P. UHPLC-HRMS analysis of Coccoloba cowellii, an endemic endangered plant from Cuba., doi:10.3390/mol2net-07-09603.
2. De Oliveira, G.G.; Carnevale Neto, F.; Demarque, D.P.; De Sousa Pereira-Junior, J.A.; Sampaio Peixoto Filho, R.C.; De Melo, S.J.; Da Silva Almeida, J.R.G.; Lopes, J.L.C.; Lopes, N.P. Dereplication of Flavonoid Glycoconjugates from Adenocalymma imperatoris-maximilianii by Untargeted Tandem Mass Spectrometry-Based Molecular Networking. *Planta Med.* **2017**, *83*, 636–646, doi:10.1055/s-0042-118712.
3. Nothias, L.F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst, M.; Tsugawa, H.; Fleischauer, M.; et al. Feature-based molecular networking in the GNPS analysis environment. *Nat. Methods* **2020**, *17*, 905–908, doi:10.1038/s41592-020-0933-6.
4. Bijttebier, S.; Auwera, A. Van Der; Voorspoels, S.; Noten, B.; Hermans, N.; Pieters, L. A First Step in the Quest for the Active Constituents in Filipendula ulmaria (Meadowsweet): Comprehensive Phytochemical Identification by Liquid Chromatography Coupled to Quadrupole-Orbitrap Mass Spectrometry. *Planta Med.* **2016**, 559–572, doi:https://doi.org/10.1055/s-0042-101943.
5. Baldé, M.A.; Tuenter, E.; Matheeussen, A.; Traoré, M.S.; Balde, A.M.; Pieters, L.; Foubert, K. Bioassay-guided isolation of antiplasmodial and antimicrobial constituents from the roots of Terminalia albida. *J. Ethnopharmacol.* **2020**, 113624, doi:10.1016/j.jep.2020.113624.
6. Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; Vandergheynst, J.; Fiehn, O.; Arita, M. MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* **2015**, *12*, 523–526, doi:10.1038/nmeth.3393.
7. Tsugawa, H.; Kind, T.; Nakabayashi, R.; Yukihira, D.; Tanaka, W.; Cajka, T.; Saito, K.; Fiehn, O.; Arita, M. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. *Anal. Chem.* **2016**, *88*, 7946–7958, doi:10.1021/acs.analchem.6b00770.
8. Wang, M.; Carver, J.J.; Phelan, V. V.; Sanchez, L.M.; Garg, N.; Peng, Y.; Nguyen, D.D.; Watrous, J.; Kapono, C.A.; Luzzatto-Knaan, T.; et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* 2016, *34*, 828–837.
9. Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; et al. MassBank: A public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **2010**, *45*, 703–714, doi:10.1002/jms.1777.