Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

An integrated strategy for the comprehensive profiling of the chemical constituents of *Aspongopus chinensis* using UPLC-QTOF-MS combined with molecular networking

Fengyu Zhang[a,b], Bichen Li[a,b], Ying Wen[a,b], Yanyang Liu[a,b], Rong Liu[a,b], Jing Liu[a,b], Shao Liu[a,b] and Yueping Jiang[a,b]

[a]Department of Pharmacy, Xiangya Hospital, Central South University, Changsha, China; [b]National Clinical Research Centre for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China

ABSTRACT

Context: The extracts of *Aspongopus chinensis* Dallas (Pentatomidae), an insect used in traditional Chinese medicine, have a complex chemical composition and possess multiple pharmacological activities.

Objective: This study comprehensively characterizes the chemical constituents of *A. chinensis* by an integrated targeted and untargeted strategy using UPLC-QTOF-MS combined with molecular networking.

Materials and methods: The ultra-performance liquid chromatography-tandem quadrupole time-of-flight mass spectrometry (UPLC-QTOF-MS) combined with molecular networking-based dereplication was proposed to facilitate the identification of the chemical constituents of aqueous and ethanol extracts of *A. chinensis*. The overall strategy was designed to avoid the inefficiency and costliness of traditional techniques. The targeted compounds discovered in the *A. chinensis* extracts were identified by searching a self-built database, including fragment ions, precursor ion mass, and other structural information. The untargeted compounds were identified by analyzing the relationship between different categories, fragmentation pathways, mass spectrometry data, and the structure of the same cluster of nodes within the molecular network. The untargeted strategy was verified using commercial standard samples under the same mass spectrometry conditions.

Results: The proposed integrated targeted and untargeted strategy was successfully applied to the comprehensive profiling of the chemical constituents of aqueous and ethanol extracts of *A. chinensis*. A total of 124 compounds such as fatty acids, nucleosides, amino acids, and peptides, including 74 compounds that were reported for the first time, were identified in this study.

Conclusions: The integrated strategy using LC tandem HRMS combined with molecular networking could be popularised for the comprehensive profiling of chemical constituents of other traditional insect medicines.

Introduction

Aspongopus chinensis Dallas is an insect belonging to the Pentatomidae family (Hemiptera: Heteroptera: Pentatomidae); it has a shape resembling a turtle. It is widely distributed in China, especially in the provinces of Guizhou, Sichuan, Guangxi, and Yunnan (Tan et al. 2019). According to previous reports, the main chemical components in *A. chinensis* are fatty acids, proteins, amino acids, and other nutrients, as well as odour components, nucleosides, and dopamine compounds (Li et al. 2020). It is commonly used as a traditional medicine to relieve pain, warm the stomach, and treat nephropathy (Yan et al. 2019). Modern pharmacology shows that the extracts of *A. chinensis* possess strong protective effects including anticancer (Hou et al. 2012), antibacterial (Wu and Jin 2005), anti-inflammatory (Shi et al. 2014), antioxidant, anticoagulation (Xu 2019), antiulcer, and antifatigue activities (Li et al. 2020). The aqueous extract of *A. chinensis* improves the reproductive ability and protects against reproductive damage (He et al. 2016). A study showed that the fatty oil from the extract of *A. chinensis* has an antiulcer

effect (Hou et al. 2012). Previous studies showed that trichloromethane extracts from *A. chinensis* as well as serum of this insect inhibit the proliferation of human gastric, colon, and breast cancer cells (Fan et al. 2011; Wei et al. 2019; Zhao et al. 2021). Moreover, the aqueous extracts from *A. chinensis* have a significant inhibitory effect on the proliferation, migration, and invasion of the human lung adenocarcinoma cell line A549 (Wu et al. 2020). However, in addition to these widely present components, unique small molecules in the insect are largely unknown. Therefore, in this work, a rapid method was used to examine the compounds included in the aqueous and ethanol extracts of *A. chinensis* and their differences were discussed. Our findings represent a further development in the use of *A. chinensis* extracts.

The Global Natural Products Social Molecular Networking (GNPS, https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp) is an open-access online platform that generates automated molecular networking (MN) for analyzing mass spectrometry (MS/MS) datasets (Wang et al. 2016). MN is useful to speculate

---

CONTACT Shao Liu ✉ liushao999@csu.edu.cn; Yueping Jiang ✉ jiangyueping@csu.edu.cn

on the properties of multiple unannotated compounds in complex matrices by relating them to known compounds, especially structural analogues, based on the mass spectrum information and fragmentation pathway (Zhao et al. 2022). This platform shows strong integration and classification abilities of information obtained with MS, playing an important role in the rapid identification of compounds on a large scale and the discovery of novel framework compounds. Since molecular networking could identify the types of new compounds, experiments could be designed to isolate target compounds based on the characteristics of the identified type of compound. Thereby, the procedure for discovering new compounds can be simplified and the purpose of targeted separation can be achieved. Furthermore, the time necessary to discover new active compounds is greatly shortened using MN. At present, MN has been successfully used in the discovery of novel natural products, microbes, fungi as well as in marine life research (Chen et al. 2021; Rodrigues et al. 2022; Wang et al. 2022). The MN combined with UPLC-QTOF-MS is mostly used to analyze the metabolic components of fungi and one type of components in plants used in traditional Chinese medicines (TCMs) (Messaili et al. 2020; Santos et al. 2021; Wang et al. 2021), for example, dihydrochalcones in Star Fruit (Wang et al. 2021).

UPLC-QTOF-MS is a powerful technique enabling the rapid characterization of the components in herbal medicines, resulting in an accurate mass determination and comprehensive MS data. This technique is widely used in the analysis of complex samples thanks to its high resolution and sensitivity (Huang et al. 2021). However, the processing of a large amount of data obtained from MS is time-consuming and complex. Several processing strategies have been applied to handle large amounts of data for a rapid elucidation of the structure of the non-targeted and complex constituents of TCMs including multifold characteristic ion filtering combined with statistical analysis (Jiang et al. 2019), key ion filtering, high-resolution neutral loss filtering, diagnostic ion filtering, high-resolution diagnostic product ions/neutral loss filtering, mass defect and fragment filtering, and pharmacophore filtering (Jiang et al. 2018). However, none of these strategies can rapidly, directly and comprehensively profile non-targeted constituents or metabolomics in non-targeted mass spectral data without the help of reference standards. Some strategies of processing data, such as combined diagnostic fragment ion filtering with reverse diagnostic fragment loss filtering and structural features guided 'fishing' can profile non-targeted constituents or metabolomics in non-targeted mass spectral data without the need for reference standards (Liu et al. 2019; Wang et al. 2020). However, while these strategies are accurate when analyzing a certain type of compound in TCMs, they tend to be time-consuming and unclear when multiple classes of components in TCMs are analyzed.

At present, most mass spectra in the GNPS database are natural products of microorganisms and marine organisms, while the mass spectra of plant-source compounds are limited. Therefore, the database may not be effective for some natural products only derived from plants. However, the GNPS database was quite suitable for the identification of chemical constituents in traditional insect medicines, because the insects' TCMs mainly contain these compounds such as fatty acids, amino acids nucleosides as well as dopamine compounds that were contained in the GNPS database. The chemical composition analysis of insect TCMs is few investigated. MN combined with UPLC-QTOF-MS is a powerful method for rapid analysis of the chemical constituents of medicines from insects in TCMs. To our knowledge, this is the first systematic and comprehensive method for analytical screening which allows the rapid analysis of the chemical constituents of medicines from insects in TCM.

Therefore, in this study, a rapid and practical UPLC-QTOF-MS method was combined with a reliable and powerful data processing approach, molecular network and an in-house database. This method simultaneously recognized known and unknown chemical constituents of the traditional Chinese medicine from the insect *A. chinensis* in a short time. Our results demonstrated that this strategy could be suitable for the comprehensive profiling of complex components of TCMs without the help of reference standards.

Materials and methods

Reagents and materials

Acetonitrile (LC-MS grade, Cat. no. JA 105030) and methanol (HPLC grade, Cat. no. I1136107 107) were purchased from Merck (Darmstadt, Germany). LC-MS grade formic acid was purchased from Sigma-Aldrich (St. Louis, MO, USA, Cat. no. 197147). Distilled water was purchased from Watsons Distilled Water Company (Hong Kong, China, Cat. no.202107291357070E3). Methanol at 99.5% (Analytical reagent, Cat. no. 20211020801) and ethanol at 99.5% (Analytical reagent, Cat. no. 2020032602) were purchased from Chengdu Kelong Chemical Co., Ltd. (Chengdu, Sichuan, China). *Aspongopus chinensis* was purchased from the medicinal materials market of Bozhou City in 2020 (Bozhou City, Anhui, China) and authenticated by Prof. Shao Liu, Department of Pharmacy, Xiangya Hospital, Central South University. The voucher specimen of *A. chinensis* (ID 20200320) was deposited at the authors' laboratory. Cyclo (Val-Val-) (Cat. no. S68110) at 95% was purchased from Acmec Technology Co. Ltd. (Shanghai, China); kynurenic acid (Cat. no. B50754, HPLC $\geq$ 98%) was purchased from Yuanye Technology Co. Ltd. (Shanghai, China). *N*-(2-Hydroxyethyl) adenosine (Cat. no. Lc0328024) at 98.39% was purchased from Leyan (Shanghai Haohong Biomedical Technology Co. Ltd., Shanghai, China). 1,2,3,4-tetrahydro-β-carboline-3-carboxylic acid (Cat. no. C12570529) at 97% was purchased from Macklin Technology Co. Ltd. (Shanghai, China).

Sample preparation

The processed product from *A. chinensis* (1 kg) was pulverized, and extracted with 4 L water two times, each time for 1 h (40 kHz, 210 W, 40 °C). Next, distilled water was filtered to obtain the aqueous extract and the medicine dregs. Then, the medicine dregs were extracted with 80% ethanol two times, each time for 1 h. The 80% ethanol extracts were filtered. The aqueous extracts and ethanol extracts were vacuum freeze-dried in a freeze dryer (Chaist, Alpha 1-2 LD plus, Germany) and quantified. Each of the extracts of *A. chinensis* powder (10 mg) was accurately weighed, 1 mL methanol was added, and then the mixture was shaken for 1 min at 25 °C. A test solution with a mass concentration of 10 mg/mL was obtained by filtration through a 0.22 μm membrane (Millipore, Merck Millipore Ltd., Germany).

UPLC-QTOF MS analysis

The chromatographic evaluation was performed using an Agilent 1290 series UPLC system (Agilent Corp., Santa Clara, CA, USA)

equipped with a binary pump, micro degasser, autosampler, and temperature-controlled column compartment. The sample was separated on an Agilent C clipsePlusC18 (1.8 μm, 2.1 × 100 mm, Agilent Corp., Santa Clara, USA). The mobile phase was composed of two solvents (A and B). The mobile phase (A) was aqueous formic acid (0.1%, v/v), and the mobile phase (B) was acetonitrile. The following mobile phase gradient was selected: 0–2 min: 2% B; 2–5 min: 2–10% B; 5–20 min: 10–30% B; 20–23 min: 30–50% B; 23–25 min: 50–60% B; 25–30 min: 60% B. *Aspongopus chinensis* mainly contain alkaloids and amino acids which are easier to ionize in positive ion mode. Therefore, a positive mode was used for the ionization of *A. chinensis* components. The flow rate was 0.3 mL/min, the injection volume was 4 μL, and the column temperature was 30 °C.

The separated components were passed through an Agilent 6545 A Q-TOF mass spectrometer (Agilent Corp., Santa Clara, CA, USA), equipped with an ESI interface. The operating parameters were as follows. The drying N_2 gas flow rate was 8 L/min, the temperature was 325 °C, the nebulizer was at 35 Psi, the capillary was 4000 V, MS/MS data acquisition mode was Auto MS/MS, the max precursor per cycle was 5, and the acquisition time was 250 ms/spectrum. The samples were analyzed in positive ion mode, and mass spectra data were recorded across a range of 50–1700 m/z. The reference mass of 121.0509 (purine) and 922.0098 (HP-0921) were utilized for the internal mass calibration during runs in positive ion mode. Fixed collision energies of 10.00, 20.00, 40.00 V were chosen at a scan rate of 4.0 spectra/s using Auto MS/MS data acquisition with a medium MS/MS isolation width.

Molecular network design

The original data of MS/MS spectra were converted into an mzML format that contains all the information regarding the analysis. Then, these data were uploaded to the GNPS platform for analysis (Chen et al. 2021; Rodrigues et al. 2022; Wang et al. 2022). Finally, the MS/MS spectra were compared pairwise to search for spectral similarities, including the same fragment ions and/or neutral losses. The optimum parameters were the following, including parent mass and fragments: tolerance 0.02 Da; cosine score ≥0.7; matched peaks ≥6; network TopK 10; maximum connected component size 100; minimum cluster size 1, No run MSCluster. The results were downloaded and exported

to be visualized on Cytoscape 3.8.2 software [https://cytoscape.org].

The identification of the compounds was supported by spectral libraries of the GNPS. The MS/MS spectra of *A. chinensis* compounds were compared to the MS/MS spectra of the compounds contained in the GNPS library platform, using the following parameters, including library search min matched peaks of 6; score threshold of 0.7; maximum analog search mass difference of 100.

Results and discussion

Frame of integrated targeted and untargeted strategy

The objective of this work was to target and identify, tentatively, components from *A. chinensis* extract. The total ion chromatogram obtained with MS is shown in Figure 1. This study proposed an integrated interpretation strategy illustrated in Figure 2 in order to reach the purpose of efficient organization and annotation of data. A comprehensive analysis based on UPLC-QTOF-MS and GNPS platform, easy to perform, was proposed for targeted screening of known compounds. MN was utilized for structural classification by non-targeted data organization based on MS/MS spectral similarity. Starting with the nodes of known compounds, the adjacent unreported compounds in the visualized networks were identified by detailed analysis of the MS/MS spectra, and online database searching.

The targeted identification strategy was performed as follows. First, Pubmed (https://pubmed.ncbi.nlm.nih.gov/), PubChem (https://pubchem.ncbi.nlm.nih.gov/), China National Knowledge Infrastructure (CNKI; https://www.cnki.net/) and other Chinese and international databases. The related data were researched in literature to build an information database by results from high-resolution chromatography and mass spectrometry, including m/z ratio, ultraviolet absorption characteristics, relative molecular mass, fragments ion mass spectrometry, and accurate $[M + H]^+$. Second, the library matching function of the Agilent MassHunter Qualitative Analysis B08.00 software was used to match the $[M + H]^+$ and the fragment ion with the data from the TCM and Metlin databases that are built into the software. Chemical formulas with a score of ≥90 were selected for the next step of identification. Third, the MS/MS data obtained from the aqueous extract/ethanol extract matched with the data in the information
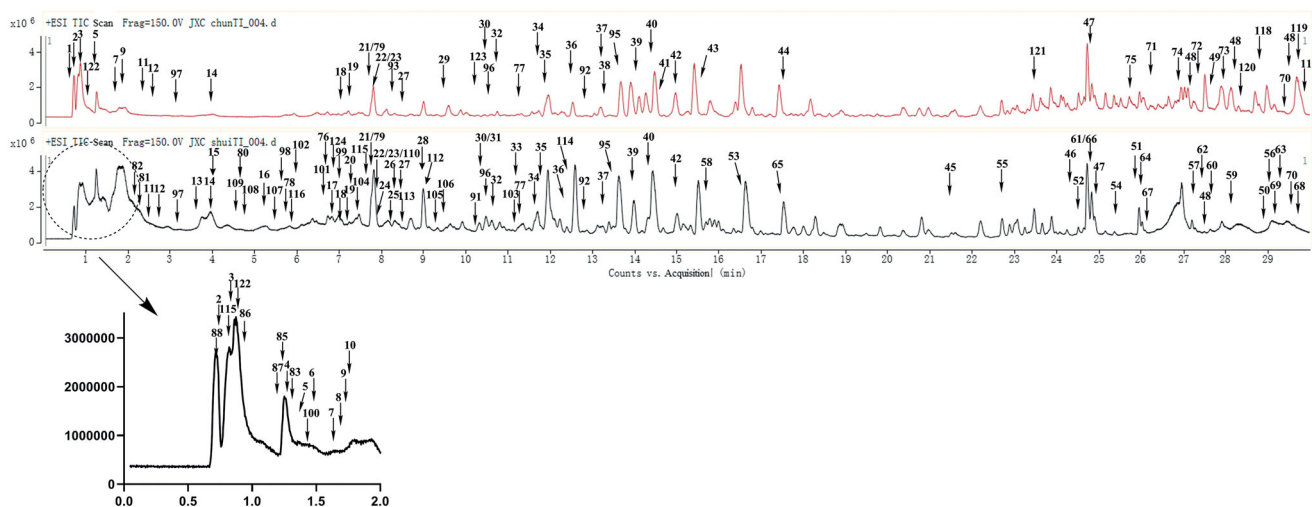
Figure 1. Total ion Chromatogram (TIC) of extract of *Aspongopus chinensis* in positive ion mode.
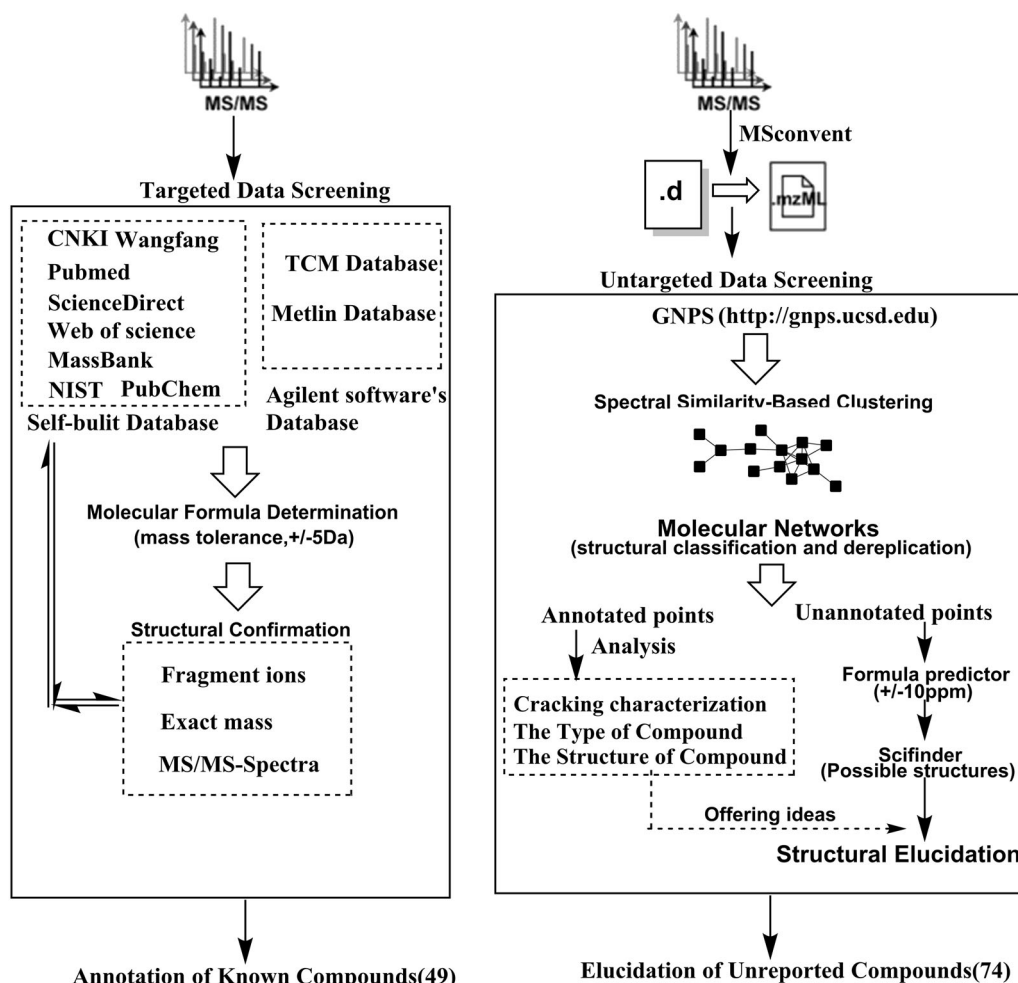
Figure 2. *Aspongopus chinensis* compounds identification strategy.

database established in steps (1) and (2) identified the compounds found in the *A. chinensis* extract.

The untargeted identification strategy was as follows. The identification of unannotated points was based on the characterization of a similar mass spectrum of the same type of compound, and each MN was performed using the same type of compound. The classification of annotated points was analyzed in order to determine the type of network where they were located. Then, the unannotated peaks were processed according to the following steps. First, the element composition of unannotated points was automatically deduced by the formula predictor of the Qualitative Navigator B.08.00 software. The elements carbon (C), hydrogen (H), oxygen (O), nitrogen (N), phosphorus (P), and sulphur (S) were selected to calculate the elemental composition of the compounds, and only consistent chemical formulas were considered. The maximum MS mass tolerance was set at $\leq 5\,ppm$. Second, the molecular formula predicted by the molecular predictor was uploaded into the SciFinder database (https://origin-scifinder.cas.org/scifinder/) to search for potential matching compounds. If they were found, the compound category was set to further filter the potential structural formula. Then, the fracture modes of the annotated points were summarised in the same network. The chromatographic characteristics and the MS/MS data, as well as the bibliographic information, were utilized for the identification of compounds. Potential structural formulas were verified by fragment ions and fracture modes to determine the structural formula of the unknown compound. Finally, Cytoscape 3.8.2 was used to visualize and edit the entire MN.

Verification was performed as follows. Four potentially active compounds deduced by MN were chosen as standard reference substances. Their primary/secondary MS/MS data were obtained under the corresponding mass spectrometry conditions. Then, the retention time of the substance used as standard reference and the second-level mass spectrometry fragments were compared to the deduced unannotated points to confirm whether it was the same compound as the unannotated points or not.

UPLC-QTOF MS analysis result

Targeted identification result

UPLC-QTOF-MS was utilized to evaluate components of *A. chinensis* extracts. The *A. chinensis* extracts could be analyzed within 30 min. The molecular formula was accurately assigned, within a mass error of 5 ppm. The fragment ions were used to further confirm the chemical structure. In total, 49 compounds were identified from the *A. chinensis* extracts on the basis of the MS^2 fragmentation information, TCM database, Metlin database, and literature. These compounds included nucleosides and their analogues, dopamine derivatives, sesquiterpenes, alkaloids, and cyclic peptides. The total ion chromatogram of the fraction of *A. chinensis* extracts in positive ion mode is shown in Figure 1. The information on these 49 compounds, including the compound name, retention time, formula, precursor ions, and fragment ions are listed in Table 1.

Table 1. Characterisation of constituents in aqueous/ethanol extract of *Aspongopus chinensis* Dallas by UPLC-Q-TOF-MS/MS and MS^2 analysis in positive ion mode.

Number	t_R (min)	[M + H]$^+$	MS^2	Formula	Identification	Mode	Source
1	0.757	243.0975	60.0812/84.080/1104.1069/124.9913	$C_{10}H_{14}N_2O_5$	Thymidine	[M + H]$^+$	Alcohol
2	0.828	148.0610	56.0495/57.0336/84.0441/ 85.0501/102.0547	$C_5H_9NO_4$	Glutamic acid	[M + H]$^+$	Water/alcohol
3	0.839	252.1084	136.0378/136.0745	$C_{10}H_{13}N_5O$	Deoxyadenosine	[M + H]$^+$	Water/alcohol
4	1.238	126.0550	53.0389 68.0131 108.0442	$C_6H_7NO_2$	6-(Hydroxymethy1) pyridin-3-ol	[M + H]$^+$	Water
5	1.318	123.0553	53.0390 78.0340	$C_6H_6N_2O$	Nicotinamide	[M + H]$^+$	Water/alcohol
6	1.466	259.0925	56.0501 84.0444 124.0764 170.0779	$C_{10}H_{14}N_2O_6$	2′-O-Methyluridine	[M + H]$^+$	Water
7	1.655	153.0407	55.0294 69.0082 81.0091 93.0069 110.0348 136.0144	$C_5H_4N_4O_2$	Xanthine	[M + H]$^+$	Water/alcohol
8	1.707	113.0346	53.9384 70.0656 95.0314	$C_4H_4N_2O_2$	Uracil	[M + H]$^+$	Water/alcohol
9	1.767	245.0768	55.0550 70.0291 113.0347 140.1088 173.1052 214.1260	$C_9H_{12}N_2O_6$	Uridine	[M + H]$^+$	Water/alcohol
10	1.894	86.0600	61.0287 69.0351	C_4H_7NO	2-Pyrrolidinone	[M + H]$^+$	Water
11	2.490	268.1037	55.0290/57.0326/69.0339/87.0017/ 94.0391/109.0488/135.1923	$C_{10}H_{13}N_5O_4$	Adenosine	[M + H]$^+$	Alcohol
12	2.746	144.0655	55.0186 70.0662 104.0496	$C_6H_9NO_3$	N-(2-Hydroxyethyl) succinimide	[M + H]$^+$	Water/alcohol
13	3.632	100.0757	55.0553 72.9383	C_5H_9NO	Valerolactam	[M + H]$^+$	Water
14	3.943	166.0860	51.0233/65.0389/77.0392/91.0548/ 103.0547/120.0802	$C_9H_{11}NO_2$	ʟ-Phenylalanine	[M + H]$^+$	Water/alcohol
15	4.018	138.0912	53.0378/137.0781	$C_8H_{11}NO$	2-Ethyl-3-hydroxy-6-methylpyridine	[M + H]$^+$	Water
16	5.256	127.0502	54.0333/84.0438/110.0222/127.0502	$C_5H_6N_2O_2$	Thymine	[M + H]$^+$	Water
17	6.793	171.1128	55.0551 67.0550 90.9066 113.0350	$C_8H_{14}N_2O_2$	Cyclo(gly-ʟ-leu)	[M + H]$^+$	Water
18	7.071	205.0973	142.0650/130.0651/117.0572/103.0550/ 91.0543/77.0385	$C_{11}H_{12}N_2O_2$	Tryptophan	[M + H]$^+$	Water/alcohol
19	7.337	146.0600	117.0566/90.0499/78.0420	C_9H_7NO	2-Quinolinol	[M + H]$^+$	Water/alcohol
20	7.343	146.0600	51.0238/65.0385/77.0386/91.0543/ 101.0358/118.0641/132.0776/146.0577	C_9H_7NO	Indole-3-aldehyde	[M + H]$^+$	Water/alcohol
21	7.812	162.0550	65.0386/89.0381/116.0488/149.1066	$C_9H_7NO_2$	Indole-β-carboxylic acid	[M + H]$^+$	Water/alcohol
22	7.812	190.0499	190.0489/144.0449	$C_{10}H_7NO_3$	Transtorine	[M + H]$^+$	Water/alcohol
23	7.815	162.0554	51.0226/64.0187/77.0387/89.0389/ 116.0503/117.0519/144.0446	$C_9H_7NO_2$	Indole-3-carboxylic acid	[M + H]$^+$	Water/alcohol
24	7.990	237.1346	55.0549/79.0537/106.0657/118.0653/ 134.0981/144.0817/163.0911/ 178.1226/206.1183	$C_{11}H_{16}N_4O_2$	Asponguanine B	[M + H]$^+$	Water
25	8.364	264.1455	69.0691/97.0648/134.0976/150.0780/ 169.0351/202.0975/247.1415	$C_{12}H_{17}N_5O_2$	Aspongadenine A	[M + H]$^+$	Water
26	8.492	331.1288	62.9825/84.0810/104.9935/153.0236/ 177.0507/199.0298/233.1527/268.1346	$C_{17}H_{18}N_2O_5$	Aspongamides C	[M + H]$^+$	Water
27	8.513	236.1506	55.0536/67.0274/91.0534/123.0911/ 147.1026/164.1044/186.1149	$C_{11}H_{17}N_5O$	Aspongadenine B	[M + H]$^+$	Water/alcohol
28	8.999	136.0616	55.0288/65.0137/67.0297/77.0410/ 92.0243/109.0478	$C_5H_5N_5$	Adenine	[M + H]$^+$	Water
29	9.291	385.1393	355.2017	$C_{20}H_{20}N_2O_6$	Aspongopusamide B	[M + H]$^+$	Alcohol
30	10.571	251.1139	70.0659/86.0962/114.9779/137.0452/ 153.0388/180.1031/206.0983/233.1251	$C_{11}H_{14}N_4O_3$	Asponguanine A	[M + H]$^+$	Water/alcohol
31	10.607	137.0457	55.0292/65.0378/67.0550/82.0409/ 83.0233/92.0236/94.0395/109.0502/ 119.0345120.0383/	$C_5H_4N_4O$	Hypoxanthine	[M + H]$^+$	Water
32	10.704	172.0968	55.0179/84.9577/142.0607	$C_8H_{13}NO_3$	Asponglactam A	[M + H]$^+$	Water/alcohol
33	11.330	226.1074	57.0331/72.9373/105.0670/122.0279/ 156.0803/168.0794/184.1138/210.0652	$C_{11}H_{15}NO_4$	N-[2-(3,4-Dihydroxyphenyl)-2-methoxyethyl]-acetamide	[M + H]$^+$	Water
34	11.755	382.1721	57.0332/96.0800/135.0286/152.0554/ 189.0741/306.1020/232.1181	$C_{16}H_{23}N_5O_6$	Asponguanosines A	[M + H]$^+$	Water/alcohol
35	11.787	382.1721	/	$C_{16}H_{23}N_5O_6$	Asponguanosines B	[M + H]$^+$	Water/alcohol
36	12.291	192.0655	53.0385/65.0378/77.0385/94.0659/ 104.0490/123.0428/146.0927/ 158.0979/176.1336	$C_{10}H_9NO_3$	Aspongopusin	[M + H]$^+$	Alcohol
36	12.303	192.0655	同上	$C_{10}H_9NO_3$	Aspongopusin	[M + H]$^+$	Water
37	13.176	387.1542	150.0561/193.0746/328.1190/269.0816/ 123.0435/137.0586/328.1190	$C_{20}H_{22}N_2O_6$	trans-2-(3′4′-Dihydroxyphenyl)-3-acetylamino-6-(N-acetyl-2″-aminoethyl)-1,4-benzodioxane	[M + H]$^+$	Water/alcohol
38	13.201	194.0788	117.0380/140.9607/178.0494/194.1551	$C_8H_{13}NO_3$	(±)-Asponglactam A	[M + Na]$^+$	Alcohol
39	13.985	385.1393	150.0561/193.0746/328.1190/269.0816/ 123.0435/137.0586/328.1190	$C_{20}H_{20}N_2O_6$	trans-2-(3′4′-Dihydroxyphenyl)-3-acetylamino-7-(N-acetyl-2-amino-ethylene)-1,4-benzodioxane	[M + H]+	Water/alcohol
40	14.442/ 14.398	189.1114	190.15	$C_9H_{16}O_4$	Aspongester A	[M + H]$^+$	Water/alcohol
41	14.851	385.1393	150.2973	$C_{20}H_{20}N_2O_6$	Aspongopusamide A	[M + H]$^+$	Alcohol
42	15.021	387.1542	150.0561/193.0746/328.1190/269.0816/ 123.0435/137.0586/328.1190	$C_{20}H_{22}N_2O_6$	trans-2-(3′4′-Dihydroxyphenyl)-3-acetylamino-7-(N-acetyl-2″-aminoethyl)-1,4-benzodioxane	[M + H]$^+$	Water/alcohol

(continued)

Table 1. Continued.

| Number	t_R (min)	$[M+H]^+$	MS^2	Formula	Identification	Mode	Source
43	15.644	385.1393	150.0558/122.0605/94.0652	$C_{20}H_{20}N_2O_6$	*trans*-2-(3′4′-Dihydroxyphenyl)-3-acetylamino-6-(*N*-acetyl-2″-aminoethylene)-1,4-benzodioxane	$[M+H]^+$	Alcohol
44	17.547	578.2135	60.0438/89.0369/123.0444/150.0552/164.0689/181.0612/193.0720/206.0814/239.0705/269.0791/286.1037/312.0851/328.1181/345.1466/366.7094/401.1060/417.1175/443.1217/459.1297/521.1740	$C_{30}H_{31}N_3O_9$	(±)-Aspongamide A	$[M+H]^+$	Alcohol
45	21.796	111.0441	55.9330 72.9369	$C_6H_6O_2$	1,2-Benzenediol	$[M+H]^+$	Water
46	24.680	285.170	64.0157/213.1140/285.1694/287.6971	$C_{15}H_{24}O_5$	Aspongnoid C	$[M+H]^+$	Water
47	24.925	283.1520	55.0188/139.0764/171.1024	$C_{18}H_{34}O_2$	Oleic acid	$[M+H]^+$	Water/alcohol
48	27.107	255.2323	137.1306/69.0702/55.0546	$C_{16}H_{30}O_2$	(*Z*)-Hexadec-11-enoic acidl	$[M+H]^+$	Alcohol
48	27.368	255.2323	137.1306/69.0702/55.0546	$C_{16}H_{30}O_2$	(*Z*)-Hexadec-11-enoic acid II	$[M+H]^+$	Water
49	27.702	281.2484	55.0546/69.0696/95.0858/133.0987/147.1141	$C_{18}H_{32}O_2$	Linoleic acid	$[M+H]^+$	Alcohol
48	28.454	255.2323	137.1306/69.0702/55.0546	$C_{16}H_{30}O_2$	(*Z*)-Hexadec-11-enoic acid III	$[M+H]^+$	Alcohol
48	29.509	255.2323	137.9628/69.0690/55.0550	$C_{16}H_{30}O_2$	(*Z*)-Hexadec-11-enoic acid IV	$[M+H]^+$	Alcohol
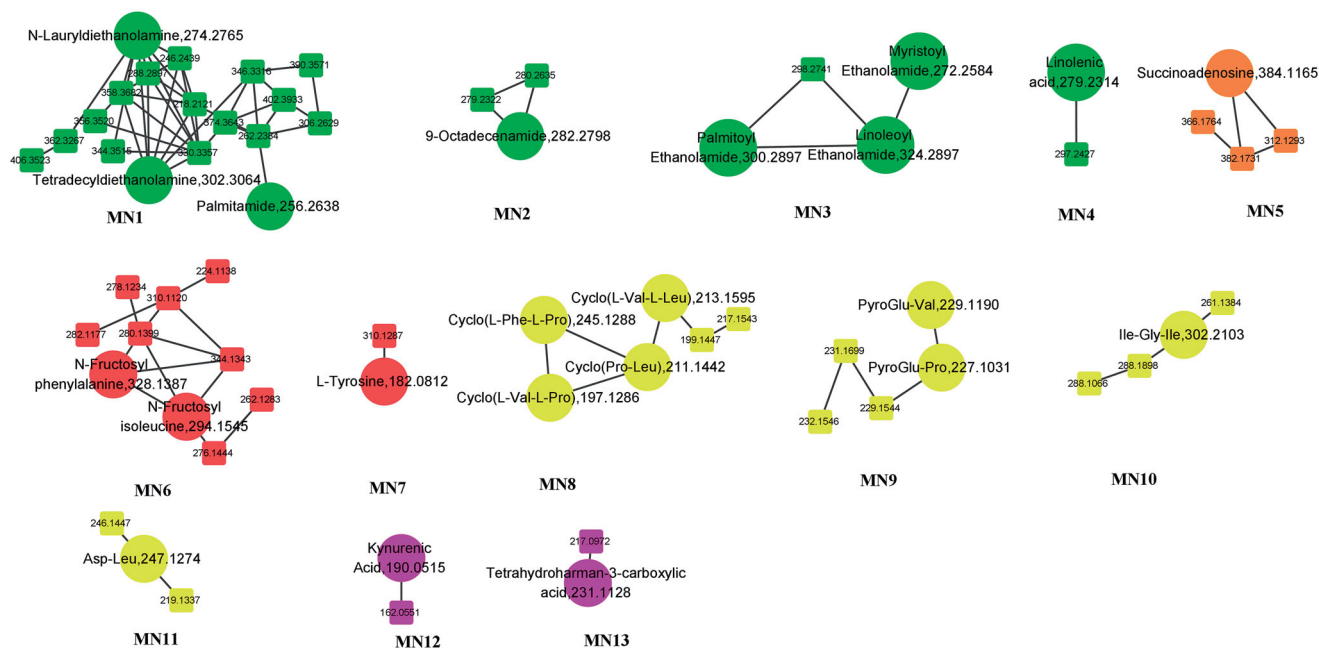
Figure 3. Thirteen analysable clusters obtained through molecular networks. Green represents fatty acid; orange represents nucleoside; red represents amino acid; yellow represents peptide; purple represent other type. The circles represent the annotated point; the squares represent unannotated point.

Untargeted identification results

MN was formed by GNPS, which is able to identify analogues or derivatives of known compounds, based on the relationship between the nodes within the same network, and the relationship between the spectra of the nodes. The relationship among network nodes was visualized using Cytoscape 3.8.2. Thus, MN was used to discover complex crude extracts. The disadvantages of low efficiency, the high cost of using traditional techniques involved in the discovery of traditional compounds, and the repeated separation of known compounds could effectively be avoided. In this study, A total of 13 clusters of *A. chinensis* aqueous and ethanol extracts were analyzed and shown in Figure 3. 75 compounds were obtained by MN analysis, and among them, 74 were reported for the first time. The compound names, retention times, formula, precursor ions, and fragment ions are shown in Table 2.

Analysis of the unannotated points of fatty acid clusters

The common characteristic of fatty acids is the successive loss of H$_2$O (18 Da) in their MS^2 spectra and the cleavage of the carbon chain. MN1 is the representative cluster of fatty acids. Thus, tetradecyl diethanolamine (**51**) was chosen to be the compound that could help investigate the MS^2 fragmentation patterns of fatty acids (Figure 4). The $[M+H]^+$ quasi-molecular ion of *m/z* 302 of tetradecyl diethanolamine (C$_{18}$H$_{39}$NO$_2$) in the positive ion mode can easily be formed. The parent ion could be dehydrated (losing H$_2$O) to produce the fragment ion of *m/z* 284. Then the carbon chain of *m/z* 284 ion can be broken to obtain a fragment at *m/z* 102. The quasi-molecular ion of $[M+H]^+$ could be losing C$_{14}$H$_{30}$NO$_2$ to obtain the fragment of *m/z* 57 or losing the carbon chain to obtain the fragment of *m/z* 106 directly. Then the latter ion could be dehydrated to obtain a fragment of *m/z* 88 or further dehydrate for the production of the fragment

Table 2. The constituents in aqueous/ethanol extract of *Aspongopus chinensis* Dallas annotated by UPLC-Q-TOF-MS/MS and molecular network.

Number	Identification	Formula	$[M + H]^+$	MS^2	Identification	Type	Cluster
50	Palmitamide*	$C_{16}H_{33}NO$	256.2635	57.0703/71.0500/83.0879/88.0749/ 102.0935	Water/alcohol	Fatty acids	MN1
51	Tetradecyldiethanolamine*	$C_{18}H_{39}NO_2$	302.3054	57.0700/70.0651/71.0855/88.0755/ 102.0906/106.0856/284.2933	Water/alcohol	Fatty acids	MN1
52	N-Lauryldiethanolamine*	$C_{16}H_{35}NO_2$	274.2741	57.0710/70.0659/88.0760/ 106.0869/274.2710	Water/alcohol	Fatty acids	MN1
53	2-[2-Hydroxyethyl(2-octoxyethyl)amino]ethanol	$C_{14}H_{31}NO_3$	262.2377	57.0699/70.0638/88.0753/102.0903/ 176.1427/200.2001	Water/alcohol	Fatty acids	MN1
54	N,N-bis(2-Hydroxyethyl)tridecylamine	$C_{17}H_{37}NO_2$	288.2897	57.0703/70.0654/88.0753/102.0916/ 106.0870/270.2827	Water/alcohol	Fatty acids	MN1
55	2,2'-(Decylimino)bisethanol	$C_{14}H_{31}NO_2$	246.2428	57.0699/70.0657/88.0756/ 106.0853/228.2307	Water/alcohol	Fatty acids	MN1
56	Stearyldiethanolamine	$C_{22}H_{47}NO_2$	358.3680	57.0708/70.0664/88.0766/106.0870/ 128.0656/340.3619	Water/alcohol	Fatty acids	MN1
57	N-Hexadecyl diethanolamine	$C_{20}H_{43}NO_2$	330.3367	57.0705/88.0759/106.0868/312.3272	Water/alcohol	Fatty acids	MN1
58	2,2'-(Octylazanediyl)diethanol	$C_{12}H_{27}NO_2$	218.2115	57.0703/70.0652/88.0755/ 106.0861/200.2008	Water/alcohol	Fatty acids	MN1
59	Ethanol,2,2'-(heptadecylimino)bis-	$C_{21}H_{45}NO_2$	344.3523	59.0318/70.0648/88.0750/102.0874/ 160.9885/267.2069 /268.2992/287.1601/312.3217	Water/alcohol	Fatty acids	MN1
60	n-Oleyldiethanolamine	$C_{22}H_{45}NO_2$	356.3523	55.0546/69.0698/70.0652/83.0854/ 88.0759/106.0861/338.3512	Water/alcohol	Fatty acids	MN1
61	16-[bis(2-Hydroxyethyl)amino]hexadecane-1,15-diol	$C_{20}H_{43}NO_4$	362.3265	57.0698/70.0650/88.0751/132.1014/ 256.2611/300.2884	Water/alcohol	Fatty acids	MN1
62	1-[bis(2-Hydroxyethyl)amino]octadecan-2-ol	$C_{22}H_{47}NO_3$	374.3629	57.0698/70.0653/88.0758/ 102.0902/312.3245	Water/alcohol	Fatty acids	MN1
63	1-[bis(2-Hydroxyethyl)amino]icosan-2-ol	$C_{24}H_{51}NO_3$	402.3942	57.0703/70.0657/88.0762/ 102.0918/340.3586	Water/alcohol	Fatty acids	MN1
64	4-[bis(2-Hydroxyethyl)amino]-1-octoxybutan-2-ol	$C_{16}H_{35}NO_4$	306.2639	57.0706/70.0661/86.0970/132.1026/ 146.1172/244.2266	Water/alcohol	Fatty acids	MN1
65	1-[bis(2-Hydroxyethoxy)amino]octadecan-2-ol	$C_{22}H_{47}NO_5$	406.3527	57.0697/70.0644/88.0750/ 300.2846/344.3121	Water/alcohol	Fatty acids	MN1
66	1-[2-[bis(2-Hydroxyethyl)amino]ethoxy]hexadecan-2-ol	$C_{22}H_{47}NO_4$	390.3578	57.0701/70.0659/88.0761/132.1031/ 146.1177/276.1986/328.3216	Water/alcohol	Fatty acids	MN1
67	1-[bis(2-Hydroxyethyl)amino]hexadecan-2-ol	$C_{20}H_{43}NO_3$	346.3316	57.0700/70.0655/88.0755/ 102.0907/284.2947	Water/alcohol	Fatty acids	MN1
68	9-Octadecenamide*	$C_{18}H_{35}NO$	282.2791	57.0706/69.0709/83.0871/97.1022/ 111.1188/121.1017/135.1174/149.1365/ 240/247.2439/282.2825	Water	Fatty acids	MN2
69	Linoleamide(9,12-octadecadienamide)	$C_{18}H_{33}NO$	280.2635	55.0540/57.0696/69.0697/83.0850/ 97.1005/135:52/149.1296/247.2398	Water	Fatty acids	MN2
70	Linolenic acid	$C_{18}H_{30}O_2$	279.2319	57.0701/69.0698/83.0853/97.1006/ 149.1310/163.1474/247.2412	Water	Fatty acids	MN2
71	Myristoyl Ethanolamide*	$C_{16}H_{33}NO_2$	272.2584	57.0698/62.0598/67.0542/71.0852/95.0845	Alcohol	Fatty acids	MN3
72	Linoleoyl Ethanolamide*	$C_{20}H_{37}NO_2$	324.2897	62.0608/67.0552/83.0474/95.0867	Alcohol	Fatty acids	MN3
73	Palmitoyl Ethanolamide*	$C_{18}H_{37}NO_2$	300.2897	57.0700/62.0602/71.0857/283.2630	Alcohol	Fatty acids	MN3
74	Palmitoleoyl Ethanolamide	$C_{18}H_{35}NO_2$	298.2741	55.0550/62.0605/81.0701/281.2434	Alcohol	Fatty acids	MN3
70	Linolenic acid*	$C_{18}H_{30}O_2$	279.2319	55.0545/67.0545/81.0699/95.0855/ 109.1012/123.1175/137.1299/149.0237	Alcohol	Fatty acids	MN4
75	Coriolic acid	$C_{18}H_{32}O_3$	297.2424	55.0530/67.0546/81.0695/95.057/ 147.1158/279.2315	Alcohol	Fatty acids	MN4
76	Succinoadenosine*	$C_{14}H_{17}N_5O_8$	384.1150	136.0606/162.0765/188.0560/192.0500/ 206.0669/234.0618/252.0725	Water	Nucleosides	MN5
77	N-(2-Hydroxyethyl)adenosine	$C_{12}H_{17}N_5O_5$	312.1302	136.0750/162.0753/180.0875/255.2666	Water	Nucleosides	MN5
78	6-[[9-[(2R,3R,4S,5R)-3,4-Dihydroxy-5-(hydroxymethyl)oxolan-2-yl]purin-6-yl]amino]hexanoic acid	$C_{16}H_{23}N_5O_6$	382.1721	136.0538/152.0557/190.0729/ 206.1031/250.1279	Water	Nucleosides	MN5
79	Isopentenyl-Adenine-7-glucoside or Isopentenyl-Adenine-9-glucoside	$C_{16}H_{23}N_5O_5$	366.1772	136.0610/190.1083/234.1340	Water	Nucleosides	MN5
80	N-Fructosyl phenylalanine*	$C_{15}H_{21}NO_7$	328.1391	91.0561/120.0805/132.0805/166.0852/ 178.0855/264.1218/292.1168/310.1255	Water	Amino acid	MN6
81	N-Fructosyl isoleucine*	$C_{12}H_{23}NO_7$	294.1547	86.0966/212.1286/230.1387/ 258.1340/276.1449	Water	Amino acid	MN6
82	Boc-Glu(OMe)-Ome	$C_{12}H_{21}NO_6$	276.1442	86.0958/132.1012/144.1011/161.0663/ 212.1267/230.1370/258.1313	Water	Amino acid	MN6
83	N-Fructosyl valine	$C_{11}H_{21}NO_7$	280.1391	72.0816/84.0811/118.0868/130.0854/ 138.0523/150.0759/198.1108/216.1244/ 221.0196/244.1173/262.1264	Water	Amino acid	MN6
84	beta-D-Glucopyranosyl-O-L-tyrosine	$C_{15}H_{21}NO_8$	344.1340	97.0277/136.0761/165.0563/182.0821/ 194.0817/280.1192/308.1148/326.1252	Water	Amino acid	MN6
85	N-(1-Deoxy-D-sorbopyranose-1-yl)-L-glutamic acid	$C_{11}H_{19}NO_9$	310.1133	97.0296/130.0508/138.0550/148.0600/ 160.0597/180.0668/210.0740/226.0724/ 246.0976274.0929/292.1032	Water	Amino acid	MN6
86	N-(1-Deoxy-D-fructosyl)-L-proline	$C_{11}H_{19}NO_7$	278.1234	72.0799/118.0852/128.0692/216.1216/ 232.1176/242.1014/244.1187/260.1122	Water	Amino acid	MN6
87	O-Mannopyranosylthreonine	$C_{10}H_{19}NO_8$	282.1183	57.0327/102.0551/120.0642/132.0654/ 138.0556/144.1018/186.0755/200.0895/ 246.0968/264.1082	Water	Amino acid	MN6
88	2-[(1-Methylethyl)oxy]-8-(methyloxy)-1H-purin-6-amine	$C_9H_{13}N_5O_2$	224.1142	62.0593/74.0594/86.0596/97.0279/ 158.0797/188.0921/206.1014	Water	Amino acid	MN6
89	Allyl 2-acetamido-2-deoxy-beta-D-glucopyranoside	$C_{11}H_{19}NO_6$	262.1285	55.0163/57.0324/72.0798/84.0795/ 112.0370/118.0847/130.0854/143.0552/ 198.1075/216.1204/244.1141	Water	Amino acid	MN6

(continued)

Table 2. Continued.

Number	Identification	Formula	$[M+H]^+$	MS^2	Identification	Type	Cluster
90	L-Tyrosine*	$C_9H_{11}NO_3$	182.0812	91.0549/119.0497/123.0447/136.0764/ 147.0444/165.077	Water	Amino acid	MN7
91	N-(4-Carboxy-2-methylbutanoyl)-L-tyrosine	$C_{15}H_{19}NO_6$	310.1285	57.0697/113.0595/139.0385/181.0511/ 204.1026/240.1330/247.0066/251.0915/ 264.1225/293.1020	Water	Amino acid	MN7
92	Cyclo(L-Val-L-Pro)*	$C_{10}H_{16}N_2O_2$	197.1285	55.0543/70.0650/72.0803/98.0593/ 124.1111/154.0722/169.1306	Water/alcohol	Peptides	MN8
93	Cyclo(L-Phe-L-Pro)*	$C_{14}H_{16}N_2O_2$	245.1285	55.0532/70.0656/98.0595/ 120.0802/154.0741	Water/alcohol	Peptides	MN8
94	Cyclo(Pro-Leu)*	$C_{11}H_{18}N_2O_2$	211.1441	55.0535/70.0658/86.0964/98.0593/ 138.1273/194.1200	Water/alcohol	Peptides	MN8
95	Cyclo(L-Val-L-Leu)*	$C_{11}H_{20}N_2O_2$	213.1598	55.0537/72.0806/86.0962/140.1440/ 168.1375/185.1640	Water/alcohol	Peptides	MN8
96	Cyclo(Val-Val-)	$C_{10}H_{18}N_2O_2$	199.1441	55.0546/72.0804/84.9592/100.0757/ 126.1269/171.1481	Water/alcohol	Peptides	MN8
97	Val-val	$C_{10}H_{20}N_2O_3$	217.1547	55.0552/72.0811/84.9598/114.9827	Water/alcohol	Peptides	MN8
98	PyroGlu-Pro*	$C_{10}H_{14}N_2O_4$	227.1026	70.0653/84.0447/116.0703/181.0970/ 209.0887/227.1382	Water	Peptides	MN9
99	PyroGlu-Val*	$C_{10}H_{16}N_2O_4$	229.1183	57.0347/72.0814/84.0045/86.0966/ 116.0709/138.0922/183.1109	Water	Peptides	MN9
100	Leucylproline	$C_{11}H_{20}N_2O_3$	229.1547	70.0656/86.0960/126.0898/ 140.0817/142.0868	Water	Peptides	MN9
101	Leu-Val	$C_{11}H_{22}N_2O_3$	231.1703	55.0553/72.0809/86.0964/126.1272/ 172.1321/186.1490	Water	Peptides	MN9
102	Boc-L-leucine/Boc-L-isoleucine	$C_{11}H_{21}NO_4$	232.1543	55.0546/69.0707/72.0811/97.0653/ 126.1283/136.0631/172.1333/186.150	Water	Peptides	MN9
103	Ile-Gly-Ile*	$C_{14}H_{27}N_3O_4$	302.2074	86.0968/132.1021/143.1178/171.1125	Water	Peptides	MN10
104	Gamma-Glu-leu	$C_{11}H_{20}N_2O_5$	261.1445	86.0949	Water	Peptides	MN10
105	Leu-Gly-Val	$C_{13}H_{25}N_3O_4$	288.1918	55.0556/72.0811/86.0969/132.1009/ 189.1240/229.0979	Water	Peptides	MN10
106	(E,2R)-2-[(2-Acetyloxyacetyl)amino]-6-ethoxy-6-oxohex-4-enoic acid	$C_{12}H_{17}NO_7$	288.1078	72.08170/86.0952/132.1025/143.1185/ 189.1211/229.0923	Water	Peptides	MN10
107	Asp-Leu*	$C_{10}H_{18}N_2O_5$	247.1288	69.07/86.0968/88.0386/132.0627/ 141.1004/201.1427	Water	Peptides	MN11
108	L-Asparaginyl-L-Isoleucine	$C_{10}H_{19}N_3O_4$	246.1448	55.0540/57.0335/69.0698/86.0954/ 110.0227/132.0981/141.1012/166.0851/ 201.1230/212.0925/229.1163	Water	Peptides	MN11
109	N(6)-(2-Carboxyethyl)lysine	$C_9H_{18}N_2O_4$	219.1339	57.0694/60.0444/84.9589/86.0959/ 132.1008/143.0017/161.0965/ 176.1043/173.1263	Water	Peptides	MN11
110	Kynurenic Acid*	$C_{10}H_7NO_3$	190.0499	89.0387/116.0494/144.0440	Water/alcohol	Others	MN12
111	Indole-3-carboxylic acid	$C_9H_7NO_2$	162.0550	89.0380/116.0488/144.0438	Water/alcohol	Others	MN12
112	Tetrahydroharman-3-carboxylic acid*	$C_{13}H_{14}N_2O_2$	231.1128	74.0224/130.0646/143.0726/158.0959/ 168.0796/188.0680/214.0830	Water	Others	MN13
113	1,2,3,4-Tetrahydro-beta-carboline-3-carboxylic acid	$C_{12}H_{12}N_2O_2$	217.0972	144.0807/77.0385/74.0233143.0715	Water	Others	MN13
114	Undecaethylene glycol*	$C_{22}H_{46}O_{12}$	503.3062	87.0438/89.0595/133.0848/177.1104	Water/alcohol	Fatty acids	
115	sn-Glycero-3-phosphocholine*	$C_8H_{20}NO_6P$	258.1101	60.0811/86.0962/104.1067/184.0730	Water	Fatty acids	
116	2-(3-Carboxypropanoylamino)-3-methylpentanoic acid*	$C_{10}H17NO5$	232.1179	55.0188/69.0710/86.9565/132.0785	Water	Fatty acids	
117	1-(9Z-Octadecenoyl)-sn-glycero-3-phosphocholine*	$C_{26}H_{52}NO_7P$	522.3554	60.0811/86.0957/104.1066/ 184.0731/504.34	Alcohol	Fatty acids	
118	1-Hexadecanoyl-sn-glycero-3-phosphocholine*	$C_{24}H_{50}NO_7P$	496.3398	60.0811/86.0957/104.1066/ 184.0731/478.3244	Alcohol	Fatty acids	
119	1-Stearoyl-2-hydroxy-sn-glycero-3-phosphoethanolamine*	$C_{23}H_{48}NO_7P$	482.3241	57.0692,62.0509,227.2013,155.01, 310.3108,341.3047,464.3073	Alcohol	Fatty acids	
120	1-Palmitoyl-2-hydroxy-sn-glycero-3-phosphoethanolamine*	$C_{21}H_{44}NO_7P$	454.2928	57.0347/62.0344/71.0861/85.1024/ 98.9848/155.0098/282.2801/313.2770	Alcohol	Fatty acids	
121	N,N-Diethyl-3-methylbenzamide*	$C_{12}H_{17}NO$	192.1383	91.0528/109.0720/119.0720	Alcohol	Fatty acids	
122	2-Methyl-4′-(methylthio)-2-morpholinopropiophenone*	$C_{15}H_{21}NO_2S$	280.1366	70.0647/88.0389/150.0764	Alcohol	Alkaloids	
123	Riboflavin*	$C_{17}H_{20}N_4O_6$	377.1456	69.0347/99.0440/172.0876/200.0824/ 243.0886/377.1502	Alcohol	Alkaloids	
124	Xanthurenic acid*	$C_{10}H_7NO_4$	206.0448	51.0237/77.0385/104.0490/ 132.0447/160.0397	Alcohol	Alkaloids	

*Means the points which are annotated by GNPS in molecular network.

ions of m/z 70. The cleavage of the carbon chain within the fragment at m/z 302 easily formed the fragment at m/z 106, which was a characteristic fragment ion of this fatty acid cluster. Moreover, the fragment ions at m/z 88, m/z 70 and m/z 57 were obtained from the fragment ion at m/z 106, and they were also characteristic fragment ions. N-lauryl diethanolamine (52) and tetradecyl diethanolamine have the same fragmentation pathway. The carbon chains of both of them were broken, then continuously dehydrated, thereby producing fragment ions of the same mass. Their fragmentation pattern is shown in Figure 4. Other 15 fatty acids were identified by the fragmentation pattern of

tetradecyl diethanolamine, and their chemical structures along with that of tetradecyl diethanolamine (51) are shown in Figure 4, with the associated information listed in Table 2.

The m/z 106/88/70/55 fragment ion peaks appeared in the compounds 53∼60. The $[M+H]^+$ quasi-molecular ion peaks differed by an integer multiple of 14 Da, suggesting that they were consistent with the tetradecyl diethanolamine skeleton. The deduction of the unannotated point N-hexadecyl diethanolamine (57, $[M+H]^+$ 330.3367) was utilized as an example. First, compound 57 was directly connected to tetradecyl diethanolamine in the cluster, and the $[M+H]^+$ was 28 Da more than the
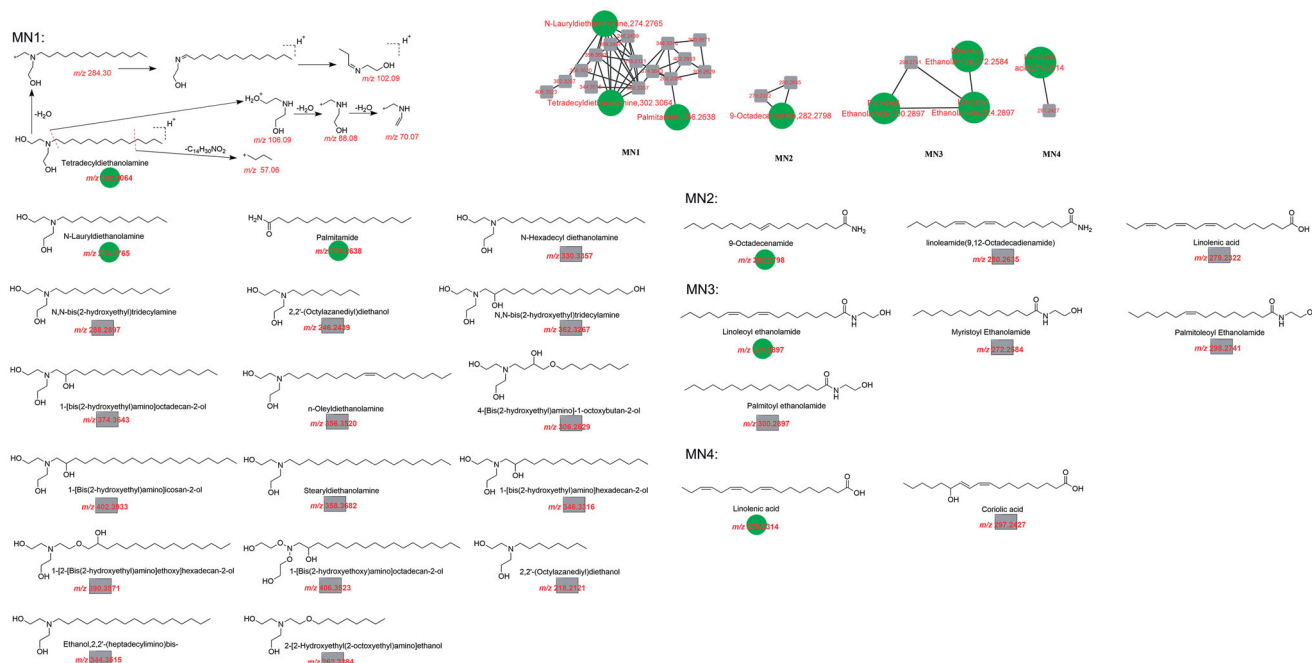
Figure 4. Analysis of fatty acid. Green circles are annotated points; grey squares are inferred unannotated point.

quasi-molecular ion peak of compound **51**. The basic skeleton was consistent with **51**. The results of the formula predictor (ppm ≤ 5) led to two different molecular formulae: $C_{20}H_{43}NO_2$ (ppm = 1.99) and $C_{18}H_{41}N_4O$ (ppm = −2.08). The above two molecular formulae were imported into Scifinder in order to search for potential compounds. Then, the type of the compounds was set and the Chemical Abstracts Service Registry Number of the potential compound was obtained. The structure of the potential compound was obtained from PubChem, which led to the discovery of N-hexadecyl diethanolamine. The self-build database, which included mass spectrum, chromatographic data, CNKI, MassBank (http://www.massbank.jp/RecordDisplay?id=PR300821), National Institute of Standards and Technology (NIST, https://www.nist.gov/) and other databases, as well as the literature, were used to search the N-hexadecyl diethanolamine. The comparison of the mass spectrometry data of compound **51**, the mass splitting, structural formula and other information finally led to the discovery that the unannotated point 330.336 was the compound N-hexadecyl diethanolamine. The m/z 106/88/70/57 fragment ions were shown on compounds **61**∼**57**. There was one more dehydration (H_2O) fragment ion peak in their structures compared to the structure of tetradecyl diethanolamine, indicating that there were more hydroxyl (OH) molecules than in compounds **53**∼**60**, as determined by Scifinder. Their structure can be determined by combining their quasi-molecular ion peak of $[M + H]^+$ and other information. All 13 compounds were identified from the *A. chinensis* extracts for the first time (Table 2). The mass splitting of the annotated points in the three MNs (MN2, MN3, and MN4) had continuous methylene fragment ions, which was in line with the typical fatty acid fragmentation pattern. Through the structural formula and fracture pattern of these annotated points, the unannotated points can be deduced (Figure 4). These four compounds were identified from the *A. chinensis* extracts for the first time.

Analysis of the unannotated points of nucleoside clusters
Adenosine derivatives easily lose their glycosyl group ($C_5H_9O_4$) to produce the characteristic ion $[M–C_5H_9O_4+H]^+$.

Succinoadenosine (**76**, $C_{14}H_{17}N_5O_8$) is a succinic acid derivative of adenosine. It was selected as the reference compound to study the MS^2 fragmentation pathway of the nucleoside MN5. The $[M + H]^+$ quasi-molecular ion of **76** was m/z 384 in the positive ion mode. The glycosyl group ($C_5H_9O_4$) could be easily released to form the characteristic fragment ions $[M–C_5H_9O_4+H]^+$ of m/z 252. Then, the fragment ions passed through a series of cracking, thereby forming the intermediate ions of m/z 188 and m/z 162. Finally, the succinic acid group ($C_4H_6O_4$) was fully removed in order to form the fragment ion of m/z 136. The ion of $[M–C_5H_9O_4+H]^+$ and $[M–C_5H_9O_4–C_4H_6O_4+H]^+$ were two characteristic fragment ions of this cluster. Three more nucleoside compounds were annotated in MN5 (Figure 5) by the screening of the online databases, including GNPS, MassBank, METLIN, and ChemSpider, with related information listed in Table 2.

Analysis of the unannotated points of amino acid clusters
The amino acid cluster was divided into MN6 and MN7. N-fructosyl phenylalanine (**80**) and N-fructosyl isoleucine (**81**) in MN6, and L-tyrosine (**90**) in MN7 were selected for analysis of the fragmentation pathway of amino acids (Figure 6). MN6 is an amino acid with glycosyl substitution. The neutral loss of H_2O (18 Da) and the loss of glycosyl ($C_5H_{10}O_5$) were the features of MN6. The $[M + H]^+$ quasi-molecular ion m/z 328 of N-fructosyl phenylalanine could be formed easily in the positive ion mode. The quasi-molecular ion was easily and continuously dehydrated to produce $[M–H_2O + H]^+$ of m/z 310 and $[M–2H_2O + H]^+$ of m/z 292. Additionally, the glycosyl group ($C_5H_{10}O_5$) was removed to form the characteristic fragment ion of m/z 178 or the $C_6H_{11}O_5$ was removed to generate the m/z 166 ion. The quasi-molecular ion could also generate the $[M + H–2H_2O–CO_2]^+$ fragment ion of m/z 264 firstly. Then, the dissociation of the $C_5H_8O_4$ formed the ion of m/z 132. Likewise, the cleavage of the formic acid group, the continuous dehydration, and the loss of the glycosyl group from the N-fructosyl isoleucine was also observed. The structures of the other eight

MN5

Succinoadenosine
m/z **384.1165**

m/z 252.07

m/z 162.08

m/z 188.06

m/z 136.06

6-[[9-[(2R,3R,4S,5R)-3,4-Dihydroxy-5-(hydroxymethyl)oxolan-2-yl]
purin-6-yl]amino]hexanoic acid

m/z 382.1731

N-(2-Hydroxyethyl)adenosine

m/z 312.1293

Isopentenyl-Adenine-7-glucoside

m/z 366.1764

or

Isopentenyl-Adenine-9-glucoside

m/z 366.1764
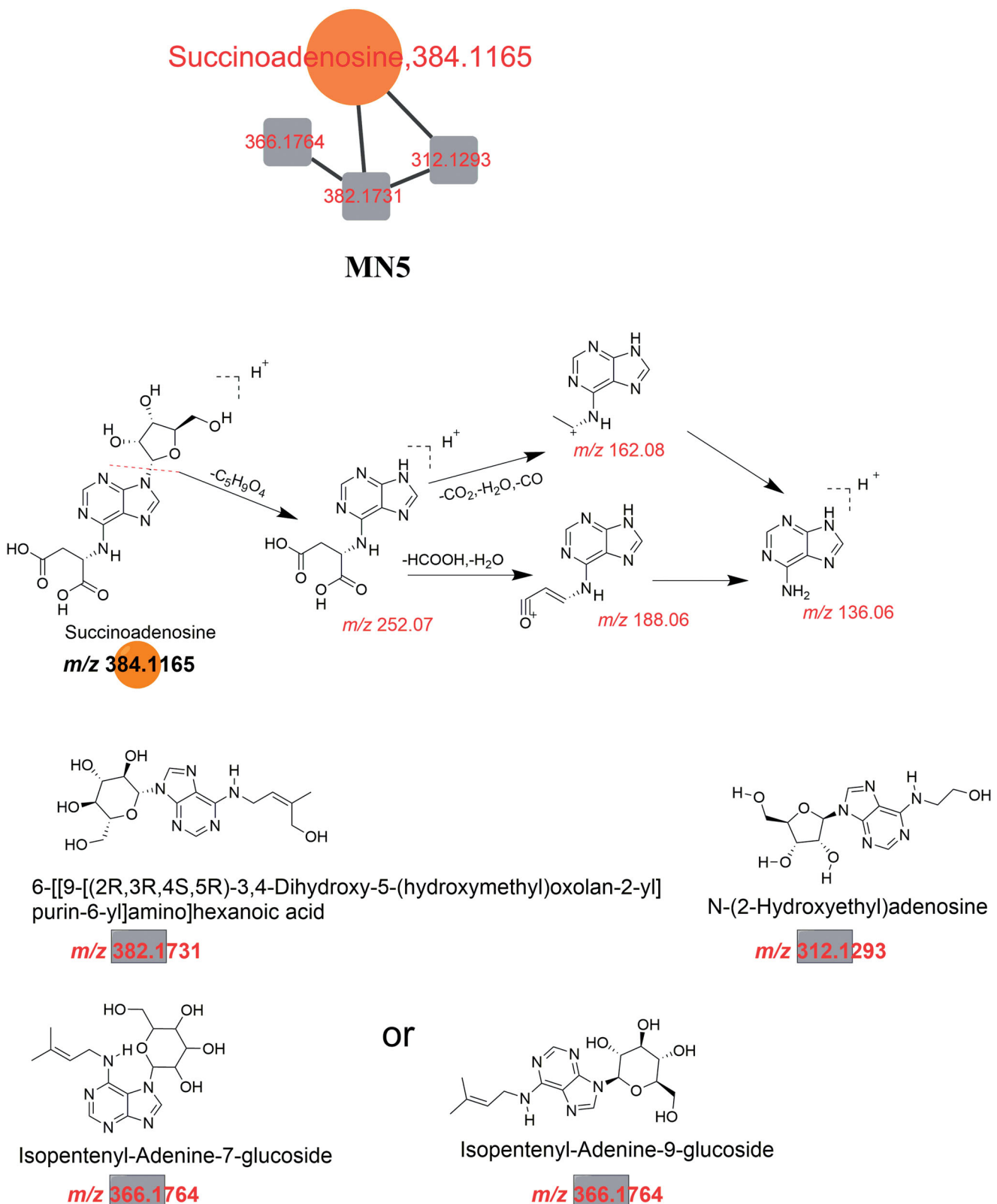
Figure 5. Analysis of nucleosides. Orange circles are annotated points of nucleosides; grey squares are inferred unannotated points.

compounds in MN6 could be deduced on the basis of the pathway of these two compounds. N-(4-carboxy-2-methylbutanoyl)-L-tyrosine is a compound that is characterized by the replacement of one hydrogen on the amino group of tyrosine with 4-carboxy-2-methylbutanoyl. The substituent was easily removed to produce an ion at *m/z* 204 in the positive ion mode, and [M + H]⁺

quasi-molecular ion *m/z* 310. This was conformed by the fragmentation pattern of tyrosine. Thus, the unannotated point at *m/z* 310.128 was recognized as N-(4-carboxy-2-methylbutanoyl)-L-tyrosine. The chemical structures of the compounds in MN6 and MN7 are shown in Figure 6, and the relevant information is listed in Table 2.
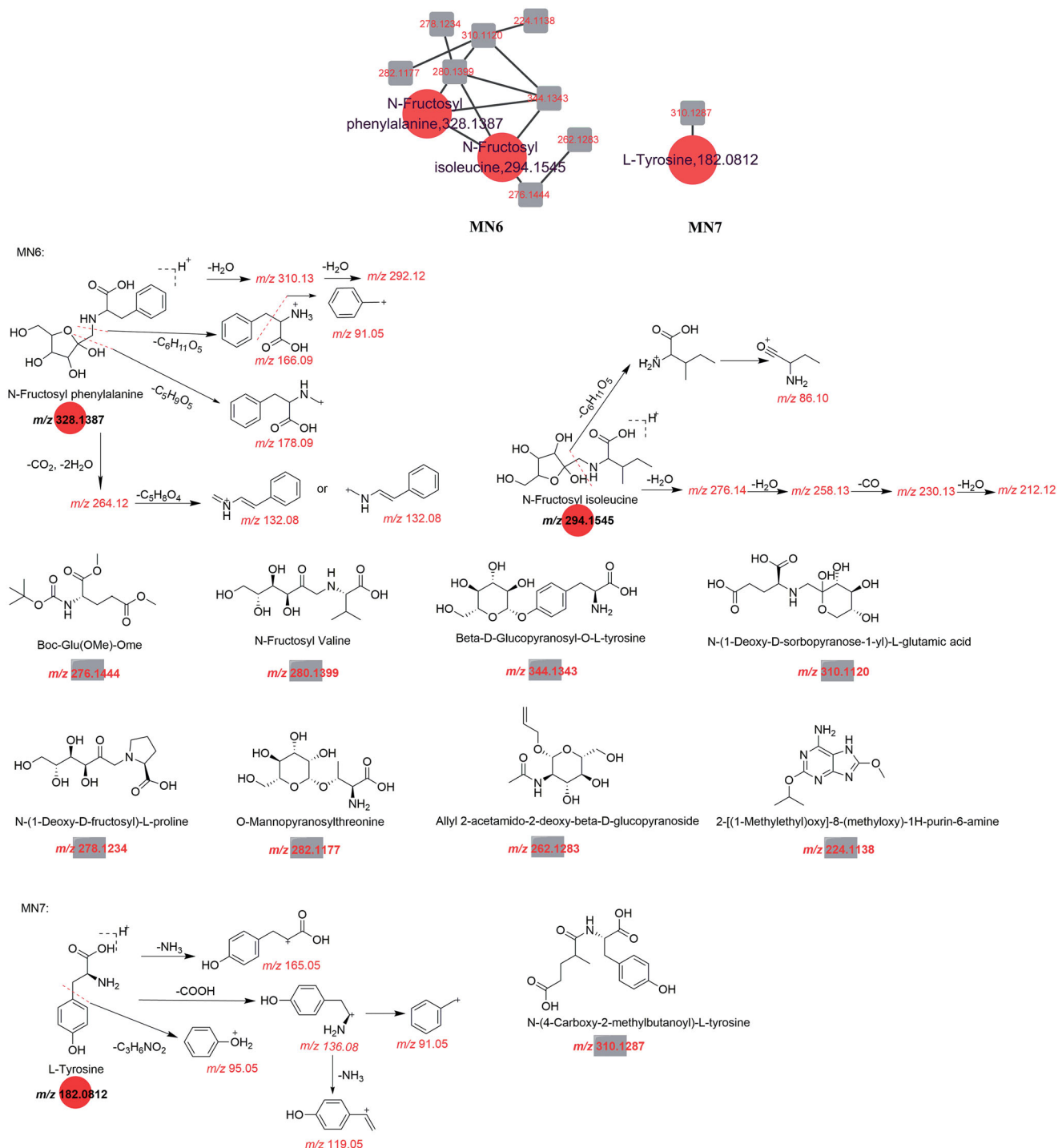
Figure 6. Analysis of amino acids. The red circle is the annotated point of nucleosides; the grey square is inferred unannotated point.

Analysis of the unannotated points of peptide clusters

Peptides were included in four clusters (MN8, MN9, MN10, and MN11). Cyclo (L-Phe-L-Pro) (**43**) in MN8 and PyroGlu-Val (**99**) in MN9 were selected as the reference compounds to investigate the MS^2 fragmentation patterns of the peptide (Figure 7). The points in MN9 that exhibited the highest frequency of ions at m/z 55/57, m/z 70/72, m/z 84/86 were the elementary characteristic ions (Figure 7). The cyclic dipeptides could be easily open from the $[M + H]^+$ quasi-molecular ion m/z 245 of the cyclo (L-Phe-L-Pro) in the positive ion mode to form the fragment of m/z 120 and m/z 98, which is one of the characteristic fragment ions of this MN. In addition, the carbonyl (CO) compound could

dissociate from m/z 98 to generate the characteristic fragment ions of m/z 70. Finally, the opening of the tetrahydropyrrole could lead to the production of the characteristic fragment ions of m/z 55. Similar fragmentation pathways were observed on MN9, MN10 and MN11, including mainly deamination, decarboxylation, decarboxylation and peptide chain scission (i.e., PyroGlu-Val). Their chemical structure is shown in Figure 7, and the relevant information is listed in Table 2.

Analysis of the unannotated points of other clusters

Other clusters were included in MN12 and MN13. Kynurenic acid (**110**) in MN12 and tetrahydroharman-3-carboxylic acid
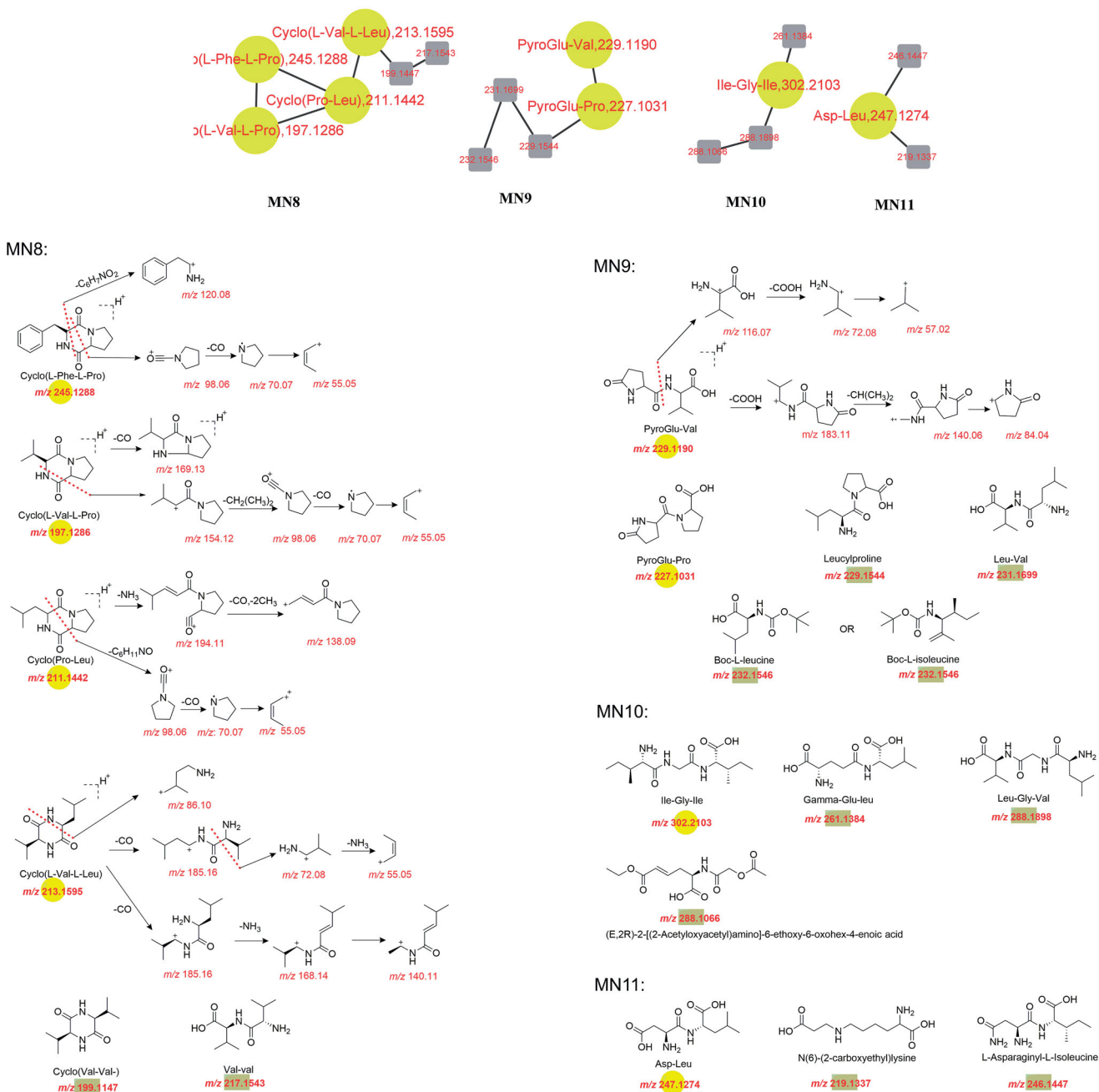
Figure 7. Analysis of peptides. The yellow circle is the annotated point of peptides; the grey square is inferred unannotated point.

(112) in MN13 were selected as reference compounds to investigate the MS^2 fragmentation patterns. In MN12, the carboxyl group (COOH) could be easily lost from the $[M + H]^+$ quasi-molecular ion m/z 190 of kynurenic acid to form the fragment at m/z 144 in the positive mode. Then, the carbonyl group (CO) dissociated from the m/z 144 to form a fragment ion at m/z 116. Finally, the amino group was removed from the m/z 116 to generate the fragment ion of m/z 89. The fragment ions, m/z 144, m/z 116, and m/z 98, appeared in two other compounds belonging to MN12, which suggests that their skeletons are similar. Two this two unannotated components in the cluster were identified. Their chemical structure and that of kynurenic acid are shown in Figure 8, and the relevant information is listed in Table 2.

MN is an online method that is continuously updated and allows for the visualization and targeting of natural products,

thereby enabling biological research and biotechnology applications in a wide range of fields. The generation of MN is based on the analysis of mass spectrometry of crude extracts. The MS/MS spectra of compounds are compared in pairs to find similarities in fragmentation pathways, that is, the same fragment ion or similar neutral loss. The nodes in MN represent spectra or compounds, and the edges between nodes represent the similarity between the two spectra. Compounds with similar fragmentation patterns are grouped into the same cluster. Additionally, those with different MS/MS spectra are not relevant (Wang et al. 2019). In other words, the two nodes connected by a line in MN represent two compounds that have similar structures and fracture patterns. The identification of certain compounds by direct use of standard compounds to generate MS/MS spectra, or the use of public spectral libraries (i.e., GNPS or referencing known documents) is sufficient to determine the molecular family of
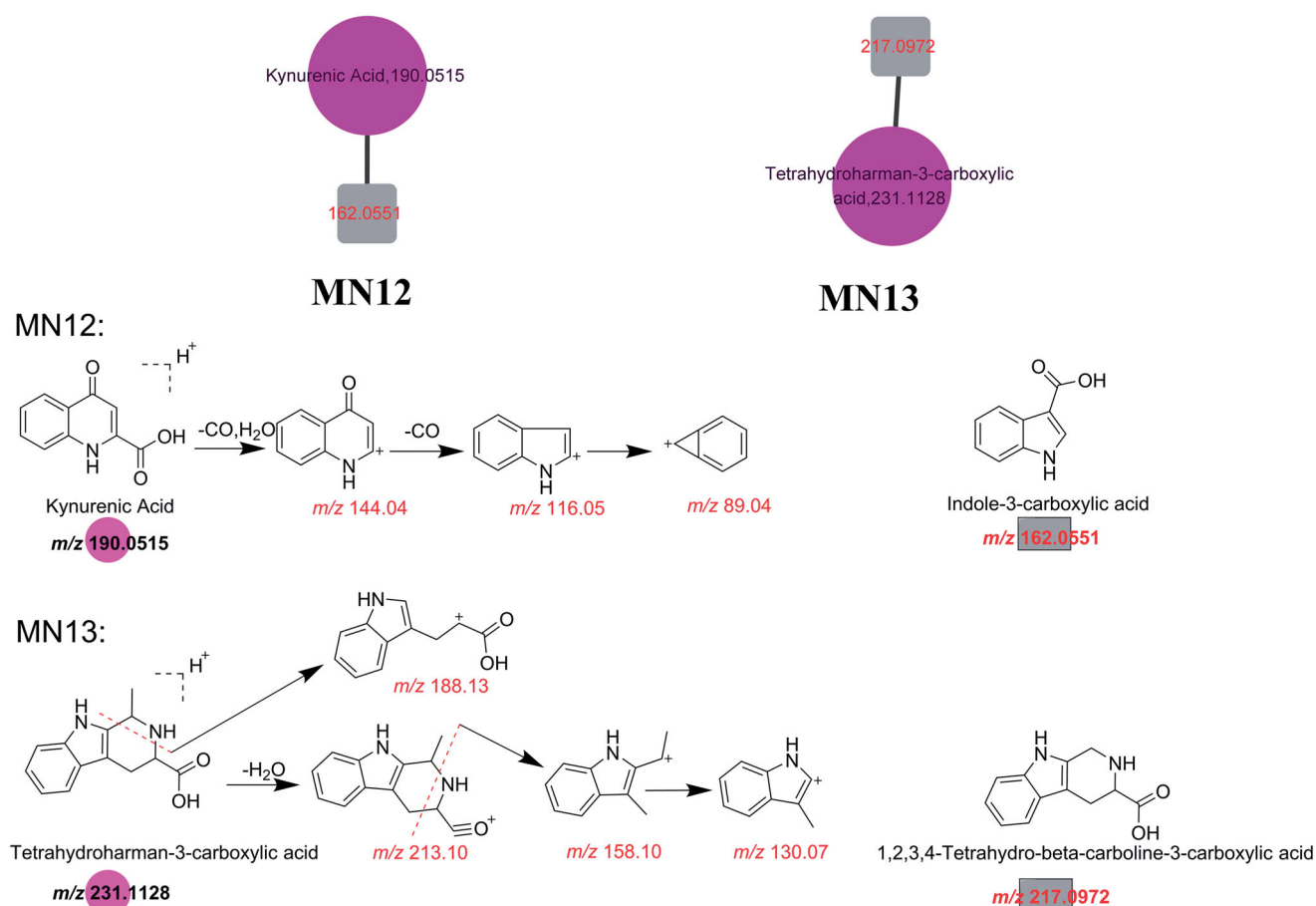
Figure 8. Analysis of other categories. Purple circles are annotated points; grey squares are inferred unannotated points.

Table 3. The mass spectrum/chromatographic information of the four compounds in the standard reference substance and *Aspongopus chinensis* Dallas extract.

Identification	Formula	$[M+H]^+$	MS^2	t_R (min)	Source	
N-(2-Hydroxyethyl)adenosine	$C_{12}H_{17}N_5O_5$	312.1302	180.0875/136.0750/162.0753/255.2666	5.850	Water	Inference
N-(2-Hydroxyethyl)adenosine	$C_{12}H_{17}N_5O_5$	312.1302	180.0882/136.0623/162.0776	5.853	Water	Reference
Cyclo(Val-Val-)	$C_{10}H_{18}N_2O_2$	199.1441	55.0546/72.0804/84.9592/100.0757/ 126.1269/171.1481	10.686	Water/alcohol	Inference
Cyclo(Val-Val-)	$C_{10}H_{18}N_2O_2$	199.1441	55.0538/72.0809/84.9597/100.0768/ 126.1274/171.1490	10.702	Water/alcohol	Reference
Kynurenic Acid	$C_{10}H_7NO_3$	190.0499	89.0387/116.0494/144.0440	7.815	Water/alcohol	Inference
Kynurenic Acid	$C_{10}H_7NO_3$	190.0499	89.0393/116.0500/144.0449	7.789	Water/alcohol	Reference
1,2,3,4-Tetrahydro-beta-carboline-3-carboxylic acid	$C_{12}H_{12}N_2O_2$	217.0972	144.0807/77.0385/74.0233/143.0715	8.567	Water	Inference
1,2,3,4-Tetrahydro-beta-carboline-3-carboxylic acid	$C_{12}H_{12}N_2O_2$	217.0972	144.0813/77.0390/74.0252/143.0726	8.533	Water	Reference

related compounds, thereby facilitating the identification of unknown compounds (Li et al. 2020).

The MN of *A. chinensis* was generated using GNPS. The aqueous extract produced a total of 76 clusters, of which 11 were analyzable. Moreover, a total of 47 compounds matched with compounds within the database. The ethanol extract of *A. chinensis* produced 54 clusters, and seven of them were analyzable. A total of 34 compounds matched with compounds in the database. Since the five clusters obtained from the aqueous and ethanol extracts were exactly the same, 13 clusters of *A. chinensis* crude extract were visualized by Cytoscape 3.8.2 and shown in Figure 3. The annotated point was indicated by a circle, while the deduced unannotated point was indicated by a square. Each node represented a compound, and inside the node, the molecular mass of the parent ion, such as the molecular ion, fragment ion, or adduct mass of the compound, was indicated, linked by a line indicating

the similarity among the compounds. Moreover, fatty acids were represented by a green node; nucleosides were represented by an orange node; amino acids were represented by a red node. Peptides and other clusters were represented by yellow and purple points, respectively. A total of 34 fatty acids, 6 nucleosides, 18 peptides, 12 amino acids and 4 other compounds were initially identified, among which 74 compounds were reported for the first time. The relevant information is listed in Table 2.

Previous studies on *A. chinensis* have mainly focussed on nucleosides and their analogues, as well as the dopamine derivatives contained in its extracts. In this study, 124 compounds were found in the extracts of *A. chinensis*, and they were subdivided into five categories, which represented an informative supplement to fill the gap in previous research. Among the ingredients reported for the first time in this study, fatty acids, peptides, and amino acids were the majority. Nucleosides and amino acids were

detected only in the aqueous extract. The aqueous extract contained a higher number of peptides as compared to the amount in the ethanol extract. The ethanol extract contained a higher content of fatty acids as compared to the amount in the aqueous extract. The difference in composition was related to the polarity of the compound itself. Nucleosides, amino acids and peptides mostly had polar carboxyl groups and glycosyl groups. Thus, they were mostly present in the aqueous extract, while the polarity of fatty acids was lower. Therefore, fatty acids were found in the ethanol extract. Previous studies have revealed that the haemolymph of
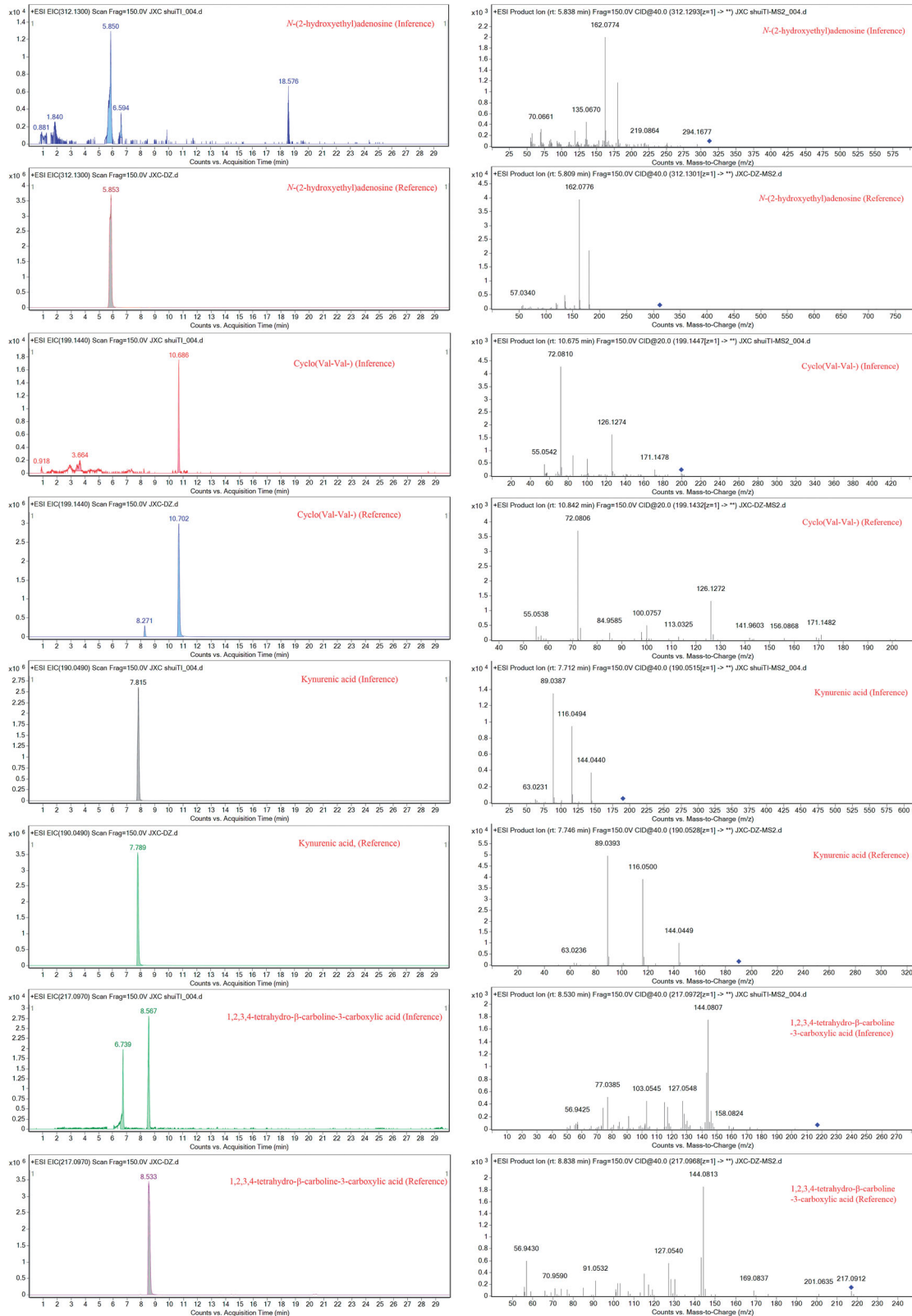
Figure 9. The extracted ion chromatogram (EIC) and MS^2 spectrum of the four compounds in the standard reference substance and *Aspongopus chinensis* extract.

A. chinensis has a high protein content and physiologically active peptides, which effectively inhibit the growth of gastric cancer and human breast cancer. In addition, a small peptide with antibacterial properties was purified from its haemolymph. The protective effect of *A. chinensis* insect extract on nephropathy is related to its rich content in dopamine, indole, and piperidine, while its anti-ulcer effect resides in the fatty oil extract (Hou et al. 2012). However, the biologically active substances in *A. chinensis* are largely unknown. Therefore, in this study, the compounds in the extracts of *A. chinensis* were subdivided into five categories. Our findings revealed a further development in the utilization of *A. chinensis*.

Validation of the molecular network result

N-(2-Hydroxyethyl) adenosine (**28**, adenosine), cyclo (Val-Val-) (**47**, peptide chain), kynurenic acid (**61**, other categories: quinoline acid) and 1,2,3,4-tetrahydro-β-carboline-3-carboxylic acid (**64**, other types: carboxylic acid) were the unannotated points deduced from the molecular network. Hence, these compounds were selected to verify whether the molecular network speculation was correct. These four compounds were used as reference substances and were purchased on the China Reagent website. They were dissolved in methanol to prepare a mixed control solution of 1.0 mg/mL. Then, the primary/secondary mass spectrum data were obtained according to the method described in the section 'UPLC-QTOF MS analysis'. The results in Table 3 and Figure 9 demonstrate that the retention time and ion fragments of the reference substances matched with the predicted compound, which confirms the accuracy of our method, which used MN analogues to cluster compounds with similar fragmentation pathways.

Conclusions

UPLC-QTOF-MS performed in full MS/dd MS^2 mode represented a rapid and reliable method to determine the composition of *A. chinensis* aqueous and ethanol extracts. Furthermore, a practical, integrated, targeted and untargeted strategy of data processing was used, in combination with an in-built database with MN, enabling, for the first time, the rapid profiling of non-target constituents of TCMs from insects. The efficiency in the identification of unknown compounds was greatly improved by multiple database matching and fragmentation rules, which avoids the high cost and inefficiency of traditional techniques that are used to discover compounds in traditional Chinese medicine, being helpful for the discovery of new compounds. A method using the retention time of the substance used as standard reference and the mass spectrometry fragments to verify the accuracy of the unannotated compound was proposed, which avoids the complexity of the identification of first-report compounds. This study reported a total of 124 compounds in *A. chinensis* extracts, which provides a reference for the future development and use of *A. chinensis*. In short, this might be a promising study providing a convenient and powerful processing method of the data for the rapid profiling of non-target chemical constituents of insects used in TCMs. Due to the limitations of MN database, compounds in some clusters cannot be identified because there are no annotated compounds in these clusters. These identified compounds in this study need to be further isolated for activity screening *in vitro* or *in vivo*.

References

Chen S, Huang G, Liao W, Gong S, Xiao J, Bai J, Wendy Hsiao WL, Li N, Wu JL. 2021. Discovery of the bioactive peptides secreted by bifidobacterium using integrated MCX coupled with LC-MS and feature-based molecular networking. Food Chem. 347:129008.

Fan Q, Wei H, Cai HB, Sun XG, Diao JX, Wang QR. 2011. Effects of the serum containing *Aspongopus chinensis* Dallas on the expression of SW480 apoptosis-associated factors. J Anhui Agri. 39:7828–7831.

He ZQ, Sun ZC, Zhang L. 2016. Study on the repair mechanism of *Aspongopus chinensis* Dallas on rat reproductive injury. Chin Trad Patent Med. 38:924–927.

Hou XH, Sun T, Li XF. 2012. Effects of *Aspongopus chinensis* Dallas extracts on cell proliferation and cell cycle of SGC-7901 and HepG2 cell lines. Chin Trad Patent Med. 34:2278–2281.

Huang Q, Zhang F, Liu S, Jiang Y, Ouyang D. 2021. Systematic investigation of the pharmacological mechanism for renal protection by the leaves of *Eucommia ulmoides* Oliver using UPLC-Q-TOF/MS combined with network pharmacology analysis. Biomed Pharmacother. 140:111735.

Jiang Y, Liu R, Chen J, Liu M, Liu M, Liu B, Yi L, Liu S. 2019. Application of multifold characteristic ion filtering combined with statistical analysis for comprehensive profiling of chemical constituents in anti-renal interstitial fibrosis I decoction by ultra-high performance liquid chromatography coupled with hybrid quadrupole-orbitrap high resolution mass spectrometry. J Chromatogr A. 1600:197–208.

Jiang Y, Liu R, Liu M, Yi L, Liu S. 2018. An integrated strategy to rapidly characterize non-targeted benzylisoquinoline alkaloids from *Plumula nelumbinis* ethanol extract using UHPLC/Q-orbitrap HRMS. Int J Mass Spectrom. 432:26–35.

Li J, Li S, Li L. 2020. Advances in studies on chemical constituents pharmacological effects and clinical application of *Aspongopus chinensis*. China J Chin Mater Med. 45:303–311.

Li Y, He N, Luo M, Hong B, Xie Y. 2020. Application of untargeted tandem mass spectrometry with molecular networking for detection of enniatins and beauvericins from complex samples. J Chromatogr A. 1634:461626.

Liu M, Jiang Y, Liu R, Liu M, Yi L, Liao N, Liu S. 2019. Structural features guided "fishing" strategy to identification of flavonoids from lotus plumule in a self-built data "pool" by ultra-high performance liquid chromatography coupled with hybrid quadrupole-orbitrap high resolution mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci. 1124:122–134.

Messaili S, Colas C, Fougere L, Destandau E. 2020. Combination of molecular network and centrifugal partition chromatography fractionation for targeting and identifying *Artemisia annua* L. antioxidant compounds. J Chromatogr A. 1615:460785.

Rodrigues A, Lami R, Escoubeyrou K, Intertaglia L, Mazurek C, Doberva M, Pérez-Ferrer P, Stien D. 2022. Straightforward *N*-acyl homoserine lactone discovery and annotation by LC–MS/MS-based molecular networking. J Proteome Res. 21(3):635–642.

Santos AL, Soares MG, de Medeiros LS, Ferreira MJP, Sartorelli P. 2021. Identification of flavonoid-3-*O*-glycosides from leaves of *Casearia arborea* (Salicaceae) by UHPLC-DAD-ESI-HRMS/MS combined with molecular networking and NMR. Phytochem Anal. 32(6):891–898.

Shi YN, Tu ZC, Wang XL, Yan YM, Fang P, Zuo ZL, Hou B, Yang TH, Cheng YX. 2014. Bioactive compounds from the insect *Aspongopus chinensis*. Bioorg Med Chem Lett. 24(22):5164–5169.

Tan J, Tian Y, Cai R, Yi T, Jin D, Guo J. 2019. Antiproliferative and proapoptotic effects of a protein component purified from *Aspongopus chinensis* Dallas on cancer cells *in vitro* and *in vivo*. Evid Based Complement Alternat Med. 2019:1–12.

Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, et al. 2016. Sharing and

community curation of mass spectrometry data with global natural products social molecular networking. Nat Biotechnol. 34(8):828–837.

Wang X, Serrano R, González-Menéndez Z, Mackenzie T, Ramos M, Frisvad J, Larsen T. 2022. A molecular networking based discovery of diketopiperazine heterodimers and aspergillicins from *Aspergillus caelatus*. J Nat Prod. 85(1):25–33.

Wang Y, Shi J, Gong L. 2020. Gamma linolenic acid suppresses hypoxia-induced proliferation and invasion of non-small cell lung cancer cells by inhibition of HIF1alpha. Genes Genomics. 42(8):927–935.

Wang Z, Kim U, Liu J, Cheng C, Wu W, Guo S, Feng Y, Quinn RJ, Hou Y, Bai G. 2019. Comprehensive TCM molecular networking based on MS/MS *in silico* spectra with integration of virtual screening and affinity MS screening for discovering functional ligands from natural herbs. Anal Bioanal Chem. 411(22):5785–5797.

Wang Z, Li J, Chambers A, Crane J, Wang Y. 2021. Rapid structure-based annotation and profiling of dihydrochalcones in Star Fruit (*Averrhoa carambola*) using UHPLC/Q-Orbitrap-MS and molecular networking. J Agric Food Chem. 69(1):555–567.

Wei J, Cheng SH, Qi XM. 2019. Comparative study of effect of extracts from *Aspongopus* on experimental gastric ulcer rats. J Shanxi Coll Trad Chin Med. 20:11–16.

Wu ML, Jin DC. 2005. The antibacterial activity of the haemolymph and the purified haemo-protein from *Asporgopus chinesis*. Chin Bull Entomol. 42: 315–318.

Wu X, Lai XD, Mo JM. 2020. Molecular mechanism of *Aspongopus chinensis* aqueous extract regulating proliferation, invasion and migration of non-small cell lung cancer through mir-1287-5p/ubiquitin specific peptidase 22 pathway. Pharm Biotechno. 27:506–512.

Xu ZJ. 2019. HPLC fingerprint identification and anticoagulant activity of *Aspongopus chinensis*. Wuhan (China): Hubei University of Chinese Medicine; p. 1–61.

Yan YM, Luo Q, Di L, Shi YN, Tu ZC, Cheng YX. 2019. Nucleoside and *N*-acetyldopamine derivatives from the insect *Aspongopus chinensis*. Fitoterapia. 132:82–87.

Zhao S, Tan J, Yu HM, Tian Y, Wu YF, Luo R, Guo JJ. 2021. *In vivo* and *in vitro* antiproliferative and antimetastatic effects of haemolymph of *Aspongopus chinensis* Dallas on breast cancer cells. J Tradit Chin Med. 41: 523–529.

Zhao X, Hengchao E, Dong H, Zhang Y, Qiu J, Qian Y, Zhou C. 2022. Combination of untargeted metabolomics approach and molecular networking analysis to identify unique natural components in wild *Morchella* sp. by UPLC-Q-TOF-MS. Food Chem. 366:130642.