Malaysian Journal of Chemistry, 2022, Vol. 24(4), 150-160
(Special Issue: *International Conference on Natural Products 2022*)

Shedding Light on *Piper*'s Identity via Computational Mass Spectrometry

Muhamad Faris Osman[1,2]*, Siti Munirah Mohd Faudzi[1,3], Shamsul Khamis[4], Shahrul Razid Sarbini[5] and Khozirah Shaari[1]

[1]Natural Medicines and Products Research Laboratory (NaturMeds), Institute of Bioscience, Universiti Putra Malaysia, Serdang, 43400 Serdang, Selangor, Malaysia
[2]Department of Pharmaceutical Chemistry, Kulliyyah of Pharmacy, International Islamic University Malaysia, Kuantan, 25200 Kuantan, Pahang, Malaysia
[3]Department of Chemistry, Faculty of Science, Universiti Putra Malaysia, Serdang, 43400 Serdang, Selangor, Malaysia
[4]Department of Biological Sciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, 43600 Bangi, Selangor, Malaysia
[5]Department of Crop Science, Faculty of Agricultural Science and Forestry, Universiti Putra Malaysia, Bintulu, 97008 Bintulu, Sarawak, Malaysia
*Corresponding author (e-mail: farisosman@iium.edu.my)

Morphological characteristics of *Piper rubro-venosum* hort. ex Rodigas bear a close resemblance to a plant identified as *Piper crocatum* Ruiz & Pav. in the literature. Hence, this study aimed to investigate whether both names describe the same species using data-dependent acquisition (DDA) LC-MS/MS analysis of methanol leaf extracts of *P. rubro-venosum* in positive and negative electrospray ionization modes. The data were analyzed using two computational mass spectrometry methods: spectral libraries search implemented in Global Natural Product Social (GNPS) molecular networking and molecular structure databases search implemented in SIRIUS. Classical molecular networking implied that the metabolites giving rise to two features with the highest intensities in the positive ionization chromatograms had distinct MS/MS spectra. De novo molecular formula annotations and machine learning predictions in SIRIUS suggested that both features were sodiated precursor ions of neolignans. Based on the accurate mass of the precursor ions, the two features were annotated as crocatin A and B, which are bicyclooctanoid neolignans previously isolated in relatively large amounts from *P. crocatum* leaves. Based on the findings, it can be concluded that *P. crocatum* and *P. rubro-venosum* could be two names in the literature used to refer to one species of *Piper*.

Key words: *Piper rubro-venosum*; *Piper crocatum*; GNPS; SIRIUS; crocatin

Received: September 2022; Accepted: November 2022

Our previous study of *Piper betle* L. variants in Malaysia has revealed the occurrence of an introduced species identified as *Piper rubro-venosum* hort. ex Rodigas based on high similarity (99.39%) of its internal transcribed spacer 2 nuclear ribosomal DNA (ITS2) sequences to that of *P. rubro-venosum* in Gen-Bank [1]. Morphological characteristics of the species depicted in [1] explain its local name, "sirih merah", which means red betel leaf, owing to the remarkable red colouration of the leaf underside. Combination of this characteristic with silver-pink pattern on the adaxial surface of the leaf makes "sirih merah" a unique ornamental plant [2]. Other than that, in Indonesia, "sirih merah" leaves are traditionally prepared by infusing a leaf in hot or warm water. The infusion is consumed twice daily to help reduce blood sugar levels and prevent coronary heart diseases [3,4]. The traditional uses of the plant part have become a basis for further investigations of its potential use as an anti-hyperglycemic [5-7] and anti-inflammatory agent [8-11].

The medicinal significance of "sirih merah" leaves is apparent from their inclusion in the Indonesian herbal monograph, along with *P. betle* leaves and *P. cubeba*, *P. nigrum*, and *P. retrofractum* fruits [12]. In the monograph, *P. crocatum* Ruiz & Pav. is the sole scientific name used to describe "sirih merah". From the year 2008 until 2022, the scientific name is also used in reviews and studies on diverse functional aspects of "sirih merah" (*e.g.*, botany, phytochemistry, and pharmacology) [2-20]. The specific epithet *crocatum* is a Latin word, which means saffron yellow (orange-yellow) [21] as in *Tritonia crocata* (L.) Ker Gawl. [22]. Hence, the appropriateness of using *crocatum* to describe "sirih merah" is questionable. Moreover, from our communication (via email) with an Indonesian researcher who studied the phylogeny of Malesian-Pacific *Piper* [23], we were informed that the identification of "sirih merah" as *P. crocatum* is incorrect because *P. crocatum* is a Neotropical *Piper*

†Paper presented at the International Conference on Natural Products 2022 (ICNP 2022)

(R. Asmarayani, personal communication, March 23, 2021).

Given the inconsistency in scientific naming of "sirih merah" either as *P. rubro-venosum* [1,23] or *P. crocatum* in the literature [2-20], this study aimed to assess the similarity of the specialized metabolite profile of a species previously identified as *P. rubro-venosum* [1] with that of *P. crocatum* in the literature using computational liquid chromatography tandem mass spectrometry (LC-MS/MS) methods to annotate the metabolites. To this end, we hypothesized that if highly abundant specialized metabolites of *P. crocatum* leaves are detected in our leaf samples of *P. rubro-venosum*, then *P. rubro-venosum* and *P. crocatum* could be two scientific names referring to one *Piper* species.

EXPERIMENTAL

Chemicals and Materials

LC-MS grade methanol (0.1 µm filtered, ≥99.9% purity) and PTFE syringe filters (13 mm, 0.22 µm) were purchased from Fisher Scientific (Canada) and Orioner Hightech (Malaysia), respectively. LC-MS grade water, acetonitrile, and formic acid were purchased from Merck KGaA (Germany).

Samples and Extracts Preparation

Two (2) variants of leaf samples, *i.e.*, green-underside leaf (n = 1) and maroon-underside leaf (n = 1) were collected at 3:30 p.m. on 17th April 2019 from the same *P. rubro-venosum* studied in [1]. A voucher specimen (UKMB404018) was deposited in the herbarium of Universiti Kebangsaan Malaysia (UKMB). Collection of plant materials and preparation of samples and extracts were carried out adapting the procedures described in [1]. Briefly, the leaf samples were ground in liquid nitrogen using mortar and pestle and lyophilized. Next, 30 mg of lyophilized samples were weighed and sonicated in 1.2 mL LC-MS grade methanol for 3 min. The supernatant was filtered using 0.22 µm PTFE syringe filters into a 1.5 mL screw top vial (Agilent Technologies, USA) prior to LC-MS/MS analysis (see Figure S1).

LC-MS/MS Data Acquisition

The method of LC-MS/MS data acquisition was adapted from previous studies [24,25] and the difference in instrumentation and acquisition parameters is described as follows. The LC-MS/MS data were acquired using a Dionex UltiMate 3000 Standard UHPLC system equipped with photodiode array (PDA) detector and coupled to an Orbitrap Q-Exactive Focus mass spectrometer (Thermo Fisher Scientific, USA). The PDA recorded spectra from 190-600 nm with resolution 2.0 nm, data acquisition rate 5.0 Hz, and response time 2.0 s. The instruments were controlled using Dionex Chromatography Mass Spectrometry Link (DCMS^{Link}) 2.11 (Thermo Fisher Scientific, USA).

An aliquot of 5 µL extract was injected into an ACQUITY UPLC BEH C18 column (particle size 1.7 µm, pore size 130 Å, 2.1 mm × 100 mm; Waters Corporation, USA) and eluted (0.4 mL/min, 40 °C) with water (A) and acetonitrile (B), both containing 0.1% formic acid with the following gradient: 5% B from 0.00-0.45 min, 5-100% B from 0.45-30.05 min, isocratic at 100% B until 34.27 min, 100%-5% B from 34.27-34.70 min, and isocratic at 5% B until 38.50 min. Extraction solvent blank (LC-MS grade methanol) was injected at the beginning of the run sequence and after every extract injection.

The LC-MS/MS data were acquired in both positive (+) and negative ionization (–) modes. The ion source parameters were set as follows: spray voltage 4.0 kV, sheath gas (N_2) flow rate 80.5 units, auxiliary gas flow rate 40.0 units, and capillary temperature 355.0 °C. The mass spectrometer was operated in Full MS/dd-MS^2 Discovery mode with resolution 70,000, scan range 150-2000 *m/z*, automatic gain control (AGC) target 3.0×10^6, and maximum injection time 250 ms. The data-dependent MS/MS events were set to resolution 17,500 FWHM, isolation window 4.0 *m/z*, stepped normalized collision energy (NCE) of 15, 30, and 45, default charge state 1, AGC target 1.0×10^5, minimum AGC target 8.0×10^3, maximum injection time 60 ms, loop scan 3 scans, intensity threshold 1.3×10^5, and dynamic exclusion 3.0 s. Centroid MS and MS/MS spectra were acquired for all extracts and blanks. The chromatograms and spectra were visualized using Qual Browser Thermo Xcalibur 4.0.27.19, MZmine 2 [26], Microsoft Excel for Microsoft 365 (version 2209) 64-bit, Metabolomics Spectrum Identifier Resolver (version 6.12) [27], and GNPS Dashboard [28].

Spectral Libraries Search

Searching the experimental positive and negative ionization MS/MS spectra in spectral libraries was conducted in the current study as it facilitates dereplication of known metabolites detected in the extracts [29]. Xcalibur raw files in .raw format were converted into feature list files in .mzXML format using ProteoWizard MSConvert software (version 3.0.21321-f1d1b0f) [30]. Next, the .mzXML files of the extracts and extraction solvent blank were uploaded to the online workflow at GNPS [29,31] via FTP (WinSCP) to generate classical molecular network (MN), which was then searched against GNPS spectral libraries. Mass tolerances for precursor and fragment ion were set at 0.005 Da. Edges between two nodes were created in the MN when the alignment score (*i.e.*, cosine score) between two consensus MS/MS spectra is ≥ 0.8 and the spectra pair has ≥ 5 matched fragment ions. MS/MS similarity score threshold and minimum matched fragment ions for spectral libraries search were set to 0.8 and 5, respectively. The MN was visualized using Cytoscape (version 3.9.0) [32].

Molecular Structure Databases Search

Searching in molecular structure databases with MS/MS spectra was employed in the current study to overcome the limitation of searching in spectral libraries, which have a limited number of reference MS/MS spectra for many metabolites [33]. As a greater number of high intensity peaks was observed in the positive ionization base peak chromatograms (BPCs) compared to the negative ionization BPCs, and these peaks remained unannotated even after spectral libraries search in GNPS, only the positive ionization datasets were searched against molecular structure databases in SIRIUS 5 [34]. This allows for a more detailed manual examination of the metabolites annotation executed by the tools in SIRIUS 5. Firstly, the .raw files of the extracts were preprocessed in MZmine 2, and the parameters are described as follows. The noise level for mass detection of MS level 1 and 2 were set to 1.5×10^6 and 0.0, respectively. For ADAP chromatogram builder [35]: minimum group size in number of scans = 5, group intensity threshold = 1.3×10^5, minimum highest intensity = 3.0×10^6, and scan to scan accuracy = 0.001 *m/z* or 3.0 ppm. For 13C isotope filter: *m/z* tolerance = 0.001 *m/z* or 3.0 ppm, retention time tolerance = 0.1 min, monotonic shape = true, maximum charge = 1, representative isotope = lowest *m/z*, never remove feature with MS2 = true, and original feature list = keep.

Next, the resulting feature list file was converted to .mgf file with the following parameters: merge MS/MS = true; merge spectra across samples, *m/z* merge mode = weighted average (remove outliers), intensity merge mode = mean intensity, expected mass deviation = 0.001 *m/z* or 3.0 ppm, cosine threshold = 0%, signal count threshold = 0%, isolation window offset = 0, and isolation window width = 4.0 *m/z*. Multiple charge features were excluded from the final .mgf files. The .mgf files were exported to SIRIUS 5 for de novo molecular formula annotation using SIRIUS [36-39] and ZODIAC [40], structure annotation using CSI:FingerID [33,41–43] and COSMIC [44], and chemical class prediction using CANOPUS [45,46]. ZODIAC is a network-based algorithm that employs Gibbs sampling to re-rank molecular formula annotations by SIRIUS, by considering shared fragment ions and losses between fragmentation trees in complete LC-MS/MS datasets [40]. COSMIC is a workflow that provides a confidence score, which combines E-value estimation and linear support vector machines (SVMs) with enforced features' directionality. The COSMIC's confidence score assists the users to decide whether the first-ranked chemical structures predicted by CSI:FingerID are likely correct or incorrect [44].

The parameters for SIRIUS, ZODIAC, and CSI:FingerID are described as follows. For SIRIUS: filter by isotope pattern = false, MS/MS isotope scorer = ignore, MS2 mass accuracy = 5 ppm, candidates stored = 10, minimum candidates per ion stored = 1, possible ionizations = $[M+H]^+$, $[M+K]^+$, and $[M+Na]^+$, use heuristic above 300 *m/z*, and use heuristic only above 650 *m/z*. For ZODIAC: considered candidates 300 *m/z* = 10, considered candidates 800 *m/z* = 15, use 2-step approach = true, edge threshold = 0.95, minimum local connections = 10, and Gibbs sampling = 20,000 iterations, 2,000 burn-in, and 10 separate runs. For CSI:FingerID: all listed fallback adducts were included except $[M+C_2H_6OS+H]^+$ with score threshold = true and all non in silico molecular structure databases were selected (all databases in SIRIUS 5 were selected except the EcoCyc Mine, KEGG Mine, and YMDB Mine). Lastly, the outputs of SIRIUS, CSI:FingerID, and CANOPUS were manually examined and selected chemical structure candidates were searched in the naturaL prOducTs occUrrence databaSe (LOTUS) to assess for their occurrence in other plant species [47].

RESULTS AND DISCUSSION

Spectral Libraries Search

Compared to the negative ionization BPCs, a greater number of high intensity peaks (intensities higher than 2×10^9) was observed in the positive ionization BPCs of green- (15 peaks) and maroon-underside (13 peaks) *P. rubro-venosum* leaf extracts (Figure 1). The highest base peak intensities were recorded at retention times 11.64 (peak 1) and 14.47 min (peak 2). Zoomed-in view of the BPCs is shown in Figures S2 and S3. Metabolites that fulfilled the criteria of cosine similarity of at least 0.8 and minimum 5 matched fragment ions within the allowed mass tolerance of 0.005 Da with the reference MS/MS spectra in GNPS are listed in Table 1 (see Table S1 for a more detailed view of the MS/MS spectra).

Despite the successful annotations of 31 LC-MS/MS features in Table 1, peak 1 and 2 remained unannotated, indicating the probable absence of reference MS/MS spectra of the corresponding metabolites from the GNPS spectral libraries. Nevertheless, the MN nodes representing metabolites 1 and 2 are connected to at most one other unannotated node, signifying that both metabolites might belong to chemical class/es different from those listed in Table 1 (Figure S4).

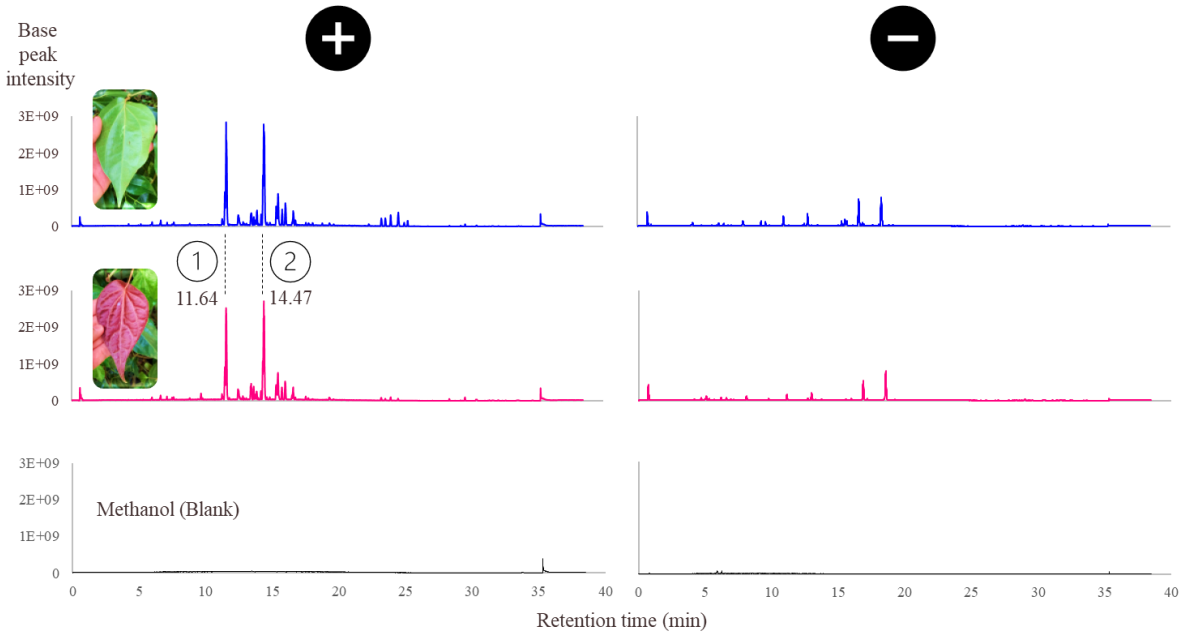

Figure 1. Positive and negative ionization LC-MS BPCs of *Piper rubro-venosum* methanolic green- and maroon-underside leaf extracts and methanol. Base peaks at 11.64 and 14.47 min are labelled as peaks 1 and 2, respectively. Overlaid and zoomed-in BPCs are shown in Figures S2 and S3.

The absence of an edge that connects the MN nodes of metabolites 1 and 2 in Figure S4 denotes modified cosine similarity of less than 0.8 between consensus MS/MS spectrum of each node. A fundamental concept behind molecular networking of LC-MS/MS data in GNPS is that two structurally related molecules (*e.g.*, same chemical class) share fragment ion patterns when subjected to conventional collision-induced dissociation (CID)-based fragmentation. Using MS-Cluster algorithm in the GNPS, the raw MS/MS spectra are aligned to each other and spectra with highly similar fragment ion patterns are collapsed into an MN node, which is described by a representative consensus MS/MS spectrum [29]. The consensus spectrum is computationally generated from the highly similar raw MS/MS spectra after a series of spectra merging, intensity normalization, and scaling (*i.e.*, natural logarithm of fragment ion peak's normalized intensity to minimize domination of high intensity fragment ion peaks in the spectrum on the outcome of spectral similarity computation), and peak filtering [48].

The assumption that metabolites 1 and 2 belong to two different chemical classes because their consensus MS/MS spectra are different does not necessarily hold because previous investigation showed that with respect to the chemical space of spectral libraries and structure databases, the MS/MS spectra of four chemical classes, *i.e.*, alkaloids, lignans, organic polymers, and hydrocarbon derivatives did not have characteristic fragment ion patterns to define the aglycone and/or unique substructure moiety, and their interpretation was not trivial [49]. Therefore, no connecting edge between the MN nodes of metabolites 1 and 2 remains possible if both metabolites belong to one these four chemical classes.

De novo Molecular Formula Annotation

Based on SIRIUS output shown in Table 2, the LC-MS features corresponding to metabolites 1 and 2 were annotated as sodiated precursor ions $[M+Na]^+$. The first-ranked molecular formulas for both metabolites *(i.e.*, metabolite 1: $C_{23}H_{30}O_7$, metabolite 2: $C_{25}H_{32}O_8$) were accepted as the correct molecular formulas as their SIRIUS and ZODIAC scores were much higher than the second- and third-ranked molecular formulas (see Figures S5 and S6 for computed fragmentation trees of both metabolites and fragment ion peaks explained by the trees). The acceptance of $C_{23}H_{30}O_7$ and $C_{25}H_{32}O_8$ as the correct molecular formulas for metabolites 1 and 2 was supported by the occurrence of protonated adduct and sodiated dimers and trimers of the metabolites with *m/z* error less than 5.0 ppm (see Figures S7 and S8 for the corresponding MS/MS spectra).

Table 1. List of metabolites annotated via spectral libraries search in GNPS.

RT (min)	Metabolite	Molecular formula	Adduct	*M/z*	*M/z* error in Da (in ppm)	No. shared MS/MS peaks*	Modified cosine similarity
0.62	Raffinose	$C_{18}H_{32}O_{16}$	$[M+Na]^+$	527.158	0.001 (1.85)	6	0.94
0.65	Adenosine	$C_{10}H_{13}N_5O_4$	$[M+H]^+$	268.100	0.004 (14.91)	5	0.94
0.67	5'-Deoxy-5'-(methylsulfinyl)adenosine	$C_{11}H_{15}N_5O_4S$	$[M+H]^+$	314.090	0.001 (3.21)	5	0.86
0.69	Sucrose	$C_{12}H_{22}O_{11}$	$[M–H]^–$	341.108	0.001 (2.95)	8	0.97
0.70	Raffinose	$C_{18}H_{32}O_{16}$	$[M–H]^–$	503.162	0.000 (0.00)	10	0.97
1.05	Citric acid	$C_6H_8O_7$	$[M–H]^–$	191.020	0.001 (5.27)	6	0.96
3.10	Syringin	$C_{17}H_{24}O_9$	$[M+Na]^+$	395.131	0.001 (2.47)	6	0.82
4.29	3-[3,4-Dihydroxy-6-(hydroxymethyl)-5-[3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxyoxan-2-yl]oxy-2-(3,4-dihydroxyphenyl)-5,7-dihydroxychromen-4-one	$C_{27}H_{30}O_{17}$	$[M–H]^–$	625.141	0.001 (1.66)	7	0.87
4.30	Luteolin-6-*C*-glucoside	$C_{21}H_{20}O_{11}$	$[M–H]^–$	447.093	0.001 (2.25)	10	0.89
4.61	Luteolin-8-*C*-glucoside	$C_{21}H_{20}O_{11}$	$[M+H]^+$	449.108	0.000 (0.00)	12	0.96
4.70	(2*R*,3*S*,4*S*,5*R*,6*R*)-2-[[(2*S*,3*R*,4*R*)-3,4-Dihydroxy-4-(hydroxymethyl)oxolan-2-yl]oxymethyl]-6-(2-phenylethoxy)oxane-3,4,5-triol	$C_{19}H_{28}O_{10}$	$[M–H]^–$	415.161	0.000 (0.00)	10	0.94
4.75	Quercetin-3-*O*-vicianoside	$C_{26}H_{28}O_{16}$	$[M–H]^–$	595.130	0.001 (1.64)	9	0.95
4.80	Vitexin-2''-*O*-rhamnoside	$C_{27}H_{30}O_{14}$	$[M–H]^–$	577.156	0.001 (1.69)	13	0.91
4.98	Quercetin-3-*O*-rutinoside	$C_{27}H_{30}O_{16}$	$[M–H]^–$	609.146	0.001 (1.60)	12	0.96
5.13	Quercetin-3-*O*-glucoside	$C_{21}H_{20}O_{12}$	$[M–H]^–$	463.088	0.001 (2.11)	11	0.95
5.14	Apigenin-8-*C*-glucoside	$C_{21}H_{20}O_{10}$	$[M–H]^–$	431.098	0.001 (2.34)	11	0.96
5.14	Apigenin-8-*C*-glucoside-2'-rhamnoside	$C_{27}H_{30}O_{14}$	$[M+H]^+$	579.172	0.001 (1.69)	14	0.95
5.28	Quercetin-3-*O*-rutinoside	$C_{27}H_{30}O_{16}$	$[M+Na]^+$	633.143	0.001 (1.54)	10	0.88
5.29	Quercetin-3-*O*-rutinoside	$C_{27}H_{30}O_{16}$	$[M+H]^+$	611.161	0.000 (0.00)	12	0.96
5.33	3-[(2*S*,3*R*,4*S*,5*S*,6*R*)-4,5-Dihydroxy-6-(hydroxymethyl)-3-[(2*S*,3*R*,4*S*,5*R*)-3,4,5-trihydroxyoxan-2-yl]oxyoxan-2-yl]oxy-5,7-dihydroxy-2-(4-hydroxyphenyl)chromen-4-one	$C_{26}H_{28}O_{15}$	$[M–H]^–$	579.136	0.000 (0.00)	7	0.80
5.58	Kaempferol-3-*O*-rutinoside	$C_{27}H_{30}O_{15}$	$[M–H]^–$	593.151	0.001 (1.65)	9	0.96
5.81	Kaempferol-3-*O*-glucoside	$C_{21}H_{20}O_{11}$	$[M–H]^–$	447.093	0.001 (2.25)	9	0.91
5.97	Isorhamnetin-3-*O*-glucoside	$C_{22}H_{22}O_{12}$	$[M–H]^–$	477.104	0.000 (0.00)	7	0.92
10.55	Rhamnetin	$C_{16}H_{12}O_7$	$[M–H]^–$	315.051	0.001 (3.20)	8	0.88
16.67	1-(9*Z*,12*Z*-Octadecadienoyl)-2-hydroxy-sn-glycero-3-phosphocholine	$C_{26}H_{50}NO_7P$	$[M–H]^–$	564.329	0.002 (3.57)	5	0.89
16.78	13*S*-Hydroxy-9*Z*,11*E*,15*Z*-octadecatrienoic acid	$C_{18}H_{30}O_3$	$[M+H–H_2O]^+$	277.216	0.000 (0.00)	8	0.87
17.77	1-Hexadecanoyl-sn-glycero-3-phospho-(1'-*myo*-inositol)	$C_{25}H_{49}O_{12}P$	$[M–H]^–$	571.294	0.004 (7.05)	8	0.80
28.42	1,2-Di-(9*Z*,12*Z*,15*Z*-octadecatrienoyl)-sn-glycero-3-phosphocholine	$C_{44}H_{76}NO_8P$	$[M+H]^+$	778.54	0.003 (3.84)	6	0.96
28.45	2,3-Dihydroxypropyl (9*Z*,12*Z*,15*Z*)-octadeca-9,12,15-trienoate	$C_{21}H_{36}O_4$	$[M+H–H_2O]^+$	335.258	0.001 (3.00)	6	0.85
32.38	Pyropheophytin a	$C_{53}H_{72}N_4O_3$	$[M+H]^+$	813.567	0.001 (1.28)	6	0.80
33.57	Pheophytin a	$C_{55}H_{74}N_4O_5$	$[M+H]^+$	871.571	0.001 (1.19)	8	0.81

*MS/MS peaks shared between the experimental and reference spectra. The URLs for a more detailed view of the shared MS/MS peaks are listed in Table S1.

Table 2. SIRIUS output of molecular formulas annotation of metabolites 1 and 2 in the positive ionization BPCs of *Piper rubro-venosum* methanolic leaf extracts.

Rank	Molecular formula	Adduct	ZODIAC score	SIRIUS score	Tree score	Explained fragment ion peaks	Total explained intensity
Metabolite 1 (RT 11.64 min, *m/z* 441.1874)							
1	$C_{23}H_{30}O_7$	$[M+Na]^+$	100.00%	99.53%	8.48	5/47	34.17%
2	$C_{19}H_{26}N_6O_5$	$[M+Na]^+$	0.00%	0.13%	1.81	1/47	0.75%
3	$C_{17}H_{31}N_4O_6P$	$[M+Na]^+$	0.00%	0.09%	1.50	3/47	3.24%
Metabolite 2 (RT 14.47 min, *m/z* 483.1979)							
1	$C_{25}H_{32}O_8$	$[M+Na]^+$	99.75%	98.22%	43.16	20/52	32.98%
2	$C_{21}H_{28}N_6O_6$	$[M+Na]^+$	0.25%	1.66%	39.08	22/52	34.69%
3	$C_{17}H_{36}N_2O_{10}S$	$[M+Na]^+$	0.00%	0.06%	35.68	23/52	36.72%

Molecular Structure Databases and Literature Search

First-ranked chemical structure candidates of metabolites 1 and 2 predicted using multiple kernels (similarity measures) machine learning and support vector machines (SVMs) behind CSI:FingerID are depicted in Figure 2. CSI:FingerID predicted 8,4′-oxyneolignan rhaphidecursinol B ($C_{23}H_{30}O_7$) as metabolite 1 and (+)-7-acetylraphidecursinol B ($C_{25}H_{32}O_8$) as metabolite 2, of which both are present in CSI:FingerID training data and several molecular structure databases, *e.g.*, COlleCtion of Open NatUral producTs (COCONUT) [50], KNApSaCK [51], and Super Natural II [52]. The CSI:FingerID prediction suggested that the mass difference of 42.0106 Da between metabolites 1 and 2 was indicative of metabolite 2 being the acetate ester of metabolite 1.

The blue highlight in Figure 2 marks the substructures that are well-predicted to be present in the unknown query metabolite and the chemical structure candidate, whereas the numbers in percent are Tanimoto similarity scores between the predicted molecular fingerprint of the unknown query metabolite and the molecular fingerprint of the chemical structure candidate. Basing the reliability of CSI:FingerID prediction on the Tanimoto similarity scores (76.86% and 71.23%) and the substructures annotation (blue highlight) in Figure 2 implied that rhaphidecursinol B and (+)-7-acetylraphidecursinol B were probably the correct chemical structure candidates. However, COSMIC's confidence scores of both candidates were found to be close to zero (0.08 and 0.23), thus signifying that the annotations of metabolites 1 and 2 as rhaphidecursinol B and (+)-7-acetylraphidecursinol B are likely incorrect. Low COSMIC's confidence scores are also suggestive that the metabolites are absent from spectral libraries and chemical structure databases [44]. Meanwhile, chemical class prediction by CANOPUS hinted that both metabolites belong to the neolignans.

In evaluating the correctness of prediction by CSI:FingerID, the confidence score of COSMIC and chemical class prediction by CANOPUS were utilized in combination with the Tanimoto similarity and the corresponding substructures annotation. The Tanimoto similarity is one the four best scores [53] used to compare molecular fingerprints because it is easy to use and computationally efficient [54,55]. Nonetheless, a Tanimoto score of 1 does not necessarily mean that two compounds are identical because it only means that the compounds have identical molecular fingerprints [56]. It is not advisable to use the Tanimoto similarity score in case of highly similar chemical structures (*e.g.*, constitutional isomers) because it is not designed to accurately measure such high similarity [44]. For this reason, the utilization of Tanimoto similarity alone was regarded as inadequate [57] and was complemented with COSMIC's confidence score and CANOPUS. CANOPUS uses deep neural network (DNN) to assign ClassyFire compound class [45] to the features detected in an LC-MS/MS dataset [46].

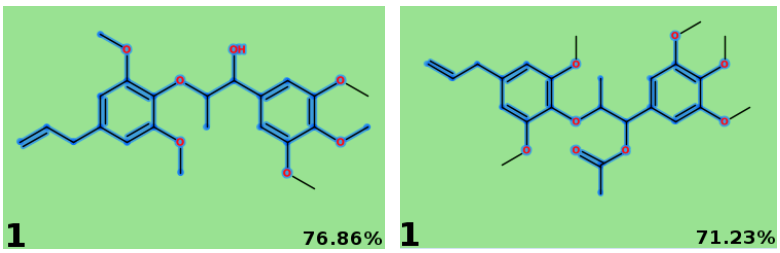

Figure 2. First-ranked chemical structure candidates of metabolites 1 (left) and 2 (right) predicted by CSI:FingerID.

Incorrect prediction by CSI:FingerID did not posit that the molecular formulas and fragmentation trees computation by SIRIUS were also incorrect. The supervised machine learning model of CSI:FingerID performs particularly well when reference MS/MS spectrum of a query metabolite is present in its training data. Poor performance of CSI:FingerID in predicting the chemical structures of certain metabolites was reported, which might be due to unusual fragmentation patterns of the metabolites or unusual instrument settings [34]. In the analysis of metabolites in complex matrices using LC-MS/MS, there is also the possibility of obtaining chimeric fragmentation spectra. Chimeric fragmentation spectrum is generated when compounds are simultaneously fragmented in the same MS/MS experiment. This is problematic especially in computational mass spectrometry because it hinders correct metabolite annotation when the chimeric fragmentation spectrum is not properly deconvoluted [58]. On the other hand, by design, SIRIUS is different from supervised machine learning as molecular structure databases are not used to train the method; they are merely used to estimate its hyperparameters of priors (*e.g.*, priors of the fragmentation tree's root, edges, and size). Hence, unlike CSI:FingerID, computations by SIRIUS do not depend on spectral or structural databases as all theoretically possible molecular formulas are considered [39].

The information obtained after MN in GNPS and structure elucidation in SIRIUS can be summarised as follows: metabolites 1 and 2 are two neolignans of a species of *Piper* that are absent from spectral libraries and molecular structure databases, with molecular formulas $C_{23}H_{30}O_7$ (418.1992 Da) for metabolite 1 and $C_{25}H_{32}O_8$ (460.2097 Da) for metabolite 2, where metabolite 2 is the acetate ester of metabolite 1. These descriptions and UV absorption maxima (λ_{max}) of both metabolites, *i.e.*, 268 nm for metabolite 1 and 266 nm for metabolite 2 (Figure 3) were found to be close to those reported for crocatin B (264 nm) and A (256 nm) in methanol [15].

Crocatin A and B are guianin-type bicyclo [3.2.1]octanoid neolignans previously isolated in relatively large amount from the ethyl acetate fraction of *P. crocatum* crude methanolic leaf extract; 3.2 g crocatin B and 6.8 g crocatin A were isolated from 120.0 g crude extract obtained from 1.1 kg dried leaves [15]. Crocatin B was also isolated from the water fraction of *P. crocatum* crude methanolic leaf extract; 0.1 g crocatin B isolated from 400.0 g crude extract obtained from 2.6 kg dried leaves [17]. Several other bicyclo[3.2.1]octanoid neolignans have been isolated from different *Piper* species, mainly from *P. puberulum* (Benth.) Maxim. and *P. kadsura* (Choisy) Ohwi, of which nine are guianin-type and five are macrophyllin-type [59]. In contrast to crocatin A and B, none of the fourteen neolignans possesses 3,4,5-trimethoxyphenyl moiety attached to C-7 position of the bicyclo[3.2.1]octanoid system, thus suggesting the occurrence of both metabolites in *P. crocatum* (or *P. rubro-venosum*) leaves is species-specific.

To check for species specificity of crocatin A and B, 2D chemical structures (disregarding the stereochemical configuration in Figure 3) of both metabolites were searched against the LOTUS database. The search result revealed that at present, both metabolites were absent from the database and two stereoisomers of crocatin A and B, namely (1*R*,5*S*,6*S*, 7*R*,8*S*)-8-hydroxy-1,3-dimethoxy-6-methyl-5-(prop-2-en-1-yl)-7-(3,4,5-trimethoxyphenyl)-bicyclo [3.2.1]oct-3-en-2-one and its acetate ester (1*S*,5*R*,6*R*, 7*S*,8*S*)-3,5-dimethoxy-7-methyl-4-oxo-1-(prop-2-en-1-yl)-6-(3,4,5- trimethoxy-phenyl)bicyclo-[3.2.1]oct-2-en-8-yl acetate have been isolated only from a plant species in the family Lauraceae Juss (order Laurales)., *i.e.*, *Licaria brasiliensis* (Nees) Kosterm. trunk [60]. With respect to the phylogeny of angiosperms, plants in the order Laurales are genetically more similar to those in the order Piperales than the other orders (with the exception of Canellales and Magnoliales) [61]. There-fore, the occurrence of crocatin A and B can be considered as specific to *P. crocatum* (or *P. rubro-venosum*) leaves. This is in contrast with quercetin that is abundant in the plant kingdom and currently a chemical marker in the pharmacognostical analysis of *P. crocatum* leaves [12].

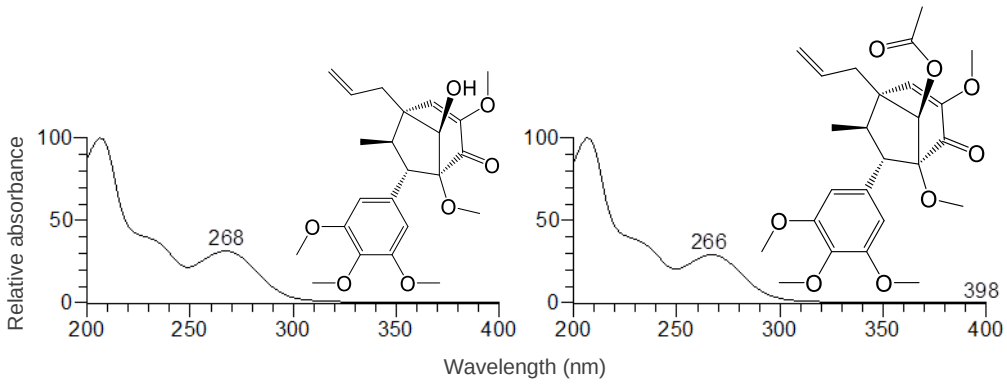

Figure 3. UV absorption spectra of metabolites 1 (left) and 2 (right), annotated as crocatin B and A, respectively.

CONCLUSION

Combination of LC-MS/MS and computational MS methods in the current study has facilitated the dereplication of two guianin-type bicyclo[3.2.1] octanoid neolignans known as crocatin A and B in two *P. rubro-venosum* leaf variants, thus supporting our hypothesis *P. rubro-venosum* and *P. crocatum* could be two scientific names in the literature used to refer to one species of *Piper*. Other than that, the findings have provided information regarding gas phase chemistry of crocatin A and B; using the LC-MS/MS conditions in the current study, both metabolites showed higher tendency to form sodiated monomers and dimers rather than other positive ionization mode adducts and despite being the acetate ester of crocatin B, the fragmentation spectrum of sodiated monomer of crocatin A was found to be highly dissimilar to that of sodiated monomer of crocatin B to the point of their fragmentation spectra being represented as two separate nodes in the GNPS molecular network. The annotation of crocatin A and B in our *P. rubro-venosum* leaf samples may serve as an impetus for revision and standardization of the scientific name of this valuable species.

DATA AVAILABILITY STATEMENT

Raw LC-MS/MS data and feature list files have been deposited in the GNPS-MassIVE online repository (dataset identifier: MSV000090428) and the public dataset is downloadable and visualizable via https://bit.ly/MSV000090428. Cytoscape session files for full view of the MN are available at https://bit.ly/Prubropositive and https://bit.ly/Prubronegative.

ACKNOWLEDGEMENTS

The authors would like to thank Muhammad Najmi Risdan for his assistance in the collection and preparation of plant materials and Universiti Putra Malaysia for the research grant GP-IPS/2018/9664600. The authors also would like to thank Ministry of Higher Education Malaysia for Skim Latihan Akademik Bumiputera (SLAB) and International Islamic University Malaysia for IIUM Postgraduate Fellowship Scheme granted to Muhamad Faris Osman.

REFERENCES

1. Osman, M. F., Lee, S. Y., Sarbini, S. R., Faudzi, S. M. M., Khamis, S., Zainudin, B. H. and Shaari, K. (2021) Metabolomics-driven discovery of an introduced species and two Malaysian *Piper betle* L. variants. *Plants*, **10(11)**, 2510.

2. Suri, M. A., Azizah, Z. and Asra, R. (2021) Traditional use, phytochemical and pharmacological review of red betel leaves (*Piper crocatum* Ruiz & Pav). *Asian Journal of Pharmaceutical Research and Development*, **9(1)**, 159–163.

3. Emrizal, Fernando, A., Yuliandari, R., Rullah, K., Indrayani, N. R., Susanty, A., Yerti, R., Ahmad, F., Sirat, H. M. and Arbain, D. (2014) Cytotoxic activities of fractions and two isolated compounds from sirih merah (Indonesian red betel), *Piper crocatum* Ruiz & Pav. *Procedia Chemistry*, **13**, 79–84.

4. Luardini, M. A., Asi, N. and Garner, M. (2019) linguistics of ethno-medicinal plants of the Dayak Ngaju community. *Language Sciences*, **74**, 77–84.

5. Safithri, M. and Fahma, F. (2008) Potency of *Piper crocatum* decoction as an antihiperglycemia in rat strain Sprague Dawley. *Journal of Bio-sciences*, **15(1)**, 45–48.

6. Safithri, M. S., Yasni, S., Bintang, M. and Ranti, A. S. (2016) Antihyperglycemic activity of *Piper crocatum* leaves and *Cinnamomum burmannii* bark mixture extract in streptozotocin-induced diabetic rats. *Journal of Mathematical and Fundamental Sciences*, **48(2)**, 178–191.

7. Shinta, D. Y. and Sudyanto (2016) Pemberian air rebusan daun sirih merah (*Piper crocatum* Ruiz & Pav) terhadap kadar glukosa dan kolesterol darah mencit putih jantan. *Journal of Sainstek*, **8(2)**, 180–185.

8. Amalia, A., Maslikah, S. I. and Lestari, S. R. (2020) Virtual screening flavonoid compounds from red betel (*Piper crocatum* Ruiz & Pav.) as inhibitor of cyclo-oxygenase-2 (COX-2). *AIP Conference Proceedings*, **2231**, 040043.

9. Atikaningrum, D. A., Ediningsih, E. and Utari, C. S. (2013) Analgesic effectiveness comparison between red betel leaf extract (*Piper crocatum*) and therapy dosage of aspirin in mice. *Biofarmasi*, **11(1)**, 1–6.

10. Gong, Y., Li, H. X., Guo, R. H., Widowati, W., Kim, Y. H., Yang, S. Y. and Kim, Y. R. (2021) Anti-allergic inflammatory components from the leaves of *Piper crocatum* Ruiz & Pav. *Biological and Pharmaceutical Bulletin*, **44(2)**, 245–250.

11. Li, H. X., Yang, S. Y., Kim, Y. H. and Li, W. (2019) Isolation of two new compounds and other constituents from leaves of *Piper crocatum* and study of their soluble epoxide hydrolase activities. *Molecules*, **24(3)**, 489.

12. Kementerian Kesehatan Republik Indonesia (2017) *Farmakope Herbal Indonesia Edisi II*, *Kementerian Kesehatan Republik Indonesia, Jakarta.*

13. Adnan, A. Z., Noer, Z. and Zulzanah (2011) Analysis of essential oil components from fresh

leaves of *Piper crocatum* Ruiz & Pav. and *Curcuma domestica* Val. *Majalah Farmasi dan Farmakologi*, **15(1)**, 17–22.

14. Afifah, S., Lukiati, B. and Maslikah, S. I. (2020) The potentials of red betel (*Piper crocatum* Ruiz & Pav) terpenoid compounds as microsomal prostaglandin E synthase-1 (mPGES-1) enzyme inhibitor of rheumatoid arthritis through virtual screening. *AIP Conference Proceedings*, **2231**, 040023.

15. Arbain, D., Nofrizal, Syafni, N., Ismed, F., Yousuf, S. and Choudhary, M. I. (2018) Bicyclo [3.2.1] octanoid neolignans from Indonesian red betle leaves (*Piper crocatum* Ruiz & Pav.). *Phytochemistry Letters*, **24**, 163–166.

16. Astuti, I. P. and Munawaroh, E. (2011) Karakteristik morfologi daun sirih merah: *Piper crocatum* Ruitz & Pav dan *Piper porphyrophyllum* N.E.Br. koleksi kebun raya Bogor. *Edisi Khusus Berkala Penelitian Hayati*, **7(A)**, 83–85.

17. Chai, Y. J., Go, Y., Zhou, H. Q., Li, H. X., Lee, S. J., Park, Y. J., Widowatib, W., Rizal, R., Kim, Y. H., Yang, S. Y. and Li, W. (2021) Unusual bicyclo[3.2.1]octanoid neolignans from leaves of *Piper crocatum* and their effect on pyruvate dehydrogenase activity. *Plants*, **10(9)**, 1855.

18. Heliawati, L., Lestari, S., Hasanah, U., Ajiati, D. and Kurnia, D. (2022) Phytochemical profile of antibacterial agents from red betel leaf (*Piper crocatum* Ruiz and Pav) against bacteria in dental caries. *Molecules*, **27(9)**, 2861.

19. Lister, I. N. E., Ginting, C. N., Girsang, E., Nataya, E. D., Azizah, A. M. and Widowati, W. (2020) Hepatoprotective properties of red betel (*Piper crocatum* Ruiz and Pav) leaves extract towards H_2O_2-induced HepG2 cells via anti-inflammatory, antinecrotic, antioxidant potency. *Saudi Pharmaceutical Journal*, **28(10)**, 1182–1189.

20. Setyawati, A., Wahyuningsih, M. S. H., Nugrahaningsih, D. A. A., Effendy, C., Fneish, F. and Fortwengel, G. (2021) *Piper crocatum* Ruiz & Pav. ameliorates wound healing through p53, E-cadherin and SOD1 pathways on wounded hyperglycemia fibroblasts. *Saudi Journal of Biological Sciences*, **28(12)**, 7257–7268.

21. Harrison, L. (2012) *Latin for gardeners*, The University of Chicago Press, Chicago.

22. *Tritonia crocata* (L.) Ker Gawl. (2022) Plants of the World Online| Kew Science. (n.d.). Retrieved from https://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:441691-1.

23. Asmarayani, R. (2018) Phylogenetic relationships in Malesian-Pacific *Piper* (Piperaceae) and their implications for systematics. *Taxon*, **67(4)**, 693–724.

24. Houriet, J., Allard, P. M., Queiroz, E. F., Marcourt, L., Gaudry, A., Vallin, L., Li, S., Lin, Y., Wang, R., Kuchta, K. and Wolfender, J. L. (2020) A mass spectrometry based metabolite profiling workflow for selecting abundant specific markers and their structurally related multi-component signatures in traditional chinese medicine multi-herb formulae. *Frontiers in Pharmacology*, **11**, 578346.

25. Norazhar, A. I., Lee, S. Y., Faudzi, S. M. M. and Shaari, K. (2021) Metabolite profiling of *Christia vespertilionis* leaf metabolome via molecular network approach. *Applied Sciences*, **11(8)**, 3526.

26. Pluskal, T., Castillo, S., Villar-Briones, A. and Orešič, M. (2010) MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*, **11**, 395.

27. Bittremieux, W., Chen, C., Dorrestein, P. C., Schymanski, E. L., Neumann, S., Meier, R., Rogers, S. and Wang, M. (2020) Universal MS/MS visualization and retrieval with the Metabolomics Spectrum Resolver web service. *BioRxiv*, 1–11.

28. Petras, D., Phelan, V. V., Acharya, D. *et al.* (2022) GNPS Dashboard: Collaborative exploration of mass spectrometry data in the web browser. *Nature Methods*, **19**, 134–136.

29. Aron, A. T., Gentry, E. C., McPhail, K. L. *et al*. (2020) Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nature Protocols*, **15**, 1954–1991.

30. Kessner, D., Chambers, M., Burke, R., Agus, D. and Mallick, P. (2008) ProteoWizard: Open source software for rapid proteomics tools development. *Bioinformatics*, **24(21)**, 2534–2536.

31. Wang, M., Carver, J., Phelan, V. *et al*. (2016) Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology*, **34**, 828–837.

32. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, **13(11)**, 2498–2504.

33. Dührkop, K., Shen, H., Meusel, M., Rousu, J.

and Böcker, S. (2015) Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proceedings of the National Academy of Sciences of the United States of America*, **112(41)**, 12580–12585.

34. Dührkop, K., Fleischauer, M., Ludwig, M. A., Aksenov, A., V.Melnik, A., Meusel, M., Dorrestein, P. C., Rousu, J. and Böcker, S. (2019) SIRIUS 4: A rapid tool for turning tandem mass spectra into metabolite structure information. *Nature Methods*, **16(4)**, 299–302.

35. Myers, O. D., Sumner, S. J., Li, S., Barnes, S. and Du, X. (2017) 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. *Analytical Chemistry*, **89(17)**, 8696–8703.

36. Böcker, S. and Rasche, F. (2008) Towards de novo identification of metabolites by analyzing tandem mass spectra. *Bioinformatics*, **24(16)**, i49–i55.

37. Rasche, F., Svatoš, A., Maddula, R. K., Böttcher, C. and Böcker, S. (2011a) Computing fragmentation trees from tandem mass spectrometry data. *Analytical Chemistry*, **83(4)**, 1243–1251.

38. Rasche, F., Svatoš, A., Maddula, R. K., Böttcher, C., & Böcker, S. (2011) Correction to computing fragmentation trees from tandem mass spectrometry data. *Analytical Chemistry*, **83(17)**, 6911.

39. Böcker, S., and Dührkop, K. (2016) Fragmentation trees reloaded. *Journal of Cheminformatics*, **8**, 5.

40. Ludwig, M., Nothias, L. F., Dührkop, K., Koester, I., Fleischauer, M., Hoffmann, M. A., Petras, D., Vargas, F., Morsy, M., Aluwihare, L., Dorrestein, P. C. and Böcker, S. (2020) Database-independent molecular formula annotation using Gibbs sampling through ZODIAC. *Nature Machine Intelligence*, **2(10)**, 629–641.

41. Heinonen, M., Shen, H., Zamboni, N. and Rousu, J. (2012) Metabolite identification and molecular fingerprint prediction through machine learning. *Bioinformatics*, **28(18)**, 2333–2341.

42. Ludwig, M., Dührkop, K. and Böcker, S. (2018) Bayesian networks for mass spectrometric metabolite identification via molecular fingerprints. *Bioinformatics*, **34(13)**, i333–i340.

43. Shen, H., Dührkop, K., Böcker, S. and Rousu, J. (2014) Metabolite identification through multiple kernel learning on fragmentation trees. *Bioinformatics*, **30(12)**, i157–i164.

44. Hoffmann, M. A., Nothias, L. F., Ludwig, M., Fleischauer, M., Gentry, E. C., Witting, M., Dorrestein, P. C., Dührkop, K. and Böcker, S. (2022) High-confidence structural annotation of metabolites absent from spectral libraries. *Nature Biotechnology*, **40(3)**, 411–421.

45. Djoumbou Feunang, Y., Eisner, R., Knox, C., Chepelev, L., Hastings, J., Owen, G., Fahy, E., Steinbeck, C., Subramanian, S., Bolton, E., Greiner, R. and Wishart, D. S. (2016) Classy Fire: Automated chemical classification with a comprehensive, computable taxonomy. *Journal of Cheminformatics*, **8(1)**, 1–20.

46. Dührkop, K., Nothias, L. F., Fleischauer, M., Reher, R., Ludwig, M., Hoffmann, M. A., Petras, D., Gerwick, W. H., Rousu, J., Dorrestein, P. C. and Böcker, S. (2021) Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. *Nature Biotechnology*, **39(4)**, 462–471.

47. Rutz, A., Sorokina, M., Galgonek, J., Mietchen, D., Willighagen, E., Gaudry, A., Graham, J. G., Stephan, R., Page, R., Vondrášek, J., Steinbeck, C., Pauli, G. F., Wolfender, J. L., Bisson, J. and Allard, P. M. (2022) The LOTUS initiative for open knowledge management in natural products research. *eLife*, **11**, e70780.

48. Frank, A. M., Bandeira, N., Shen, Z., Tanner, S., Briggs, S. P., Smith, R. D. and Pevzner, P. A. (2008) Clustering millions of tandem mass spectra. *Journal of Proteome Research*, **7(1)**, 113–122.

49. Tsugawa, H., Rai, A., Saito, K. and Nakabayashi, R. (2021) Metabolomics and complementary techniques to investigate the plant phytochemical cosmos. *Natural Product Reports*, **38(10)**, 1729–1759.

50. Sorokina, M. and Steinbeck, C. (2020) Review on natural products databases: Where to find data in 2020. *Journal of Cheminformatics*, **12(1)**, 1–51.

51. Nakamura, K., Shimura, N., Otabe, Y., Hirai-Morita, A., Nakamura, Y., Ono, N., Ul-Amin, M. A. and Kanaya, S. (2013) KNApSAcK-3D: A three-dimensional structure database of plant metabolites. *Plant and Cell Physiology*, **54(2)**, e4.

52. Banerjee, P., Erehman, J., Gohlke, B. O., Wilhelm, T., Preissner, R. and Dunkel, M. (2015) Super Natural II-a database of natural products. *Nucleic Acids Research*, **43(Database issue)**, D935–D939.

53. Bajusz, D., Rácz, A. and Héberger, K. (2015) Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations?. *Journal*

of Cheminformatics, **7**, 20.

54. Huber, F., van der Burg, S., van der Hooft, J. J. J., & Ridder, L. (2021) MS2DeepScore: a novel deep learning similarity measure to compare tandem mass spectra. *Journal of Cheminformatics*, **13**, 84.

55. Rojas-Cherto, M., Peironcely, J. E., Kasper, P. T., van der Hooft, J. J. J., de Vos, R. C. H., Vreeken, R., Hankemeier, T. and Reijmers, T. (2012) Metabolite identification using automated comparison of high-resolution multistage mass spectral trees. *Analytical Chemistry*, **84(13)**, 5524–5534.

56. Backman, T. W. H., Cao, Y. and Girke, T. (2011) ChemMine tools: An online service for analyzing and clustering small molecules. *Nucleic Acids Research*, **39(2)**, W486–W491.

57. Böcker, S., Broeckling, C., Schymanski, E. and Zamboni, N. (2020) *Computational metabolomics: From cheminformatics to machine learning*, Schloss Dagstuhl – Leibniz Center for Informatics, Wadern, Germany.

58. Stancliffe, E., Schwaiger-Haber, M., Sindelar, M. and Patti, G. J. (2021) DecoID improves identification rates in metabolomics through database-assisted MS/MS deconvolution. *Nature Methods*, **18(7)**, 779–787.

59. do Nascimento Marinho, R. F., Angrisani, B. R. P., Macedo, A. L., de Lima Moreira, D., Ribeiro, C. M. R., Vasconcelos, T. R. A. and Valverde, A. L. (2020) ^{1}H and ^{13}C NMR spectral data of neolignans isolated from *Piper* species. *Current Organic Chemistry*, **24(14)**, 1527–1554.

60. Maria S. P. Guilhon, G., Conserva, L. M., Guilherme, J., Maia, S., Yoshida, M. and Gottlieb, O. R. (1992) Bicyclo[3.2.1]octanoid neolignans from *Licaria brasiliensis*. *Phytochemistry*, **31(8)**, 2847–2850.

61. Chase, M. W., Christenhusz, M. J. M., Fay, M. F., Byng, J. W., Judd, W. S., Soltis, D. E., Mabberley, D. J., Sennikov, A. N., Soltis, P. S., Stevens, P. F., Briggs, B., Brockington, S., Chautems, A., Clark, J. C., Conran, J., Haston, E., Möller, M., Moore, M., Olmstead, R. and Weber, A. (2016) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*, **181(1)**, 1–20.

Supplementary Materials

Table S1. URLs for a more detailed view of the MS/MS peaks shared between the experimental and reference spectra.

Metabolite	Adduct	URL
Raffinose	$[M+Na]^+$	https://bit.ly/Prubropositive1
Adenosine	$[M+H]^+$	https://bit.ly/Prubropositive2
5'-Deoxy-5'-(methylsulfinyl)adenosine	$[M+H]^+$	https://bit.ly/Prubropositive3
Sucrose	$[M–H]^–$	https://bit.ly/Prubronegative1
Raffinose	$[M–H]^–$	https://bit.ly/Prubronegative2
Citric acid	$[M–H]^–$	https://bit.ly/Prubronegative3
Syringin	$[M+Na]^+$	https://bit.ly/Prubropositive4
3-[3,4-Dihydroxy-6-(hydroxymethyl)-5-[3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxyoxan-2-yl]oxy-2-(3,4-dihydroxyphenyl)-5,7-dihydroxychromen-4-one	$[M–H]^–$	https://bit.ly/Prubronegative4
Luteolin-6-*C*-glucoside	$[M–H]^–$	https://bit.ly/Prubronegative5
Luteolin-8-*C*-glucoside	$[M+H]^+$	https://bit.ly/Prubropositive5
(2*R*,3*S*,4*S*,5*R*,6*R*)-2-[[(2*S*,3*R*,4*R*)-3,4-Dihydroxy-4-(hydroxymethyl)oxolan-2-yl]oxymethyl]-6-(2-phenylethoxy)oxane-3,4,5-triol	$[M–H]^–$	https://bit.ly/Prubronegative6
Quercetin-3-*O*-vicianoside	$[M–H]^–$	https://bit.ly/Prubronegative7
Vitexin-2''-*O*-rhamnoside	$[M–H]^–$	https://bit.ly/Prubronegative8
Quercetin-3-*O*-rutinoside	$[M–H]^–$	https://bit.ly/Prubronegative9
Quercetin-3-*O*-glucoside	$[M–H]^–$	https://bit.ly/Prubronegative10
Apigenin-8-*C*-glucoside	$[M–H]^–$	https://bit.ly/Prubronegative11
Apigenin-8-*C*-glucoside-2'-rhamnoside	$[M+H]^+$	https://bit.ly/Prubropositive6
Quercetin-3-*O*-rutinoside	$[M+Na]^+$	https://bit.ly/Prubropositive7
Quercetin-3-*O*-rutinoside	$[M+H]^+$	https://bit.ly/Prubropositive8
3-[(2*S*,3*R*,4*S*,5*S*,6*R*)-4,5-Dihydroxy-6-(hydroxymethyl)-3-[(2*S*,3*R*,4*S*,5*R*)-3,4,5-trihydroxyoxan-2-yl]oxyoxan-2-yl]oxy-5,7-dihydroxy-2-(4-hydroxyphenyl)chromen-4-one	$[M–H]^–$	https://bit.ly/Prubronegative12
Kaempferol-3-*O*-rutinoside	$[M–H]^–$	https://bit.ly/Prubronegative13
Kaempferol-3-*O*-glucoside	$[M–H]^–$	https://bit.ly/Prubronegative14
Isorhamnetin-3-*O*-glucoside	$[M–H]^–$	https://bit.ly/Prubronegative15
Rhamnetin	$[M–H]^–$	https://bit.ly/Prubronegative16
1-(9*Z*,12*Z*-Octadecadienoyl)-2-hydroxy-sn-glycero-3-phosphocholine	$[M–H]^–$	https://bit.ly/Prubronegative17
13*S*-Hydroxy-9*Z*,11*E*,15*Z*-octadecatrienoic acid	$[M+H–H_2O]^+$	https://bit.ly/Prubropositive9
1-Hexadecanoyl-sn-glycero-3-phospho-(1'-*myo*-inositol)	$[M–H]^–$	https://bit.ly/Prubronegative18
1,2-Di-(9*Z*,12*Z*,15*Z*-octadecatrienoyl)-sn-glycero-3-phosphocholine	$[M+H]^+$	https://bit.ly/Prubropositive10
2,3-Dihydroxypropyl (9*Z*,12*Z*,15*Z*)-octadeca-9,12,15-trienoate	$[M+H–H_2O]^+$	https://bit.ly/Prubropositive11
Pyropheophytin a	$[M+H]^+$	https://bit.ly/Prubropositive12
Pheophytin a	$[M+H]^+$	https://bit.ly/Prubropositive13

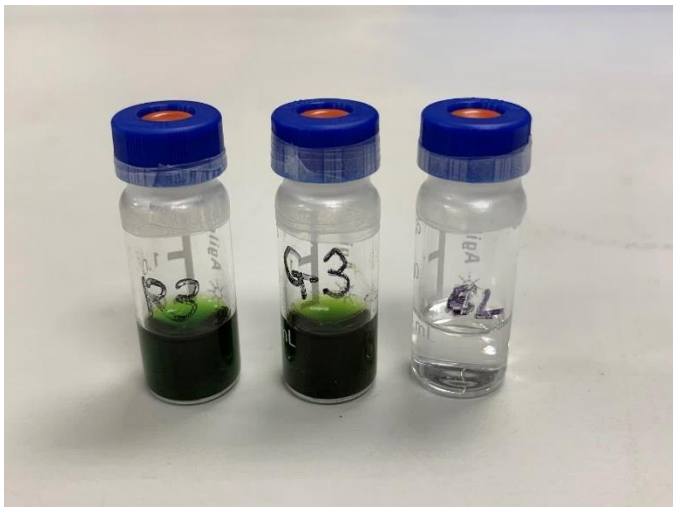

Figure S1. Filtered (0.22 µm) *Piper rubro-venosum* methanolic leaf extracts and LC-MS grade methanol (extraction solvent blank). Left: maroon-underside *P. rubro-venosum* methanolic leaf extract, middle: green-underside *P. rubro-venosum* methanolic leaf extract and right: methanol.

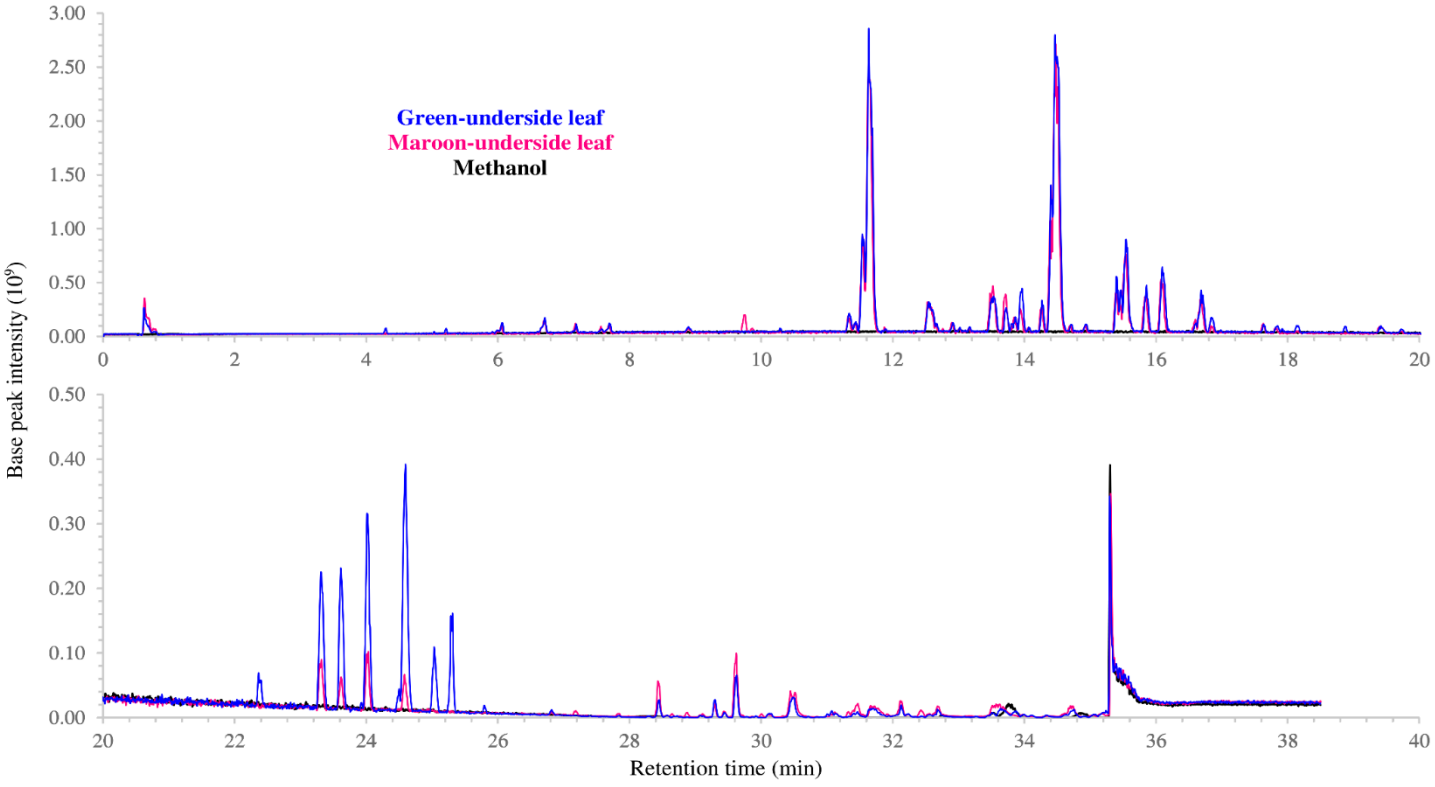

Figure S2. Overlaid and zoomed-in positive ionization LC-MS BPCs of *Piper rubro-venosum* methanolic leaf extracts and methanol.

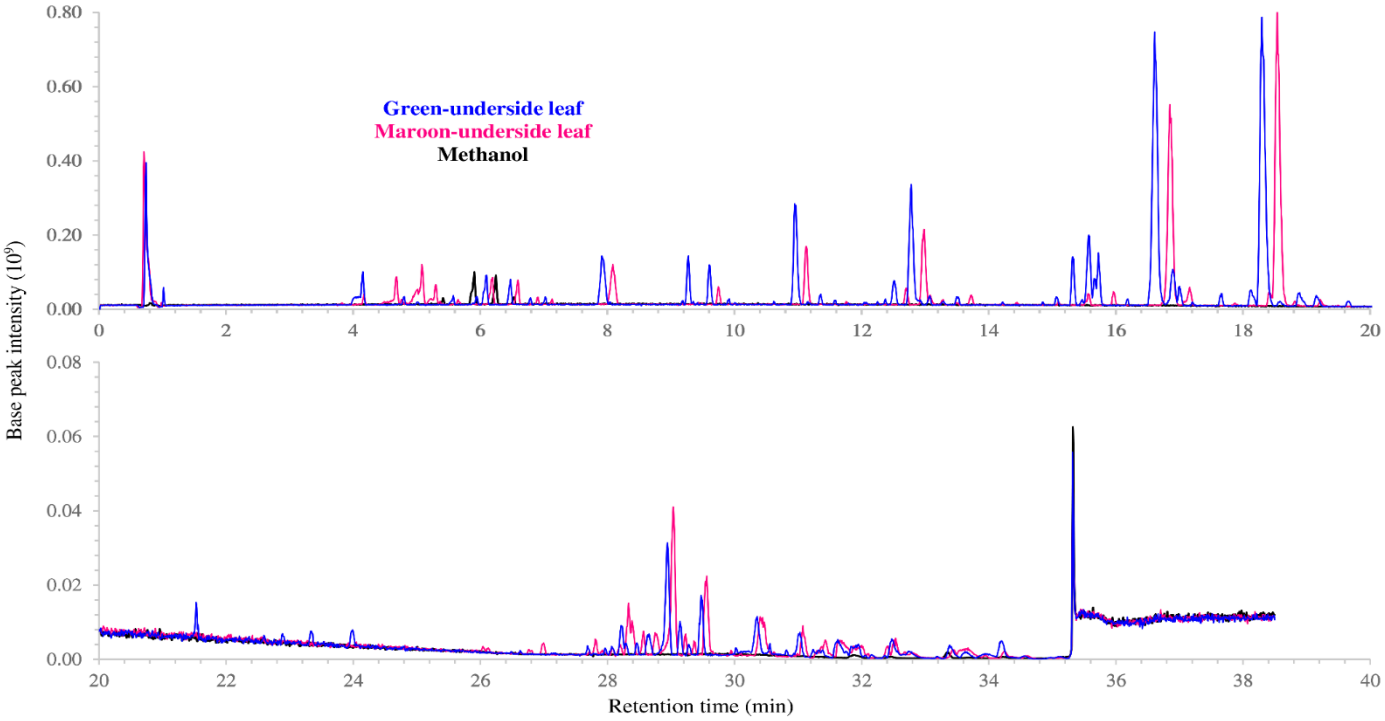

Figure S3. Overlaid and zoomed-in negative ionization LC-MS BPCs of *Piper rubro-venosum* methanolic leaf extracts and methanol.

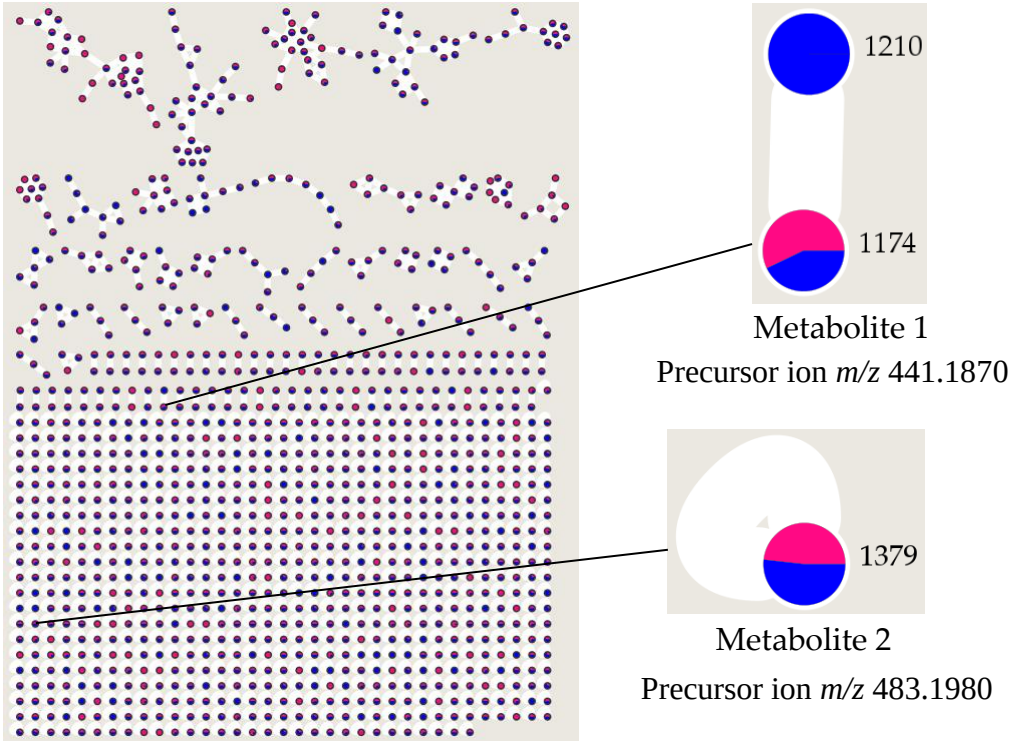

Figure S4. Nodes of metabolites 1 and 2 (node ID 1174 and 1379) in the classical molecular network (MN) of positive ionization BPC of *Piper rubro-venosum* methanolic leaf extracts. Blue and rose colours in the nodes represent green-underside and maroon-underside leaf, respectively. Cytoscape session files for full view of the MN are available at https://bit.ly/Prubropositive (for positive ionization BPC) and https://bit.ly/Prubronegative (for negative ionization BPC).

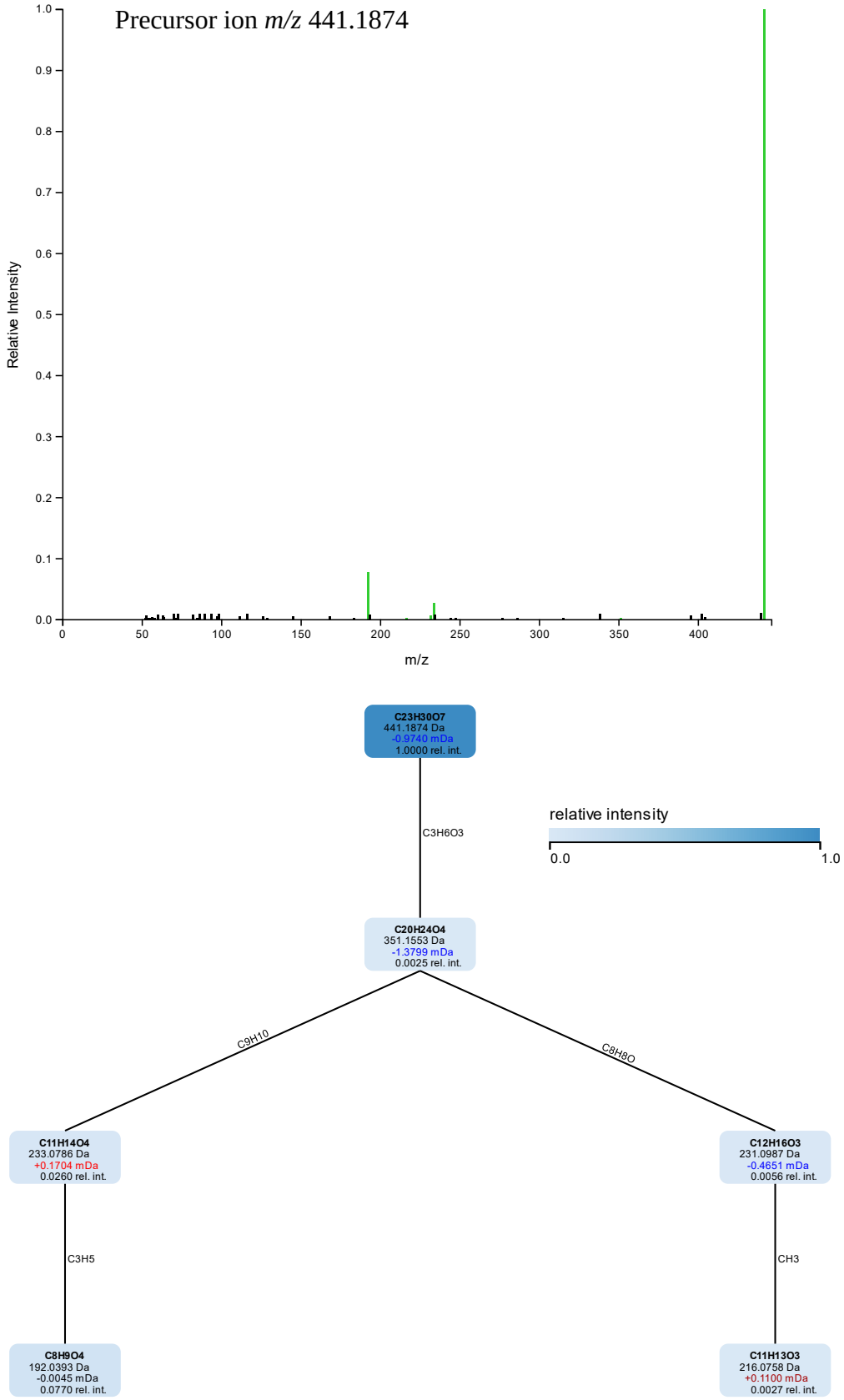

Figure S5. Pre-processed positive ionization fragmentation spectrum of metabolite 1 (top) and best fragmentation tree (bottom) that explains six (6) ion peaks (in green) in the fragmentation spectrum. *M/z* error range of explained ion peaks: -3.9–0.7 ppm.

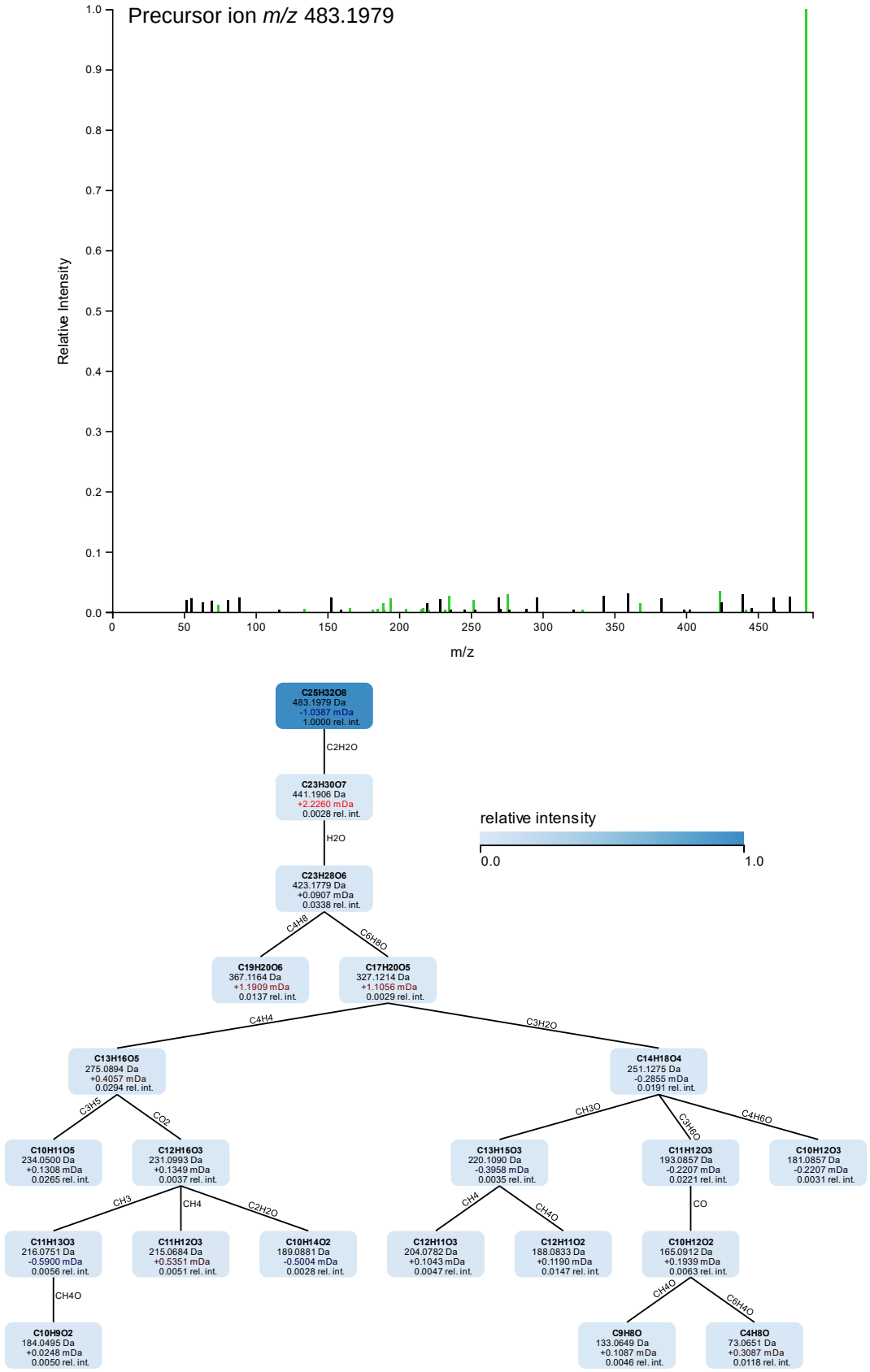

Figure S6. Pre-processed positive ionization fragmentation spectrum of metabolite 2 (top) and best fragmentation tree (bottom) that explains 21 ion peaks (in green) in the fragmentation spectrum. *M/z* error range of explained ion peaks: -2.7–5.1 ppm.

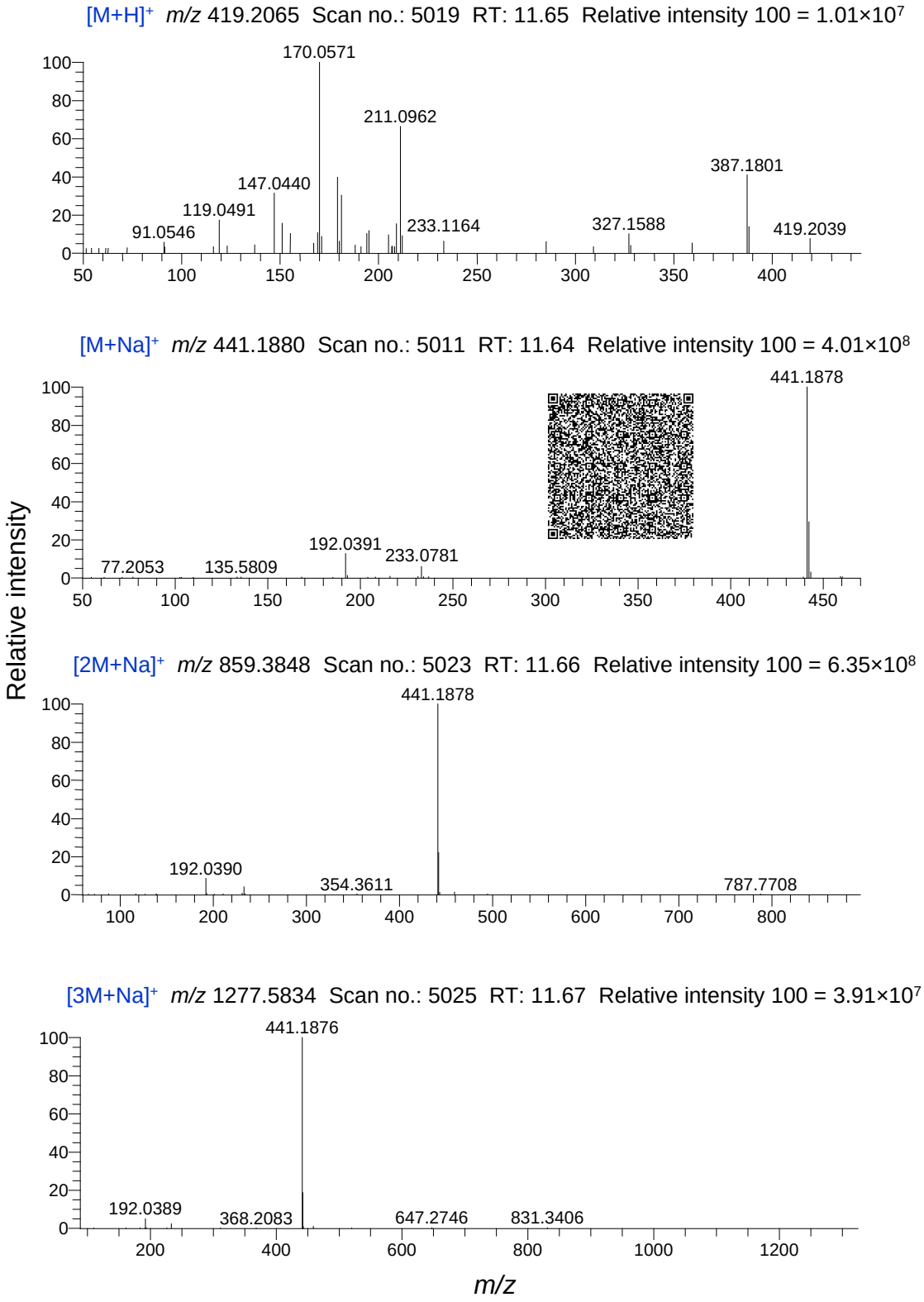

Figure S7. Raw positive ionization MS/MS spectra of protonated adduct, sodiated monomer, dimer, and trimer of metabolite 1 in *P. rubro-venosum* methanolic leaf extract.

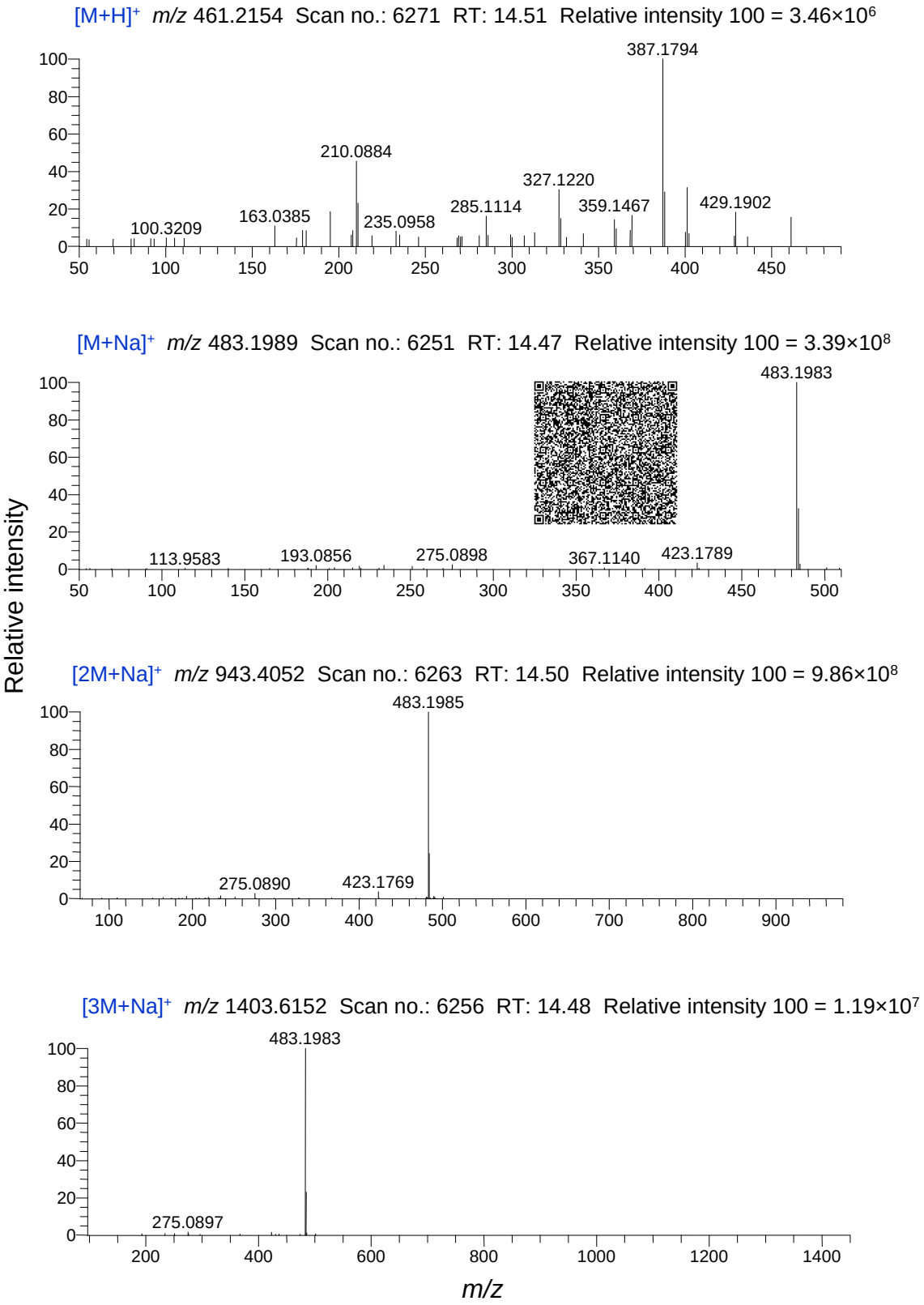

Figure S8. Raw positive ionization MS/MS spectra of protonated adduct, sodiated monomer, dimer, and trimer of metabolite 2 in *P. rubro-venosum* methanolic leaf extract.