ISSN 2462-2052 | eISSN 2600-8718 *Journal of Science and Mathematics Letters*
DOI: https://doi.org/10.37134/jsml.vol12.1.13.2024 *Volume 12, Issue 1, 111-118, 2024*

Research Article

Mining Secondary Metabolites of *Chassalia curviflora* Leaves Using SIRIUS and High-Resolution Mass Spectrometry Data

Aina Syuhada Ammar[1], Nur Ain Syahmina Abdull Aziz[1], Nurunajah Ab Ghani[1,2]*, and Nurulfazlina Edayah Rasol[1,2]

[1]Faculty of Applied Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[2]Atta-ur-Rahman Institute for Natural Product Discovery (AuRIns), Universiti Teknologi MARA Puncak Alam Campus, 42300 Bandar Puncak Alam, Selangor, Malaysia
*Corresponding author: nurunajah@uitm.edu.my

Received: 2 October 2023; **Accepted:** 25 December 2023; **Published:** 5 January 2024

ABSTRACT

Chassalia curviflora Thwaites is a flowering plant from Rubiaceae family. Previous phytochemical investigations on *Chassalia* species led to the isolation of benzoic acid, benzoquinone derivatives, flavonoids, monoterpenoid, indole alkaloids and its essential oil composition. However, these are still scarce in terms of scientific reports on the studied species. This study was carried out to investigate the fast screening of secondary metabolites by using data-dependent acquisition (DDA) LC-MS/MS analysis of methanolic extract of the leaves of *C. curviflora* in positive ionization mode. The data were analysed using the computational mass spectrometry method from the molecular structure database search implemented in Sirius. This led to the discovering of 13 known secondary metabolites from distinct tandem MS data. De novo molecular formula annotation and predictions from SIRIUS suggested five major class compounds classified as coumarins, benzofuran, fatty acids, esters, and triterpenes. Stigmasterol **12** and β-sitosterol **13** were successfully isolated, thus used to verify the data obtained from the Sirius. With our results, we advocate the MS-based approach as a useful starting method for the dereplication of compounds from the complex crude extract of plants.

Keywords: Rubiaceae; *Chassalia curviflora*; dereplication; SIRIUS; orbitrap

1. INTRODUCTION

Chassalia curviflora (Wall ex Kurz.) Thwaites is one of the flowering plants that comes from Rubiaceae family. It also known as *wan guan hua, pokok jarum-jarum, pokok beras-beras,* or *pokok beberas.* It can be found in the northern states of Peninsular Malaysia and in other Asia countries such as India, China, and Indonesia (Gopal et al., 2016). *C. curviflora* can grow as a weed and every part of this plant has been used in traditional medicine for the past few decades. The decoction of *Chassalia* leaves and roots has been proven to be effective in anti-inflammatory, analgesic, hepatoprotective, sedative, anti-epileptic, antioxidant, anti-hypertensive, and antimicrobial (Lohia et al., 2023). Thus, it has been used as a traditional medicine by the practitioners of the tribes of Kerala in treating wounds and insect bites and as

an antidote for snake bites (Lohia et al., 2023). The phytochemical screening proved that *C. curviflora* has several classes of secondary metabolites such as alkaloids, carbohydrates, phytosterols and flavonoids (Palayullaparambil et al., 2016).

This study was carried out to investigate the ethnomedical use and phytochemical constituents present in the leaves of the *C. curviflora* species using advanced techniques. MZmine and Sirius were used to annotate the compounds with the aid of a custom database extracted from DNP and literature. To verify the annotated compounds, phytoconstituents were isolated using various chromatographic techniques. Retention time, precursor ions, and fragmentation pattern of the isolated compounds were compared to verify the data obtained from the Sirius.

2. MATERIALS AND METHODS

2.1. Plant material

A sample of *C. curviflora* Thwaites was collected from Behrang, Perak in September 2019, and identified by Shamsul Khamis from Universiti Kebangsaan Malaysia (UKM). The voucher specimen (SK131/19) was deposited at UKMB Herbarium, UKM. The dried leaves were extracted with methanol to give 0.37 g of methanol crude.

2.2. Metabolites profiling

The methanol crude (20 mg) was subjected to SPE-C18 to yield 1.7 mg (8.5% recovery) of the enriched fraction. SPE is used to remove sample interference (in this case is chlorophyll) by flushing with methanol and conditioned with water: methanol ratio (5:95, v/v). The sample was then transferred to an HPLC vial through a 0.20 µm UHPLC filter.

Ultra-High-Performance Liquid Chromatography (UHPLC): The crude was transferred to the Eppendorf tube and dissolved in 1 mL of solvent mixture (95% methanol and 5% water). The tube was sonicated for a few seconds and the solution was transferred to an HPLC vial through a 0.22 µm syringe filter UHPLC filter. This sample preparation is mandatory before injection into the UHPLC system. This study used a Luna Omega C18 column (100 x 2.1 mm, 1.6 µm) as the nonpolar stationary phase with a diode array detector (DAD). The column temperature was set at 45 ºC, and the system flow rate was set to 0.2 mL/min followed by an injection volume of 1 µL. Gradient elution was conducted using deionized water as the polar mobile phase of deionized water (% A) and organic solvent, methanol (% B). The gradient elution was carried out in a mixture of deionized water (A) and methanol (B) with following solvent system: 0-30 min (10-100% of B) and 30-35 min (100% B).

Liquid Chromatography-Mass Spectrometry (LC-MS): In this study, a Thermo Scientific Vanquish system with an Accucote™ Vanquish C18 column of 2.1 × 100 mm, 1.5 µm was used. The column temperature was set at 30 ºC. The sample was prepared with the concentration of 250 ppm, dissolved in a mixture of methanol and water. The flow rate of the system was set up to 0.2 mL/min and 1 µL for the injection volume. LCMS grade water (% A) and methanol (% B) were used in the analysis, with the elution system 0-30 min (10-100% of B) and 30-35 min (100% B). The Thermo Scientific™ Orbitrap Fushion™ Tribrid™ mass spectrometer was applied together with electrospray ionization (ESI) to identify molecular ions. The sample was screened using positive ionization mode and was analyzed using multistage mass spectrometry (MS/MS) data analysis, where the product ions were subjected to defragmentation.

2.3. Data mining

In this study, the dereplication technique will be used, particularly, to compare the molecular mass and identify the known and unknown compounds in the sample. For this, a database of known molecular masses will be constructed. The data mining process for the custom database (DNP) will be carried out by extracting the information on the chemical constituents of the Rubiaceae family based on the chemical name, synonym(s), molecular formula, and accurate mass parameters from the DNP. A total of 2389 compounds of the Rubiaceae family of DNP and a total of 30 compounds were reported from the literature review on *C. curviflora*. For data processing, two free downloaded software (MZmine and Sirius) were used for data analysis. Flowchart of data processing using MZmine is shown in Figure 1.

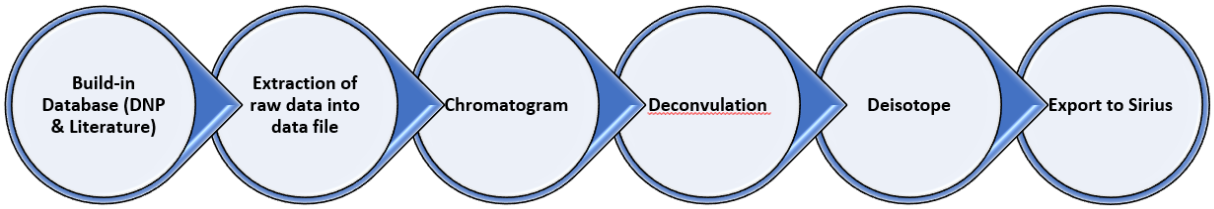


Figure 1. Steps of data processing using MZmine

2.4. Extraction and purification

Approximately 350 mg of methanol extract was subjected to an open column of size exclusion chromatography (Sephadex LH-20) using an isocratic elution of dichloromethane and methanol (1:1 ratio) for fractionation to yield 6 fractions. Fraction 3 (28.9mg) was selected for isolation and purification. Isolation and purification were performed using recycled High Performance Liquid Chromatography (RHPLC) using a reverse phase of JAIGEL-ODS-AP column (20 x 250 mm), column at 4 mL/min (isocratic solvent system acetonitrile-water, 80:20) at 210 nm to produce 3 sub-fractions. Subfractions 2 (2.48 mg, **12**) and 3 (3.25mg, **13**) were obtained as a white solid, respectively, and was subjected to NMR and LCMS for further analysis.

3. RESULTS AND DISCUSSION

The sample was subjected to UHPLC analysis to obtain the UV profile and LCMS analysis to obtain its MS profile. Figure 2 shows a chromatogram of *C. curviflora* methanol extract. The chromatogram was then analysed using Thermo Scientific Freestyle 1.6 SP1 modern data visualization software. There is one major peak that can randomly be observed. Information such as UV maximum mass spectral (MS) of the peak can be observed using this software. The data mining process began with gathering all compound IDs from the online database Dictionary of Natural Products (DNP). The IDs of the compounds were extracted using the keywords of family names and species name. The identification numbers were extracted and tabulated in an Excel file. In addition, IDs from the literature were recorded as well. Four parameters were set up for data mining purposes: chemical name, molecular formula, synonym(s), and accurate mass. Since the mass data was recorded in the positive mode, it was set to $[M+H]^+$ for the accurate mass.

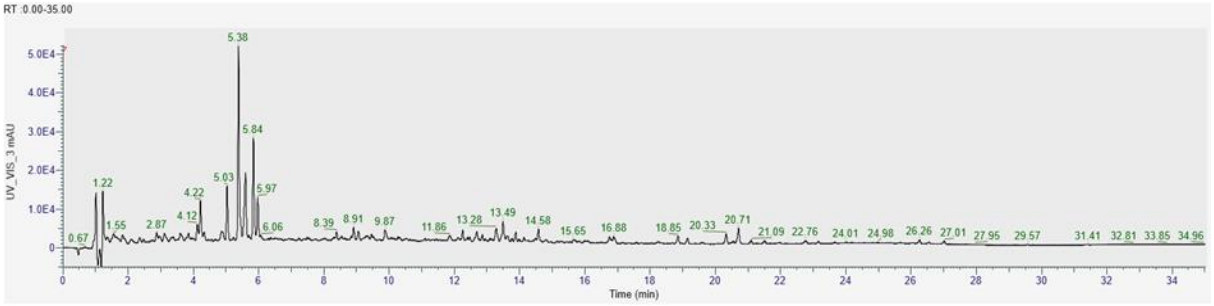


Figure 2. UV profile of methanol extract of the leaves of *C. curviflora* observed at 210 nm

Data processing started with data clean up using MZmine software. Using this software, the noise level was set at 2.0E6, with the minimum highest intensity set at 5.0 E6 and the mass tolerance was set at 0.0015 *m/z* or 5.0 ppm. The data processing continued with ADAP resolver where the overlapping peaks were defined and finally deisotope processing where all isotope peaks were identified and filtered. The process began with importing data from MZmine, followed by databases extracted from DNP, *C. curviflora* genus and family of Rubiaceae) and literature. The process continued with annotation by computing all steps. The time required is determined by the number of features on the job list. It is advised to use a slightly higher MS2 with a mass accuracy value of (10). The dataset from which the ion identity network information generated by MZmine was used to select possible ionisations and fallback adducts. The process and annotation status can be viewed in job listings. The job order can be arranged by cosmic arrangement. COSMIC is an SVM classifier that employs a large number of score differences. The higher the number, the more certain the annotation is.

For the structure annotation option, different colours of the peaks are shown in the spectrum. Purple peaks represent the substructure annotations explained by the candidate. Then, the black peaks, which were mostly small peaks behind the purple peaks, were not explained by the structure and molecular formula, while the green peaks had a formula candidate but could not be explained by the structure. The data were exported in tabulated Excel form in an Excel workbook form. The proposed compound is based on its mass, peak shape and fragmentation. All compounds proposed to be present in the *C. curviflora* sample are based on the online database (DNP) and a review of the literature. Figure 3 shows the chromatogram based on MS2 and the retention time of the LC-MS detected peaks. Table 1 shows the list of annotated compounds. A total of 13 compounds were predicted to be present in the methanolic extract of *C. curviflora.* Identified compounds were further analysed and were identified to be from coumarin, benzofuran, fatty acids, esters, and triterpenes classes of compounds.

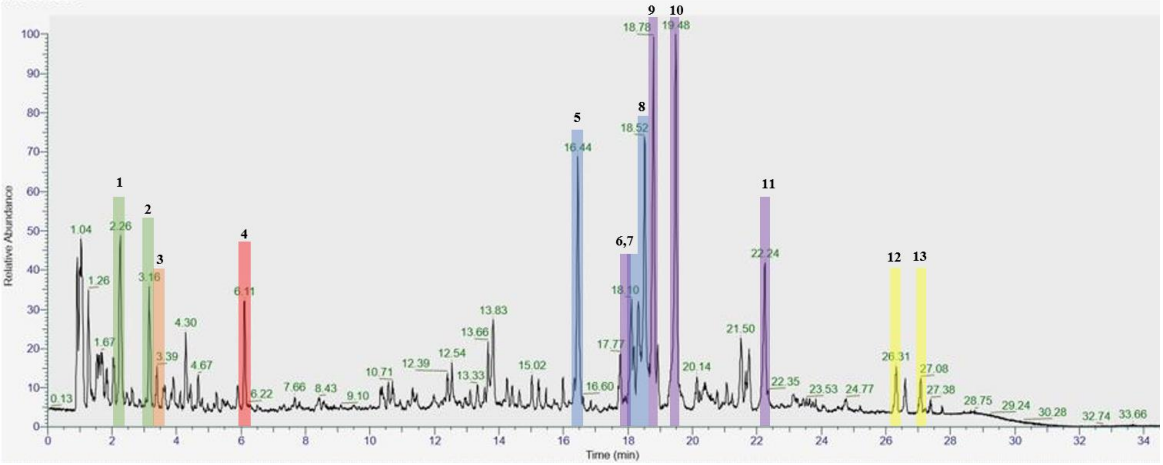
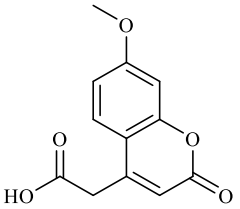


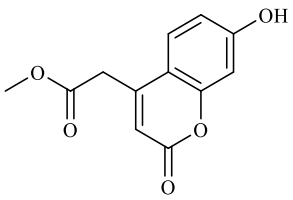
Figure 3. MS profile of methanol extract of the leaves of *C. curviflora*

Table 1. Proposed compounds from the methanol extract of the leaves of *C. curviflora*

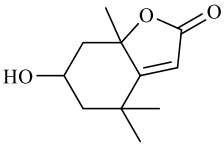
Peak	tR (min)	Compound identification	Molecular formula	Precursor ions	Ion type	Key fragments	Sirius similarity (%)
1	2.26	7-Methoxycoumarin-4-acetic acid	$C_{12}H_{10}O_5$	235.06	$[M+H]+ H^+$	175.04, 119.05	78.63
2	3.16	Methyl 7-hydroxycoumarin-4-acetate	$C_{12}H_{10}O_5$	235.06	$[M+H]+ H^+$	175.04, 147.04, 119.05	79.71
3	3.40	Octaethylene glycol	$C_{16}H_{34}O_9$	388.25	$[M+H]+H_2O^+$	371.23, 133.09	100.00
4	6.11	Loliolide	$C_{11}H_{16}O_3$	197.12	$[M+H]+ H^+$	179.11, 107.09	69.15
5	16.44	Glycine-N-{17-L-[(2'-O-.beta.-D-glucopyranosyl-.beta.-D-glucopyranosyl)oxy]-cis-9-octadecenamide} methyl ester	$C_{33}H_{59}NO_{14}$	694.39	$[M+H]+ H^+$	515.32, 353.27, 261.22	64.02
6	18.08	2-hydroxy-3-{[3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxy}propyl octadeca-9,12,15-trienoate	$C_{27}H_{46}O_9$	532.25	$[M+H]+H_2O^+$	261.22, 335.26, 353.27	100.00
7	18.10	Parinaric acid	$C_{18}H_{28}O_2$	277.22	$[M+H]+ ^+$	221.15, 149.13	86.21
8	18.52	2-{2-[(17-acetamido-4,16-dihydroxy-3,7-dimethylheptadecan-5-yl)oxy]-2-oxoethyl}butanedioic acid	$C_{27}H_{49}NO_9$	532.35	$[M+H]+ H^+$	353.27, 335.36, 261.22	34.40
9	18.78	Methyl 12,13-epoxystearate	$C_{19}H_{36}O_3$	313.27	$[M+H]+ H^+$	331.28, 239.24, 109.10, 95.08	69.54
10	19.48	Monoolein	$C_{21}H_{40}O_4$	357.30	$[M+H]+ H^+$	339.29, 265.25	99.52
11	22.24	[1-Hydroxy-2-octadecanoyloxy-5-(1,2,3-trihydroxypropoxy)pentan-3-yl]oxyboronic acid	$C_{26}H_{53}BO_{10}$	536.28	$[M+H]+ H^+$	357.30, 339.29, 265.25	68.02
12	26.31	stigmasterol	$C_{29}H_{51}O$	415.36	$[M+H]^+$	381.4, 329.3, 281.0, 213.1	100.00
13	27.38	β-sitosterol	$C_{29}H_{49}O$	413.40	$[M+H]^+$	379.3, 351.3, 300.1, 255.2	100.00



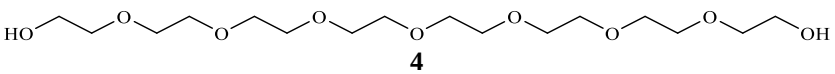
1



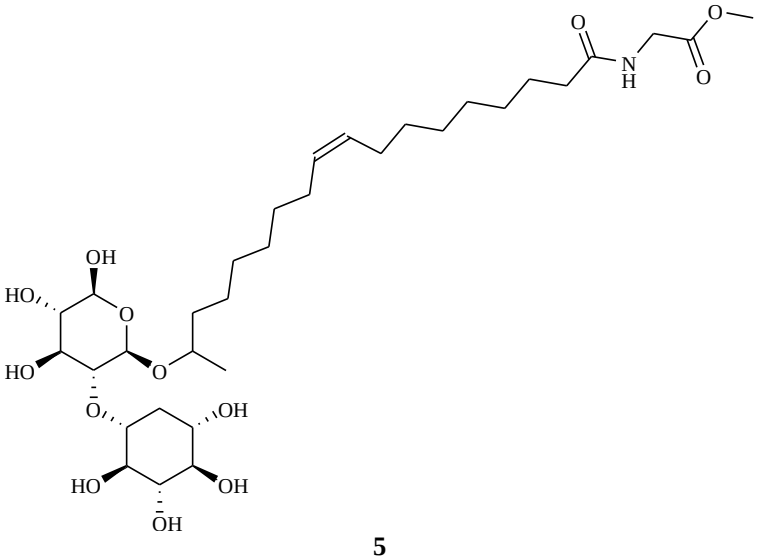
2



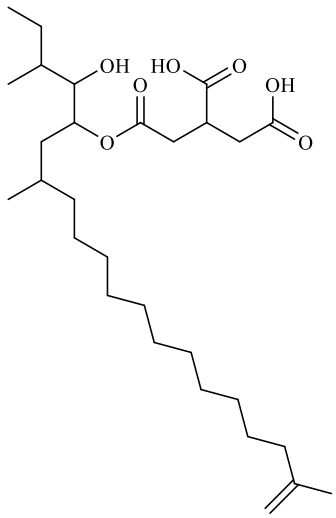
3



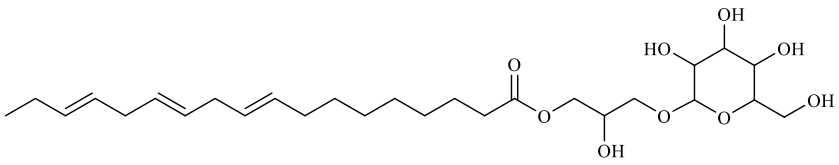
4



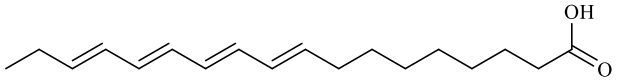
5



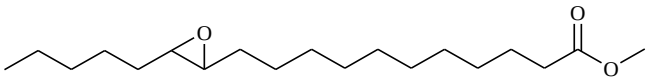
8



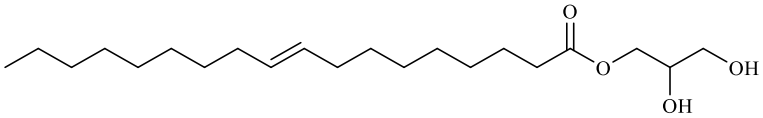
6



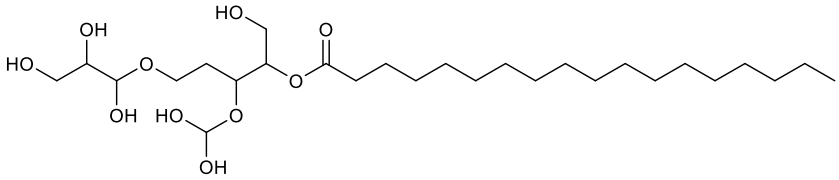
7



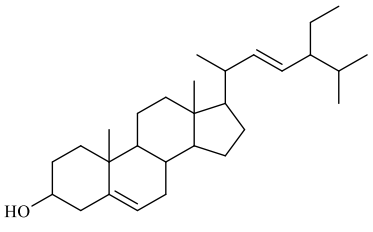
9



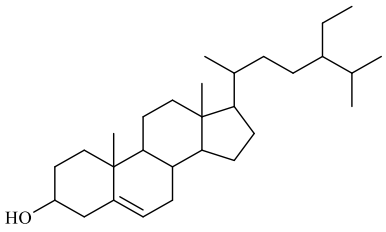
10



11



12



13

The isolated terpenoid from this sample was characterized using modern spectroscopic techniques such as ^{1}H and ^{13}C NMR. The terpenoids isolated from the leaves of *C. curviflora* are triterpene with six isoprene units. Triterpenes are the major constituents of some medicinal plants and herbs and are widely present in all parts of plants. Triterpenes labels as **12** (2.48 mg) and **13** (3.25 mg) were isolated as a white amorphous solid. The mass spectrum of compounds **12** and **13** shows the molecular ion peak $[M+H]^{+}$ at *m/z* 415 and *m/z* 413

corresponding to the molecular formula of $C_{29}H_{51}O$ and $C_{29}H_{49}O$, respectively. The ^{1}H and ^{13}C NMR data of the isolated compounds are tabulated in Table 2.

Table 2. Comparison ^{1}H (δ ppm) and ^{13}C (δ ppm) of compounds **12** and **13**

Position	12		13	
	^{1}H (δ ppm)	**^{13}C (δ ppm)**	**^{1}H (δ ppm)**	**^{13}C (δ ppm)**
1	-	37.3	-	37.3
2	-	31.6	-	31.6
3	3.55, m	71.8	3.55, m	71.8
4	-	42.3	-	42.3
5	-	140.7	-	140.7
6	5.37, d (J = 5.4 Hz)	121.7	5.37, d (J = 5.4 Hz)	121.7
7	-	31.9	-	31.9
8	-	31.9	-	31.9
9	-	50.2	-	50.1
10	-	36.2	-	36.2
11	-	21.1	-	21.1
12	-	39.8	-	39.7
13	-	42.3	-	42.3
14	-	56.7	-	56.8
15	-	24.4	-	24.3
16	-	28.9	-	28.3
17	-	55.9	-	56.0
18	0.69, d (J = 5.4 Hz)	12.0	0.69, d (J = 5.4 Hz)	11.9
19	1.13, s	19.4	1.13, s	19.4
20	-	40.5	-	36.2
21	0.94, d (J = 6.6 Hz)	21.1	0.94, d (J = 6.6 Hz)	18.8
22	5.06, dd (J = 15.0, 8.4 Hz)	138.3	5.06, dd (J = 15.0, 8.4 Hz)	33.9
23	5.15, dd (J = 15.0, 8.4 Hz)	129.3	5.15, dd (J = 15.0, 8.4 Hz)	26.0
24	-	51.2	-	45.8
25	-	31.9	-	29.1
26	0.81, d (J = 6.6 Hz)	21.2	0.81, d (J = 6.6 Hz)	19.8
27	0.85, d (J = 6.6 Hz)	19.0	0.85, d (J = 6.6 Hz)	19.0
28	-	25.4	-	23.1
29	0.87, t (J = 6.6 Hz)	12.3	0.87, t (J = 6.6 Hz)	12.1

The ^{1}H NMR shows a methine proton which was assigned as H-3 for both compounds, resonated at a more downfield region compared to the other methine signal H-3 position next to the hydroxyl group. The doublet signals at δ 5.37 with the coupling constant 5.4 Hz suggested that H-6 olefinic protons of both of compounds **12** and **13**. Six methyl signals could be observed at δ 0.69 (H-18), 1.13 (H-19), 0.94 (H-21), 0.81 (H-26), 0.85 (H-27), and 0.87 (H-29). Comparing the signal of three methyl groups H-18, H-19, and H-21, the H-18 signal is more shielded due to the influence of H-21 proton next to it. Two pairs of protons with integration of two protons were observed and labelled as H-22 and H-23, respectively, with a doublet of doublet signals (*J* = 15.0 and 8.4 Hz, 1H). H-23 is more shielded compared to H-22 due to its position next to H-21 which is more protonated. Determination of both compounds was very difficult because of the complexity of their structure that is quite similar. Moreover, the appearance of several overlapping signals corresponding to methyl, methylene, and methine protons that resonate between δ 0.50 to 2.50 make the elucidation of these compounds become more difficult. Comparison of the spectroscopic data with the literature (Chang et al., 2000; De-Eknamkul and Potduang, 2003) helped to solve this problem and of compounds **12** and **13** was confirmed to be stigmasterol and β-sitosterol, respectively.

4. CONCLUSION

To conclude, there were 13 compounds were successfully identified using LCMS-Orbitrap, after comparing with the custom data extracted from Dictionary of Natural Products and literature. Purification and isolation work were performed on the methanolic extract of the leaves of *C. curviflora* in order to verify the compounds annotated in the dereplication. Stigmasterol **12** and β-sitosterol **13** were successfully isolated. The retention time, mass, and fragmentation pattern of the isolated compounds were compared with the data obtained from the dereplication, thus verifying the data obtained.

Declaration of Interest
The authors declare that there is no conflict of interest.

Acknowledgement
This research was supported by Geran Penyelidikan Center of Excellence (CoE) UiTM Cawangan Selangor (DUCS-CoE) (600-UITMSEL (P.I5/4) (38/2022) from Universiti Teknologi MARA, Malaysia. The authors would like to express appreciation to Atta-ur-Rahman Institute for Natural Product Discovery for the facilities provided.

REFERENCES

Chang YC, Chang FR, Wu YC. (2000). The constituents of *Lindera glauca*. *Journal of Chinese Chemical Society*, 47, 373-380.

De-Eknamkul W, Potduang B. (2003). Biosynthesis of β-sitosterol and stigmasterol in *Croton sublyratus* proceeds via a mixed origin of isoprenes units. *Phytochemistry*, 62, 389-398.

Gopal. GR, Anuja GI, Latha PG, Nair AJ, Gangaprasad A, Sudhakaran PR, Muraleedharan D, Nair GM. (2016). Anti-hepototoxic property of roots and leaves of *Chasssalia carviflora* (Wall.) Thwaites against CCl_4 induced liver damage in rats. *International Journal of Pharmaceutical Sciences and Drug Research*, 8(5), 269-274.

Lohia A, Naik TI, Prabhu S. (2023). Conceptual and quality control analysis of *Chassalia curviflora* -An extrapharmacopeial Plant. *World Journal of Pharmaceutical Research*, 12(11), 882–896.

Palayullaparambil AK, Palayullaparambil AK, Juliet S, Renganathan K, Raju R, Athalathil S. (2016). Pharmaco-chemical characterization and acaricidal activity of ethanolic extract of *Chassalia curviflora* (Wall ex Kurz.) Thwaites. *Pharmacognosy Journal*, 8(3), 215-219.