SUPPLEMENTARY MATERIAL

UHPLC-HRMS/MS on untargeted metabolomics: a case study with Copaifera (Fabaceae)

Ananda da Silva Antonio, Davi Santos Oliveira, Gustavo Ramalho Cardoso dos Santos, Henrique Marcelo Gualberto Pereira, Larissa Silveira Moreira Wiedemann and Valdir F. Veiga-Junior

Chemistry Department, Institute of Exact Sciences, Federal University of Amazonas, Avenida Rodrigo Octávio, 6200, Coroado, CEP: 69.077-000, Manaus/AM-Brazil

Federal University of Rio de Janeiro, Chemistry Institute, Brazilian Doping Control Laboratory (LBCD – LADETEC / IQ – UFRJ), Avenida Horácio Macedo, 1281 - Pólo de Química - Cidade Universitária, Ilha do Fundão, CEP: 21941-598, Rio de Janeiro/RJ-Brazil.

Chemical Engineering Section, Military Institute of Engineering, Praça General Tibúrcio, 80, Praia Vermelha, Urca, CEP: 22.290-270, Rio de Janeiro/RJ-Brazil. *Email: valdir.veiga@gmail.com

Abstract

Untargeted metabolomics is a powerful tool on chemical fingerprinting. It can be applied in phytochemistry to aid species identification, systematic studies and quality control of bioproducts. This approach aims to produce as much chemical information as possible, without focusing in any specific chemical class, thus, requiring extensive chemometric effort. This study aimed to evaluate the feasibility of a untargeted metabolomic method on phytochemistry by a study case of the Copaifera genus (Fabaceae). This genus contains notorious medicinal species used worldwide. Copaifera exploitation issues includes lack of chemical data, ambiguously species identification methods and absence of quality control for its bioproducts. Different organs of five Copaifera species were analysed by UHPLC-HRMS/MS, GNPS platform and chemometric tools. Untargeted metabolomic enabled the identification of 19 chemical markers and 29 metabolites, distinguishing each sample by species, plant organs, and biome type. Chemical markers were classified as flavonoids, terpenoids and condensed tannins. The applied method provided reliable information about species chemodiversity using fast workflow with little sampling size. Untargeted approach by UHPLC-HRMS/MS proved to be a promising tool for species identification, pharmacological prospecting and in the future for the quality control of extracts used in the manufacture of bioproducts.

Keywords: Copaiba; chemophenetic; Flavonoids; Condensed tannins; GNPS
Table S1. Chromatographic operational features applied on the analysis and peak integration of UHPLC-HRMS method.

| Feature                                      | Value          |
|----------------------------------------------|----------------|
| **Dionex Ultimate 3000 (Thermo Scientific, Bremen, Germany)** |                |
| Oven Temperature (°C)                        | 40             |
| Injection system temperature (°C)            | 40             |
| Flow rate (mL min\(^{-1}\))                 | 0.400          |
| Injection volume (µL)                        | 8.0            |

Table S2. Instrumental features used on the high resolution mass spectrometer

| Feature                                      | Value          |
|----------------------------------------------|----------------|
| **Liquid chromatograph**                    | Dionex Ultimate 3000 (Thermo Scientific, Bremen, Germany) |
| Mass spectrometer                            | Q-Exactive (Thermofisher Scientific, Bremen, Germany) |
| Detector                                     | Hybrid quadrupole-orbitrap mass |
| Ionization source                            | ESI            |
| Acquisition mode                             | MS\(^1\) and MS\(^2\) Data Dependent analysis (DDA) with “TopN” algorithm set to 5 precursors for MS\(^2\) experiments |
| m/z range (Da)                               | 100 – 900      |
| Ionization modes                             | Positive and negative |
| Spray voltage (kV)                           | 2.9 (negative mode) 3.9 (positive mode) |
| Sheath gas flow (arb)                        | 60             |
| Sweep gas flow (arb)                         | 0.0            |
| Auxiliary gas flow (arb)                     | 20             |
| Capillary Temperature (°C)                   | 380            |
| Source temperature (°C)                      | 250            |
| Collision energy (eV)                        | 30             |
| Resolution (FHWM)                            | 70,000         |
| Mass error (ppm)                             | 5.0            |
| Isolation window                             | 4.0            |
| Mass spectrometer calibration                | Xcalibur v3.0.63 Lock mass as “best” for polydimethylcuclosiloxane |

Table S3. Chromatographic operational features applied on the analysis and peak integration of UHPLC-HRMS method.

**Baseline correction**

| Feature                                      | Value          |
|----------------------------------------------|----------------|
| Chromatogram type                            | TIC            |
| MS level                                     | 1              |
| M/z bin width                                | 1              |
| Correction method                            | Asymmetric baseline corrector |
| Smoothing                                    | 1.00E+05       |
| Asymmetry                                    | 0.05           |

**Mass detection**

| Feature                                      | Value          |
|----------------------------------------------|----------------|
| Mass detector                                | Exact mass     |
| MS level                                     | 1              |
| Noise level                                  | 1.00E+07       |
| MS level                                     | 2              |
| Noise level                                  | 8.00E+03       |

**ADAP chromatogram builder**

| Feature                                      | Value          |
|----------------------------------------------|----------------|
| Min group size of scan                       | 5              |
| Group intensity threshold                    | 3.00E+07       |
| Min highest intensity                        | 7.00E+06       |
### M/z tolerance (ppm)
- 10

### Chromatogram deconvolution

| Algorithm                  | Baseline cut-off |
|----------------------------|-------------------|
| Min. Peak height           | 7.00E+06          |
| Peak duration (min)        | 0 to 3            |
| Baseline level             | 8.00E+05          |
| M/z range for MS2 scan pairing (Da) | 0.01 |
| M/z range for MS2 scan pairing (min) | 0.2 |
| M/z center calculation     | Average           |

### Deisotope

| M/z tolerance (ppm)       | 10                  |
| Retention time tolerance (%) | 0.1               |
| Maximum charge            | 3                   |
| Representative isotope    | Most intense        |

### Join aligner

| M/z tolerance (ppm)       | 10                  |
| Weight for m/z            | 75                  |
| Retention time tolerance min. | 0.1             |
| Weight for RT             | 25                  |

### Filtering

| Min. Peak in a row        | 2                   |
| Min. Peak in na isotope pattern | 2              |
| Keep only with MS2 scan (GNPS) | Yes             |

### GNPS-FNMB

| Precursor ion mass tolerante (Da) | 2 |
| Fragment ion mass tolerante (Da)  | 0.5 |