



## Implementation of a MS/MS database for isoquinoline alkaloids and other annonaceous metabolites

Salemon Akpa Agnès, Timothée Okpekon, Yvette Affoué Kouadio, Adrien Jagora, Dimitri Bréard, Emmanoel Costa, Felipe da Silva, Hector Koolen, Anne-Marie Le Ray, Pascal Richomme, et al.

### ► To cite this version:

Salemon Akpa Agnès, Timothée Okpekon, Yvette Affoué Kouadio, Adrien Jagora, Dimitri Bréard, et al.. Implementation of a MS/MS database for isoquinoline alkaloids and other annonaceous metabolites. *Scientific Data* , 2022, 9 (270), pp.1-8. 10.1038/s41597-022-01345-y . hal-03783707

HAL Id: hal-03783707

<https://univ-angers.hal.science/hal-03783707>

Submitted on 22 Sep 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# scientific data



OPEN

DATA DESCRIPTOR

## Implementation of a MS/MS database for isoquinoline alkaloids and other annonaceous metabolites

Salemon Akpa Agnès<sup>1,2</sup>, Timothée Okpekon<sup>2</sup>, Yvette Affoué Kouadio<sup>1,2</sup>, Adrien Jagora<sup>1</sup>, Dimitri Bréard<sup>1</sup>, Emmanoel V. Costa<sup>4</sup>, Felipe M. A. da Silva<sup>5</sup>, Hector H. F. Koolen<sup>1</sup>, Anne-Marie Le Ray-Richomme<sup>3</sup>, Pascal Richomme<sup>3</sup>, Pierre Champy<sup>1</sup>, Mehdi A. Beniddir<sup>1</sup> & Pierre Le Pogam<sup>1</sup>✉

This data descriptor reports on the upload to a public repository (GNPS) of the IQAMDB, IsoQuinoline and Annonaceous Metabolites Data Base, comprising 320 tandem mass spectra. This project originated from our in-house collection of isoquinolines. The diversity of compounds included in this database was further extended through the contribution of two additional laboratories involved in isoquinoline alkaloids research: University of Angers and University of Manaus. The generated MS/MS data were processed and annotated on an individual basis to promote their straightforward reuse by natural product chemists interested in either the description of new isoquinoline alkaloids or the dereplication of isoquinoline-containing samples. The interest of the current repertoire for dereplication purposes has been validated based on the molecular networking of the well-investigated plant model *Annona montana* against the IQAMDB-implemented GNPS.

### Background & Summary

As one of the largest groups of plant alkaloids, isoquinolines include a significant number of well-known drugs and lead-compounds, such as morphine, codeine, noscapine, papaverine, D-tubocurarine, berberine, emetine or higenamine. These illustrious leads tantalized many laboratories to explore the phytochemistry of isoquinoline-producing plants, so that more than 3000 BenzylIsoQuinolines (BIQ) alkaloids are currently catalogued<sup>1</sup>, displaying varied and often significant bioactivities<sup>2</sup>. From a phylogenetic standpoint, (tetrahydro)benzylisoquinoline alkaloids are reported to occur in more than 40 different families with the most represented producers being Amaryllidaceae, Annonaceae, Berberidaceae, Fumariaceae, Hernandiaceae, Lauraceae, Menispermaceae, and Papaveraceae<sup>3</sup>. In the quest for alternative industrial processes of natural product supplies, a few isoquinoline alkaloids of outstanding interest succumbed to biotechnological manufacturing, such as morphine<sup>4</sup> (and codeine as its biosynthetic intermediate) or noscapine<sup>5</sup>. The very last years witnessed tremendous advances in the metabolic engineering of BIQ with the development of optimized yeast strains synthesizing a considerable amount of (S)-reticuline, a biosynthetic pillar in the TetraHydroIsoQuinoline (THIQ) series, but also a diversity of unprecedented BIQ derivatives<sup>1</sup>. These advances in THIQ metabolite production guided Courdavault *et al.* to identify this family of compounds as holding “an eminent biosynthetic potential in the field of drug discovery”<sup>6</sup>.

<sup>1</sup>Équipe “Chimie des Substances Naturelles” Université Paris-Saclay, CNRS, BioCIS, 5 rue J.-B. Clément, 92290, Châtenay-Malabry, France. <sup>2</sup>Laboratoire de Constitution et Réaction de la Matière (LCRM), UFR Sciences des Structures de la Matière et Technologie, Université Félix Houphouët-Boigny, BP 582, Abidjan 22, Côte d’Ivoire.

<sup>3</sup>Univ Angers, SONAS, SFR QUASAV, F-49000, Angers, France. <sup>4</sup>Departamento de Química, Universidade Federal do Amazonas, Av. Rodrigo Otávio 1200, 69067-005, Manaus, AM, Brazil. <sup>5</sup>Centro de Apoio Multidisciplinar (CAM), Universidade Federal do Amazonas, Av. Rodrigo Otávio 1200, 69067-005, Manaus, Brazil. <sup>6</sup>Grupo de Pesquisa em Metabolômica e Espectrometria de Massas, Universidade do Estado do Amazonas, Av. Carvalho Leal 1777, 69065-001, Manaus, Brazil. ✉e-mail: pierre.le-pogam-alluard@universite-paris-saclay.fr

Our continuous efforts towards the improvement of the molecular-networking pipeline efficiency through the upload of MS/MS data related with collections of structurally homogeneous natural products led us to implement two different databases today incorporated into the Global Natural Products Social Molecular Networking (GNPS) libraries<sup>7</sup>. A first contribution to these public repositories was dedicated to Monoterpene Indole Alkaloids<sup>8</sup> (so-called MIADB), implying an international consortium of eight different natural product laboratories that resulted in a collection of 172 initial entries, now reaching more than 220 compounds. Our second initiative in the field, the LDB (Lichen Data Base), garnered 241 structurally diverse lichen phenolic compounds based on a collaborative project with University of Rennes 1 and the Berlin Garden and Botanical Museum<sup>9</sup>. We foresaw the possibility of building and deploying a third MS/MS database on the basis of the large diversity of isoquinoline alkaloids that have been isolated in Paris-Saclay<sup>10</sup>.

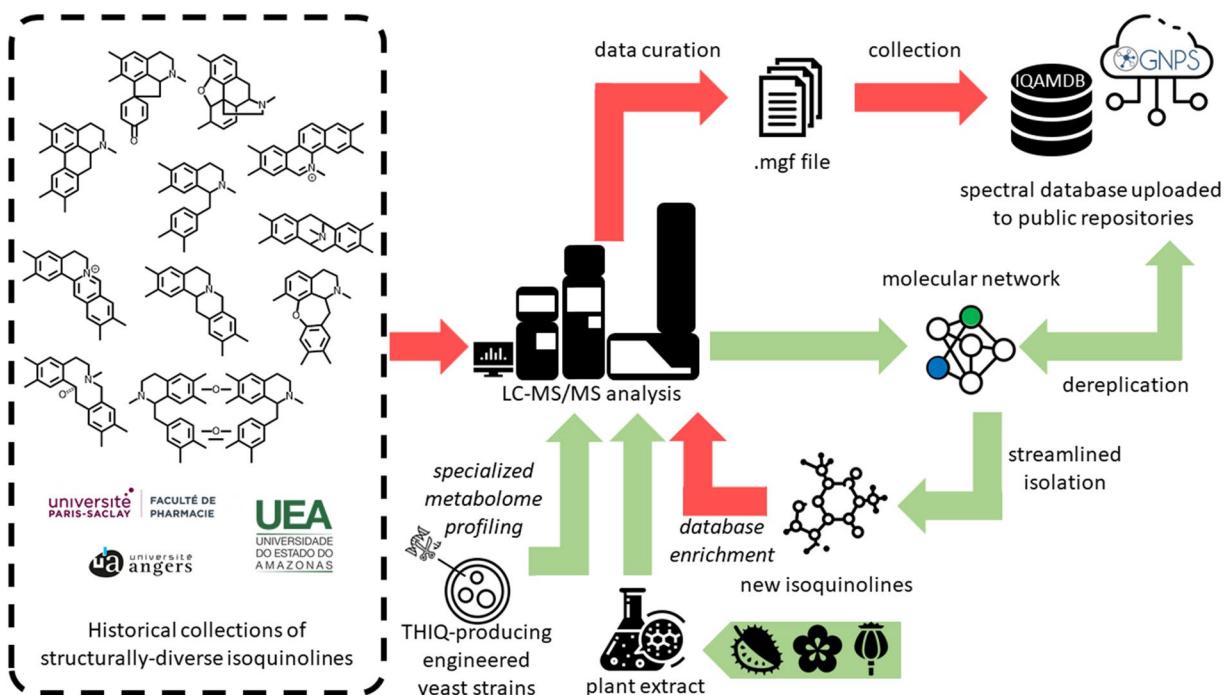
Described by Jussieu in 1789, the Annonaceae family comprises more than 120 genera and about 2100 species, all occurring in tropical and subtropical regions<sup>11</sup>. From a phytochemical standpoint, the annonaceous species have been vastly investigated, with a pronounced emphasis on their alkaloidal components<sup>12</sup>. Heretofore, more than 800 alkaloids have been reported from Annonaceous source<sup>12</sup>. Annonaceous alkaloids are vastly dominated by isoquinolines, mostly falling into benzyl- and bisbenzyltetrahydroisoquinolines, protoberberines, tetrahydroprotoberberines, proaporphines, aporphines, 7-substituted aporphines, oxoaporphines and phenanthrenes, all deriving from benzylisoquinoline precursors. A few non-isoquinoline alkaloids extend the diversity of Annonaceous metabolites through unusual structures, such as canagine, a napthyridine alkaloid<sup>13</sup>, original pyrimidine β-carboline structures referred to as annomontines<sup>14,15</sup>, a few prenylindoles and some indolosesquiterpenes<sup>16</sup>. It seems that the discovery rate of new chemical entities from Annonaceous sources decreases across decades as most of the recently published investigations in the field report known metabolites<sup>17,18</sup>. Although now uncommon, some original structures are still being unearthed from these plants, such as a few aristolactam alkaloids recently obtained from the thailandese *Dasymaschalon dasymaschalon* (Blume) I.M. Turner<sup>19</sup> and new isoquinolines, including an unprecedented 8-oxohomoaporphines from the Amazonian *Duguetia surinamensis* R. E. Fries<sup>20</sup>. It is tempting to hypothesize that the quest for new structures from these plants is made difficult by the occurrence of important amounts of recurrent structures in Annonaceous plants that are hard to obviate in the course of untargeted phytochemical investigations. In such context, molecular networking strategies have proved useful to select natural material deserving deeper chemical studies as well as to streamline the isolation of compounds of interest through a rational, hypothesis-driven workflow<sup>21</sup>. This provides new opportunities to focus on minor, interesting compounds from deeply-dug plant material<sup>22,23</sup>. Nevertheless, the success of this dereplicative strategy primarily depends on the availability of reference tandem mass spectra uploaded to the GNPS spectral libraries<sup>7</sup>. Regarding the specific example of Annonaceous plants, some molecular networking-based investigations undertaken by some of us, involving an in-house collection of alkaloid standards, efficiently assisted the targeting of new compounds for purification and subsequent structure elucidation<sup>20</sup>. These works gratifyingly outlined the trend of Annonaceous alkaloids to clusterize in a scaffold-dependent manner. Some fragmentation patterns were even proposed by the same authors<sup>24,25</sup>, leading to extend the annotation provided by this dereplication to untagged nodes based on the cursory examination of MS/MS spectra<sup>20</sup>.

We felt that our in-house collection of isoquinolines could set the ground for building a MS/MS database dedicated to Annonaceous metabolites. The core constitution of this Annonaceous and isoquinoline database was enlarged by inviting prominent natural products researchers in the field to share their collections with us. This collaborative work took advantage of additional Annonaceous isoquinolines included in the collections of Federal University of Amazonas and Amazonas State University (Manaus, Brazil). To broaden the applicability of our database, this combined collection was further extended to structurally-related isoquinolines from non-Annonaceous plant sources through the incorporation of compounds obtained from University of Angers (France) which comprised isolates from Menispermaceous, Hernandiaceous, Lauraceous and Papaveraceous species<sup>26–28</sup>. These contributions led us to constitute the largest MS/MS dataset of isoquinolines to date, namely the IQAMDB (IsoQuinolines and Annonaceous Metabolites DataBase) (<https://gnps.ucsd.edu/ProteoSAFe/gnpslibrary.jsp?library=IQAMDB>) (Fig. 1 and Table S1)<sup>29</sup>. This database should help pinpointing new chemical entities from isoquinoline-producing plants. We also hope that the upload of this database will be of interest to increase the reliability of the dereplication workflow in the frame of the very dynamic metabolic engineering efforts currently devoted to THIQ alkaloids. This data descriptor reports on the deposition of IQAMDB in the GNPS libraries and on its subsequent technical validation.

## Methods

**Sample preparation.** The different standard samples were dissolved in HPLC-MS grade methanol at 0.5 mg/mL and placed into 1500 μL vials for analysis. For the alkaloidal extract, 1 g of dried and milled *Annona montana* Macfad. root bark was extracted using 50 mL of 0.25 M H<sub>2</sub>SO<sub>4</sub> for one hour. The phase was alkalinized to a pH of 12 and counter-extracted three times using 35 mL of methylene chloride. The combined organic phases were evaporated in vacuo and dissolved in analytical grade methanol at a concentration of 0.5 mg/mL for UPLC/HRMS<sup>2</sup> analysis. Solvents were purchased from Sigma-Aldrich.

**Data acquisition.** Samples were analyzed using an Agilent 6546 Accurate-Mass Q-TOF hyphenated with a 1290 Agilent Infinity II LC system. The chromatographic system was fitted with a Zorbax RRHD Eclipse Plus C18 column (2.1 × 50 mm, 1.8 μm). Elution solvents used were Milli-Q water + 0.1% formic acid (A) and acetonitrile + 0.1% formic acid (B), eluted using the following gradient: 5% B at 0 min; 100% B at 8 min, linear gradient; 8–12 min 100% B; 12–16 min 5% B. The flow rate was 500 μL·min<sup>-1</sup>. Mass spectrometry settings were as follows: capillary temperature at 320 °C, source voltage at 3.5 kV, and a sheath gas flow rate at 10 L/min. The mass spectrometer operated in positive polarity. Mass spectrometric acquisitions were divided into four scan events: positive MS with a window from *m/z* 100 to 1200, followed by three data-dependent MS/MS scans of the



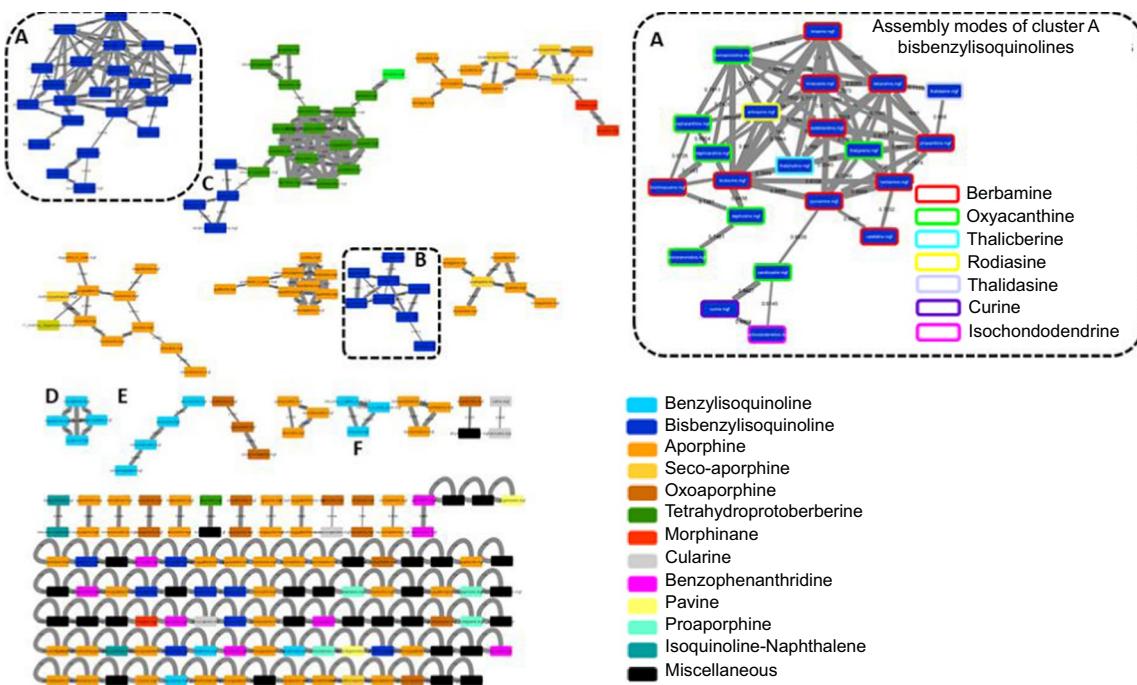
**Fig. 1** IQAMDB implementation workflow (red path) and pipeline for its subsequent use in the frame of a molecular networking-based dereplication strategy (green arrows).

three most intense ions detected through the first scan event. Tandem mass spectrometric parameters were set as follows: collision energy = 50 eV, default charge of 1, isolation width of  $m/z$  1.3. Dynamic exclusion was disabled. Purine ( $C_5H_4N_4$ ,  $m/z$  121.050873), and HP-0921 (hexakis(1*H*, 1*H*, 3*H*-tetrafluoropropoxy)-phosphazene  $C_{18}H_{18}F_{24}N_3O_6P_3$ ,  $m/z$  922.009798 were used as internal lock masses. Full scans were acquired at a resolution of 60.000 ( $m/z$  922) and 35.000 ( $m/z$  121).

**Database constitution.** The analysis of all these substances resulted in 320 files with the standard Agilent.d format. The list of features detected in every sample was generated following the auto MS/MS data mining process implemented in MassHunter software on every single file. Within the list of detected features, the exact mass of the feature of interest was identified and the other features were filtered out. The MS/MS data related to the signal of interest were subsequently converted into a .mgf file using a tailored intensity threshold thanks to the dedicated “Export” option of the MassHunter software.

**Molecular networking parameters (using IQAMDB standards as an input).** Every single MS/MS spectrum related to the IQAMDB standards was curated as indicated in the Database constitution section. A molecular network was then created using all the .mgf files as an input, using the online molecular networking workflow (version release\_28.2) at GNPS<sup>7</sup> (<http://gnps.ucsd.edu>) with a parent mass tolerance of 0.02 Da and a MS/MS fragment ion tolerance of 0.02 Da. The data were not clustered with MS-Cluster. A network was then created where edges were filtered to have a cosine score above 0.6 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. All matches kept between network spectra and library spectra were required to have a score above 0.6 and at least 6 matched peaks. The molecular networking data were analyzed and visualized using Cytoscape (ver. 3.6.0)<sup>30</sup>.

**Feature-based molecular networking parameters (dereplication of *A. montana*).** The MS2 data files related to the alkaloidal extract of *Annona montana* were converted from the .d (Agilent) standard data-format to .mzXML format using the MSConvert software, part of the ProteoWizard package<sup>31</sup>. The .mzXML file was further processed using MZmine 2 v53<sup>32</sup>. Mass detections were performed using a noise level threshold at 2E3 in MS1 and at 1.5E1 in MS2. The ADAP chromatogram builder used a minimum group size of scans of 2, a group intensity threshold of 2E3, a minimum highest intensity of 2E3 and a  $m/z$  tolerance of 10 ppm<sup>33</sup>. The chromatogram deconvolution used the Local Minimum Search algorithm with the following settings: chromatographic threshold = 1%, search minimum in RT range (min) = 0.05, minimum relative height = 2%, minimum absolute height = 1E3, min ratio of peak top/edge = 0.9, peak duration range (min) = 0.00–1. MS2 scans were paired using a  $m/z$  tolerance range of 0.02 Da and RT tolerance range of 0.15 min. Isotopes were grouped using the isotopic peaks grouper algorithm with a  $m/z$  tolerance of 10 ppm and a RT tolerance of 0.5 min keeping the most intense peak as the representative isotope. The peak list was filtered to keep only rows with MS2 features. The .mgf and .csv files were exported using the MZmine2 built-in “Export/Submit to GNPS /FBMN” option. The molecular network was finally created using the online FBMN workflow (version release\_28.2) at GNPS



**Fig. 2** Scaffold-annotated molecular network using the MS/MS spectra of the IQAMDB as an input with a cosine similarity score cut-off of 0.6.

(<http://gnps.ucsd.edu>) with a parent mass tolerance of 0.02 Da and a MS/MS fragment ion tolerance of 0.02 Da, where edges were filtered to have a cosine score above 0.6 and more than 6 matched peaks. Further edges between two nodes were maintained 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. Matches between network spectra and library spectra were required to have a cosine score above 0.6 and at least 6 matched peaks. The molecular networking data were analyzed and visualized using Cytoscape (ver. 3.6.0)<sup>30</sup>. The obtained molecular network can be accessed at: <https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=3ff5f1048c7948ff96627438ac905acd>.

## Data Records

Data reported in this article have been uploaded to the GNPS platform. Each MS<sup>2</sup> spectrum of the 320 compounds is assigned an individual accession number on the [GNPS](#). The .mgf files were deposited and are publicly available at the MassIVE repository (MSV000088909) (<https://doi.org/10.25345/C55D8NG3T>)<sup>34</sup>. The spectral collection is available for download from the library webpage of the GNPS (<https://gnps.ucsd.edu/ProteoSAFe/libraries.jsp>)<sup>29</sup>.

## Technical Validation

**Spectroscopic validation of IQAMDB compounds.** The structure elucidation of the alkaloids implemented in the IQAMDB relied on an extensive set of spectroscopic techniques comprising at least nuclear magnetic resonance spectroscopy and high-resolution mass spectrometry. Further analyses were carried out whenever needed to ensure unambiguous structure assignment. These analyses were performed in the laboratory where the product had been isolated. The manual curation of each mass spectrometric file revealed the expected elemental composition, confirming that the samples had not degraded since they were isolated.

**Retained strategies for IQAMDB validation.** The validation of the IQAMDB repertoire was achieved following two distinct strategies. At first, the topology of the molecular network obtained using the IQAMDB as an input was inspected to assess whether the uploaded MS/MS data could outline structural similarities between the standards included in the database, as a first indicator of the quality of IQAMDB spectrometric data. Then, the dereplication efficiency of the IQAMDB-implemented GNPS libraries was estimated by annotating the molecular network obtained from the alkaloidal extract of the thoroughly studied *Annona montana* Macfad.

**Topology of the molecular network obtained using the IQAMDB as an input.** The molecular network generated using the tandem mass spectra of the IQAMDB as an input is disclosed in Fig. 2, with its nodes being colored according to their phytochemical class. Generally speaking, the topology of the molecular network does not deserve to be too thoroughly studied as it vastly depends on a wealth of different parameters, ensuring a low degree of reproducibility (mass spectrometric acquisition settings, molecular networking parameters, and even the diversity of compounds included in a molecular network that could instigate edges between some nodes or not). While having in mind these limits, we nevertheless felt relevant to outline a few clustering trends which may be of use to strengthen putative identifications against the IQAMDB-implemented GNPS. Chemical structures reported below to illustrate some rational clustering trends are provided in Figure S1 (Supporting Information).

Compound	Cosine	Comment
Annomantine	0.95	Described from <i>A. montana</i> <sup>14</sup>
Anomuricine	0.75	Described from other <i>Annona</i> species <sup>53</sup>
Anomurine	0.86	Described from other <i>Annona</i> species <sup>53</sup>
Argentinine	0.95	Described from <i>A. montana</i> <sup>39</sup>
Atherosperminine	0.97	Described from <i>A. montana</i> <sup>37</sup>
Atherosperminine-N-oxide	0.97	Known from other Annonaceous genera <sup>17</sup>
Bisnorargemonine	0.86	Known from other Annonaceous genera <sup>54</sup>
Coreximine	0.94	Described from <i>A. montana</i> <sup>37</sup>
Discoguattine	0.72	Known from other Annonaceous genera <sup>17</sup>
Glaziovine	0.64	Described from other <i>Annona</i> species <sup>42</sup>
Kikemanine	0.63	Described from other <i>Annona</i> species <sup>44</sup>
Liriodenine	0.96	Described from <i>A. montana</i> <sup>37</sup>
Methoxyannomantine	0.88	Described from <i>A. montana</i> <sup>14</sup>
N-methylcoclaurine	0.91	Described from other <i>Annona</i> species <sup>40</sup>
Nandigerine	0.72	Not known from Annonaceous source
Noratherosperminine	0.93	Known from other Annonaceous genera <sup>55</sup>
Norjuziphine	0.90	Known from other Annonaceous genera <sup>56</sup>
Nornuciferine	0.77	Described from other <i>Annona</i> species <sup>47</sup>
Obovanine	0.92	Described from other <i>Annona</i> species <sup>48</sup>
O-methylisopiline	0.64	Known from other Annonaceous genera <sup>57</sup>
Pessoine	0.80	Described from other <i>Annona</i> species <sup>45</sup>
Reticuline	0.88	Described from <i>A. montana</i> <sup>37</sup>
Reticuline N-oxide	0.88	Described from other <i>Annona</i> species <sup>41</sup>
Roemerine	0.95	Described from other <i>Annona</i> species <sup>49,50</sup>
Stepharine	0.85	Described from other <i>Annona</i> species <sup>43</sup>
Stepholidine	0.81	Described from other <i>Annona</i> species <sup>46</sup>
Tembatarine	0.74	Described from other <i>Annona</i> species <sup>41</sup>

**Table 1.** Matches between the profiled compounds from an alkaloidic extract of *A. montana* root bark and IQAMDB.

The disclosed molecular network tended to cluster according to compounds structural class. The most striking examples of scaffold-based clustering are represented by tetrahydroprotoberberines (green nodes), benzylisoquinolines (light blue nodes) and bisbenzylisoquinolines (navy blue nodes)<sup>35</sup>. However, each of these scaffolds did not result in a unique and homogeneous cluster in the retained settings, sometimes conveying additional pieces of structural information. An example of rational structure-dependent clustering is that of bisbenzylisoquinolines which were split into three main clusters (A, B and C). As the largest group of isoquinoline alkaloids, bisbenzylisoquinolines are usually being classified based on their assembly modes, so that 26 different subtypes had been identified by Shamma and Moniot<sup>36</sup>. The benzylisoquinoline building blocks can indeed be joined by one or more ether bridges, sometimes accompanied by carbon-to-carbon biphenyl linkages or methyleneoxy junctions. Cluster A, exclusively, consisted of bisbenzylisoquinolines featuring two connections between their benzylisoquinoline components. Most nodes of this cluster corresponded to bisbenzylisoquinolines featuring a head-to-head ether connection and a tail-to-tail ether linkage between their monomeric units (berbamine, oxyacanthine, thalicerine and thalidamine subtypes). The rodiasine-type antioquine, featuring a head-to-head ether connection and a carbon-to-carbon biphenyl linkage as a tail-to-tail junction, connected with many representatives of the aforementioned subtypes. Notably, the southernmost part of cluster A comprised curine and isochondodendrine representatives, both featuring two head-to-tail ether connectivities. Cluster B connected bisbenzylisoquinolines featuring three intermonomeric connections. Most of the compounds fit the so-called trilobine-isotrilobine-micranthine type consisting of two ether bridges as head-to-head connections and an additional ether tail-to-tail linkage. The westernmost part of this cluster comprised tiliacorine and tiliacorinine disclosing two head-to-head ether connections and a biphenyl carbon-to-carbon bond tethering their tails. At last, four representatives loosely attached to the tetrahydroprotoberberine cluster (sub-cluster C). These molecules only featured one connection between their isoquinolines building blocks as a tail-to-tail junction (*viz.* daurisoline subtype). Likewise, the cursory examination of the clustering behaviour of other isoquinoline series revealed some tendencies which were seemingly related to some sharp structural features. Cluster D was composed of benzylisoquinolines substituted with a OH group on their benzyl component and by two oxygenated functionalities on their isoquinoline part. Benzylisoquinolines containing more than two methoxy groups on their isoquinoline component and a further methoxy functionality on their benzyl ring were detected in cluster E. At last, cluster F only collated structures disclosing a reticuline-type substitution pattern (a methoxy and a phenolic function on both isoquinoline and benzyl parts).

*Validation of the IQAMDB against the phytochemically-defined *Annona montana* alkaloidic extract.* The validation of the IQAMDB repertoire was based on the dereplication of *Annona montana*, commonly known as the mountain soursop, as a deeply-dug Annonaceous plant model. An alkaloidic extract was prepared from the historical sample of *A. montana* root bark, which had been phytochemically investigated by Lebœuf and co-workers in our laboratory in the 1980s<sup>14,37</sup> and was further analyzed in positive polarity by UPLC-HRMS<sup>2</sup>. To capture the chemical diversity of this extract, the UPLC–HRMS<sup>2</sup> data were processed using the feature-based molecular networking workflow<sup>38</sup>, and subsequently dereplicated against the IQAMDB, hosted by GNPS<sup>7</sup>. This pipeline resulted in dereplicating 27 unique and structurally-diverse compounds (Table 1). Notably, all these hits were exclusive to the IQAMDB, indicating the limited number of former Annonaceous compounds references in GNPS spectral libraries prior to the upload of the IQAMDB. Among those hits, 7 were previously reported from this plant source: Annomontine, methoxyannomontine<sup>14</sup>, argentinine<sup>39</sup>, atherosperminine<sup>37</sup>, coreximine<sup>37</sup>, liriodenine<sup>39</sup>, and reticuline<sup>37</sup>. Besides being closely related to some of these putative structures, 13 further hits had already been reported from other *Annona* species: benzyltetrahydroisoquinolines anomuricine and anomurine (both known from *A. muricata*<sup>31</sup>), *N*-methylcoclaurine (from *A. sericea*<sup>40</sup>), reticuline *N*-oxide (*A. salzmanni*<sup>41</sup>) and tembetarine (tentatively known from *A. salzmanni*<sup>41</sup>); proaporphines such as glaziovine and stepharine, both reported from *A. purpurea*<sup>42,43</sup>; tetrahydroprotoberberines kikemanine (*A. glabra*<sup>44</sup>), pessoine (from *A. spinescens*<sup>45</sup>), and stepholidine (*A. cherimolia*<sup>46</sup>); aporphines such as nornuciferine (*A. muricata*<sup>47</sup>), obovanine (*A. coriacea*<sup>48</sup>) and roemerine (from both *A. squamosa*<sup>49</sup> and *A. senegalensis*<sup>50</sup>). Most other hits corresponded to metabolites occurring in other Annonaceous genera. Among all the hits, only nandigerine was hitherto unknown from Annonaceous source. Even though this aporphine alkaloid is mainly related to Hernandiaceae<sup>12</sup> and Lauraceae<sup>51</sup>, one of its derivatives, *N*-methylnandigerine *N*-oxide, had been reported from the Annonaceous *Polyalthia longifolia*<sup>52</sup>.

*Metadata.* The MS/MS spectra of the IQAMDB library are associated to a variety of details including: LC-MS/MS acquisition parameters, RT (see Supporting Information), instrument details, smiles and InChi codes, structures, and chemical formula. These metadata are available on the GNPS website.

## Code availability

The LC-MS feature detection software (MassHunter®) used in this work is commercially available from Agilent®.

Received: 1 March 2022; Accepted: 4 May 2022;

Published online: 06 June 2022

## References

1. Pyne, M. E. *et al.* A yeast platform for high-level synthesis of tetrahydroisoquinoline alkaloids. *Nature Commun.* **11**, 1–10 (2020).
2. Nguyen, V. K. & Kou, K. G. The biology and total syntheses of bisbenzylisoquinoline alkaloids. *Org. Biomol. Chem.* **19**, 7535–7543 (2021).
3. Bisset, N. G. Plants as a source of isoquinoline alkaloids. in *The Chemistry and Biology of Isoquinoline Alkaloids* 1–22 (Springer, 1985).
4. Galanis, S., Thodey, K., Trenchard, I. J., Interrante, M. F. & Smolke, C. D. Complete biosynthesis of opioids in yeast. *Science* **349**, 1095–1100 (2015).
5. Li, Y. *et al.* Complete biosynthesis of noscapine and halogenated alkaloids in yeast. *Proc. Natl Acad. Sci.* **115**, E3922–E3931 (2018).
6. Kulagina, N., Papon, N. & Courdavault, V. Microbial Cell Factories for Tetrahydroisoquinoline Alkaloid Production. *ChemBioChem* **22**, 639–641 (2021).
7. Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 (2016).
8. Ramos, A. E. F. *et al.* Collected mass spectrometry data on monoterpene indole alkaloids from natural product chemistry research. *Sci. Data* **6**, 1–6 (2019).
9. Olivier-Jimenez, D. *et al.* A database of high-resolution MS/MS spectra for lichen metabolites. *Sci. Data* **6**, 1–11 (2019).
10. Lebœuf, M., Cavé, A., Bhaumik, P. K., Mukherjee, B. & Mukherjee, R. The phytochemistry of the annonaceae. *Phytochemistry* **21**, 2783–2813 (1980).
11. Hutchinson, J. The genera of flowering plants. Dicotyledones. Vol. 1. *The genera of flowering plants. Dicotyledones. Vol. 1.* (1964).
12. Lúcio, A. S. S. C., de Almeida JR., da-Cunha, E. V. L., Tavares, J. F. & Barbosa Filho, J. M. Alkaloids of the Annonaceae: Occurrence and a Compilation of Their Biological Activities. in *The Alkaloids: Chemistry and Biology* **vol. 74** 233–409 (Elsevier, 2015).
13. Lebœuf, M. & Cavé, A. Alkaloids of Annonaceae. XIV. Identification of canangine and eupolauridine, new alkaloids with a naphthyridine nucleus. *Lloydia* **39**, 459–460 (1976).
14. Lebœuf, M. *et al.* Alkaloids of the Annonaceae. Part 33. Annomontine and methoxyannomontine, two new pyrimidine-β-carboline-type alkaloids from *Annona montana*. *J. Chem. Soc., Perkin Trans.* 1205–1208 (1982).
15. Costa, E. V. *et al.* Full NMR analysis of annomontine, methoxy-annomontine and N-hydroxyannomontine pyrimidine-β-carboline alkaloids. *Magn. Reson. Chem.* **46**, 69–74 (2008).
16. Cave, A., Guinaudeau, H., Leboeuf, M., Ramahatra, A. & Razafindrazaka, J. Alcaloïdes des Annonacées XVIII1: Alcaloïdes des Ecories du tronc du *Polyalthia suaveolens* Engl. et Diels. *Planta Med.* **33**, 243–250 (1978).
17. de Souza Araújo, M., da Silva, F. M. A., Koolen, H. H. F. & Costa, E. V. Isoquinoline-derived alkaloids from the bark of *Guatteria olivacea*(Annonaceae). *Biochem. Syst. Ecol.* **92**, 104105 (2020).
18. Nardelli, V. B. *et al.* Isoquinoline-derived alkaloids and one terpene lactone from the leaves of *Duguetia pycnastera*(Annonaceae). *Biochem. Syst. Ecol.* **94**, 104206 (2021).
19. Suthiphasilp, V. *et al.* Dasymaschalolactams A–E, Aristolactams from a Twig Extract of *Dasymaschalon dasymaschalum*. *J. Nat. Prod.* **82**, 3176–3180 (2019).
20. Paz, W. H. *et al.* Structure-Based Molecular Networking for the Target Discovery of Oxahomoaporphine and 8-Oxohomoaporphine Alkaloids from *Duguetia surinamensis*. *J. Nat. Prod.* **82**, 2220–2228 (2019).
21. Fox Ramos, A. E., Evanno, L., Poupon, E., Champy, P. & Beniddir, M. A. Natural products targeting strategies involving molecular networking: Different manners, one goal. *Nat. Prod. Rep.* **36**, 960–980 (2019).
22. Fouotsa, H. *et al.* Voatriafricanines A and B, Trimeric Vobasine-Aspidosperma-Aspidosperma Alkaloids from *Voacanga africana*. *J. Nat. Prod.* **84**, 2755–2761 (2021).

23. Fouotsa, H. *et al.* Pyrrovolbasine, hybrid alkylated pyrraline monoterpenoid indole alkaloid pseudodimer discovered using a combination of mass spectral and NMR-based machine learning annotations. *Org. Biomol. Chem.* **20**, 98–105 (2022).
24. da Silva, F. M. *et al.* Positive electrospray ionization ion trap mass spectrometry and *ab initio* computational studies of the multi-pathway fragmentation of oxoaporphine alkaloids. *Int. J. Mass Spectrom.* **418**, 30–36 (2017).
25. Neto, F. C. *et al.* Characterization of aporphine alkaloids by electrospray ionization tandem mass spectrometry and density functional theory calculations. *Rapid Commun. Mass Spectrom.* **34**, e8533 (2020).
26. Richomme, P., Lavault, M., Jacquemin, H. & Bruneton, J. Etude des Hernandiacées VI (1): lignanes et alcaloïdes de *Hernandia guianensis*. *Planta Med.* **50**, 20–22 (1984).
27. Chalandre, M.-C., Bruneton, J., Cabalion, P. & Guinaudeau, H. Alcaloïdes de *Gyrocarpus americanus*. *J. Nat. Prod.* **49**, 101–105 (1986).
28. Lavault, M. *et al.* Alcaloïdes bisbenzylisoquinoléiques de *Albertisia* cf. *A. papuana*. *Can. J. Chem.* **65**, 343–347 (1987).
29. IsoQuinoline and Annonaceous Metabolites Data Base (IQAMDB), GNPS, <https://gnps.ucsd.edu/ProteoSAFe/gnpslibrary.jsp?library=IQAMDB>
30. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
31. Chambers, M. C. *et al.* A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnol.* **30**, 918–920 (2012).
32. Pluskal, T., Castillo, S., Villar-Briones, A. & Orešić, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 1–11 (2010).
33. Myers, O. D., Sumner, S. J., Li, S., Barnes, S. & Du, X. One Step Forward for Reducing False Positive and False Negative Compound Identifications from Mass Spectrometry Metabolomics Data: New Algorithms for Constructing Extracted Ion Chromatograms and Detecting Chromatographic Peaks. *Anal. Chem.* **89**, 8696–8703 (2017).
34. Le Pogam, P. GNPS\_IQAMDB Isoquinolines and other Annonaceous Metabolites DataBase. *MassIVE* <https://doi.org/10.25345/C55D8NG3T> (2022).
35. Shamma, M. The bisbenzylisoquinolines. in *The Isoquinoline Alkaloids. Chemistry and Pharmacology*. 115–152 (Academic Press, 1972).
36. Shamma, M. & Moniot, J. The Systematic Classification of Bisbenzylisoquinolines. *Heterocycles* **4**, 1817–1824 (1976).
37. Leboeuf, M. *et al.* Alcaloïdes des annonacées XL: Etude chimique et pharmacologique des alcaloïdes de l'*Annona montana*. *Plant. Medic. Phytother.* **3**, 169–184 (1982).
38. Nothias, L.-F. *et al.* Feature-based molecular networking in the GNPS analysis environment. *Nature Methods* **17**, 905–908 (2020).
39. Wu, Y.-C., Chang, G.-Y., Chang-Yih, D. & Shang-Kwei, W. Cytotoxic alkaloids of *Annona montana*. *Phytochemistry* **33**, 497–500 (1993).
40. Campos, F. R. *et al.* Isoquinoline alkaloids from leaves of *Annona sericea* (Annonaceae). *Biochem. Syst. Ecol.* **36**, 804–806 (2008).
41. Lima, J. M., Leme, G. M., Costa, E. V. & Cass, Q. B. LC-HRMS and acetylcholinesterase affinity assay as a workflow for profiling alkaloids in *Annona salzmannii* extract. *J. Chromatogr. B* **1164**, 122493 (2021).
42. Chang, F.-R., Wei, J.-L., Teng, C.-M. & Wu, Y.-C. Antiplatelet Aggregation Constituents from *Annona purpurea*. *J. Nat. Prod.* **61**, 1457–1461 (1998).
43. Sonnet, P. E. & Jacobson, M. Tumor inhibitors II: cytotoxic alkaloids from *Annona purpurea*. *J. Pharm. Sci.* **60**, 1254–1256 (1971).
44. Chang, F.-R., Chen, C.-Y., Hsieh, T.-J., Cho, C.-P. & Wu, Y.-C. Chemical Constituents from *Annona glabra* III. *J. Chinese Chemical Soc.* **47**, 913–920 (2000).
45. Queiroz, E. F., Roblot, F., Cavé, A., de Q. Paulo, M. & Fournet, A. Pessoine and spinosine, two catecholic berbines from *Annona spinescens*. *J. Nat. Prod.* **59**, 438–440 (1996).
46. Villar, A., Mares, M., Rios, J. L. & Cortes, D. Alkaloids from *Annona cherimolia* leaves. *J. Nat. Prod.* **48**, 151–152 (1985).
47. Hasrat, J. A., De Bruyne, T., De Backer, J.-P., Vauquelin, G. & Vlietinck, A. J. Isoquinoline derivatives isolated from the fruit of *Annona muricata* as 5-HTergic 5-HT(1A) receptor agonists in rats: Unexploited antidepressive (lad) products. *J. Pharm. Pharmacol.* **49**, 1145–1149 (1997).
48. Nogueira da Silva Avelino Oliveira Rocha, G. *et al.* Chemical constituents from the leaves and branches of *Annona coriacea* Mart. (Annonaceae). *Biochem. Syst. Ecol.* **97**, 104297 (2021).
49. Bhakuni, D. S., Tewari, S. & Dhar, M. M. Aporphine alkaloids of *Annona squamosa*. *Phytochemistry* **11**, 1819–1822 (1972).
50. You, M. *et al.* (-)-Roemerine, an aporphine alkaloid from *Annona senegalensis* that reverses the multidrug-resistance phenotype with cultured cells. *J. Nat. Prod.* **58**, 598–604 (1995).
51. Teles, M. M. R. S. *et al.* Alkaloids of the Lauraceae. *The Alkaloids: Chemistry and Biology* **82**, 147–304 (2019).
52. Wu, Y. C. Azafluorene and aporphine alkaloids from *Polyalthia longifolia*. *Heterocycles* **29**, 463–475 (1989).
53. Leboeuf, M. *et al.* Alcaloïdes des Annonacées XXIX: Alcaloïdes de l'*Annona muricata* L. *Planta Med.* **42**, 37–44 (1981).
54. Costa, E. V. *et al.* Benzylated Dihydroflavones and Isoquinoline-Derived Alkaloids from the Bark of *Diclinanona calycina* (Annonaceae) and Their Cytotoxicities. *Molecules* **26**, 3714 (2021).
55. Lu, S.-T., Wu, Y.-C. & Leou, S.-P. Alkaloids of formosan *Fissistigma* and *Goniothalamus* species. *Phytochemistry* **24**, 1829–1834 (1985).
56. Chaves, M. H., De Santos, L. A., Lago, J. H. G. & Roque, N. F. Alkaloids from *Porcelia macrocarpa*. *J. Nat. Prod.* **64**, 240–242 (2001).
57. Rasamizafy, S., Hocquemiller, R., Cavé, A. & Fournet, A. Alcaloïdes des Annonacées, 78. Alcaloïdes des Écorces d'un *Duguetia spixiana* de Bolivie. *J. Nat. Prod.* **50**, 674–679 (1987).

## Acknowledgements

The authors would like to dedicate this article to Pr. Michel Leboeuf, whose lifetime work was the basis for the chemical library of isoquinoline alkaloids digitized herein, and who supported the current project, and to the memory of Prof. André Cavé. The authors are also indebted to former researchers of the University of Angers, mostly Prof. Jean Bruneton and Prof. Hélène Guinaudeau, who contributed to the isolation of many other compounds comprised in this database.

## Author contributions

P.L.P. conceived the project. S.A.A. isolated some compounds included in the IQAMDB repertoire. S.A.A., T. O., Y.A.K., A.J. and P.L.P. contributed to sample preparation. P.L.P. performed the acquisition of the LC-MS/MS data. M.A.B and P.L.P. performed the technical validation of the IQAMDB. E.V.C., F.M.A.d.S. and H.H.F.K. isolated and collected Brasil compounds. D.B., A.M.L.R.-R. and P.R. isolated and collected compounds from Université d'Angers. P.C. and P.L.P advised on the extraction aspects of the work conducted by S.A.A. P.L.P. supervised the work at Université Paris-Saclay and wrote the manuscript. All the authors read and commented the manuscript.

## Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-022-01345-y>.

**Correspondence** and requests for materials should be addressed to P.L.P.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022