

World News of Natural Sciences

An International Scientific Journal

WNOFNS 64 (2026) 134-150

EISSN 2543-5426

UHPLC-Q-TOF-MS Phytochemical Characterization and ATR-FTIR analysis Combined with Machine Learning for *Asparagus racemosus* Leaves

Do Ngoc Thuy^{1,*}, Nguyen Thi Huong¹, Le Ngoc Hung²,
Mai Van Nam³, Tran Van Thanh⁴

¹ Center for High Technology Research and Development,
Vietnam Academy of Science and Technology, Hanoi, Vietnam

² VIPTAM Institute of Technology Application, Hanoi, Vietnam

³ Hanoi University of Science and Technology, Hanoi, Vietnam

⁴ Faculty of Medicine and Pharmacy, Trung Vuong University, Ha Dong, Hanoi, Vietnam

*E-mail address: dothuy1992831@gmail.com

ABSTRACT

The medicinal plant *Asparagus racemosus* Wild., has been widely used in traditional medicine to cure various diseases such as cough, diabetes, gastric issues, diarrhoea, headaches, and rheumatism. This paper studies phytochemicals of the species native to Gia Lai province, Vietnam, before its conservation and cultivation. *A. racemosus* leaves were extracted with water then UHPLC-Q-TOF-MS spectroscopy was used to screen for their phytochemicals. In water extract of *A. racemosus* leaves, UHPLC-Q-TOF-MS tentatively identified two flavonoids (Quercetin-3-glucuronide, Rutin), five steroidal saponins (Shatavarin I, Shatavarin IV, Shatavarin IX, Asparacoside, Asparanin A), and two steroids (β -sitosterol, Daucosterol). One objective of this study was to identify and classify the chemical differences between leaves of *Asparagus racemosus* collected in different places of Gia Lai using Attenuated Total Reflection - Fourier Transform Infrared spectroscopy (ATR-FTIR) combined with machine learning algorithms. A total of 40 samples of *A. racemosus*'s leaf powder were analyzed using ATR-FTIR spectroscopy. The Savitzky-Golay algorithm, unsupervised (Hierarchical Clustering) and supervised (Principal Component Analysis and Support Vector Machine) machine learning algorithms were employed to classify the ATR-FTIR spectra. The combination of ATR-FTIR spectral data and machine learning algorithms successfully discriminated and classified the collected places of *A. racemosus* leaves. The experimental findings confirm the *A. racemosus* species for conservation and

cultivation in Vietnam and contribute the benefits to the chemical literature of Vietnamese natural flora. *A. racemosus* should be further studied for its pharmaceutical activities.

Keywords: *Asparagus racemosus*, leaves, UHPLC-Q-TOF, ATR-FTIR, chemometrics

1. INTRODUCTION

Asparagus racemosus Willd., belonging to the family Liliaceae, is generally found in India, some regions of the Himalayas, and all over the region of Sri Lanka, Asia, Australia, etc. [1]. In Vietnam, it is only distributed in Gia Lai province [2]. The whole plant of *A. racemosus*, including leaves and roots, is especially useful in traditional medicine. Its tuberous roots are used for diabetes, dementia, bronchitis, diuretic, dyspepsia, emollient, rejuvenating, constipation, diarrhea, and stomachic activities, etc [3]. As the demand for *A. racemosus* is constantly on the rise, the supply is inadequate and now considered 'endangered' in its natural habitat. It is crucial to carry out its conservation, especially in Vietnam.

Chemical constituents from the plant roots include several steroidal saponins such as shavatarin, asparanins, asparagine, asparosides, curillins, curillosides, immunoside, racemoside A, shatavaroside C, and 9,10-dihydrophenanthrene derivative; several acids such as ferulic acid, isoferulic acid, malic acid, citric acid, asparagusic acid, caffeic acid, and fumaric acid [4, 5]. Flowers and mature fruits contain quercetin, rutin, and hyperoside.

The phytochemical evaluation of the *Asparagus racemosus* leaves was performed and showed excessive amounts of steroids and triterpenoids [6], flavonoids such as diosgenin and quercetin-3-glucuronide [7].

The ultra-performance liquid chromatography coupled with quadrupole/time-of-flight mass spectrometry (HPLC-QTOF-MS/MS) is capable of accurately measuring molecular mass by giving the elemental composition of obtained ions and is widely used in analyzing complex samples due to the high resolution and sensitivity. UHPLC-QTOF-MS was applied to characterize chemical constituents and metabolites in medicinal herbs, and obtained considerable results. HPLC-Q-TOF analysis of *A. adscendens* roots tentatively assigned the possible presence of saponins and spirostanol [8]. The HPLC-Q-TOF-MS/MS has been successfully applied to analyze phytochemicals and even to simultaneously quantify five saponin glycosides, asparacoside, shatavarin IX, shatavarin IV, asparanin A, and shatavarin V in *A. racemosus* roots [4, 9].

Fourier Transform Infrared spectroscopy (FTIR) in the middle IR wavenumber region between 4000 cm^{-1} to 400 cm^{-1} is an established time-saving method to characterize functional groups of metabolites in medicinal plants [10]. FTIR spectroscopy analysis can be used with the Attenuated Total Reflection technique (ATR) to qualitatively analyze the functional groups present in the extracts or medicinal plants [11]. Pei YF et al have used the ATR-FTIR technique to analyze the essential oil of both leaf and rhizome of *Paris yunnanensis* species and then to apply different chemometrics techniques for data treatment [12].

Several FTIR studies of *Asparagus racemosus* roots showed the presence of C-H bond stretching, alkynes group, an aromatic ring, N-H bending at primary amine, secondary amine stretching at secondary amine and alkenes group stretching compounds with major peaks at $3429\text{--}3412\text{ cm}^{-1}$, $2359\text{--}2330\text{ cm}^{-1}$, $1967\text{--}1923\text{ cm}^{-1}$, 1398 cm^{-1} , $1645\text{--}1568\text{ cm}^{-1}$, and 3149.86 cm^{-1} respectively [13-15].

Principal Components Analysis (PCA) is a technique used to condense a complex dataset into a more manageable set of underlying variables, called Principal Components (PCs). When dealing with a large number of spectral bands (input variables), reducing the number of variables to a few linear combinations of these bands is essential for meaningful data interpretation [16]. In chemometrics and machine learning, the PCA algorithm is frequently employed to reduce the dimensionality of datasets with numerous features to a more manageable number. This application of PCA helps models circumvent the issues associated with the curse of dimensionality [17].

PCA identifies PCs, which are axes in a high-dimensional space designed to maximize the variance of the projected data points. Subsequent PCs are orthogonal to the initial PC and exhibit progressively lower variance. Through orthogonal projection onto a subspace, PCA effectively preserves a significant portion of the original information with fewer features, while simultaneously mitigating noise by discarding PCs with minimal variance [18]. The Support Vector Machine (SVM) algorithm was developed to solve binary classification problems, specifically in cases where data points in a space can be linearly separated. The SVM algorithm identifies a hyperplane that divides the data clusters in the most fair and equal manner, ensuring that the distance from the hyperplane to the nearest data points of both classes (the support vectors) is maximized. The data does not necessarily need to be linearly separable to use the SVM algorithm. The SVM model that allows for noise and overlapping data is known as the soft margin SVM, while the model that maps data to a higher-dimensional space to construct a separating hyperplane is known as the kernel SVM [19].

The ATR-FTIR spectroscopy is used to establish the chemical information fingerprint of many medicinal plants, and their data are further discriminated by applying chemometrics analysis, such as PCA, HCA, and OPLS-DA [20, 21]. However, the chemometric analysis was only used to separate phenolic groups of the *Asparagus racemosus* root extracts by different extraction supercritical fluid (CO₂), soxhlet, and maceration-based methods [22].

In this research, our first step for the conservation of the plant before its cultivation in the appropriate lands is to study the phytochemistry of *A. racemosus* leaves by UHPLC-Q-TOF-MS spectroscopy. Then, we investigate the influence of ATR-FTIR spectroscopy combined with multivariate statistical analysis to identify and classify the leaves of the *A. racemosus* species collected in different places of Gia Lai province, Vietnam.

2. METHODOLOGY

2. 1. Chemicals and reagents

All reagents were purchased from Sigma-Aldrich Chem. Co. (St. Louis, MO., USA). Deionized water for HPLC; HPLC grade acetonitrile, methanol, and analytical grade formic acid (≥98%) were obtained from Fisher (USA).

2. 2. Plant material

The leaves of *A. racemosus* from the wild were collected in different places of Gia Lai Province, Vietnam, in October 2023. Samples were primarily recognized based on the morphology in the field, then preserved in silica gel-filled bags, and transferred to the Vietnam Academy of Science and Technology for final identification by Dr. Nguyen Sinh Khang, Institute of Ecology and Biological Resources, Vietnam Academy of Science and Technology

(VAST), before preservation at -30 °C. The voucher specimens have been lodged at the Center for High Technology Development (CHTD, VAST). All fresh leaves were cut off from the plants, dried, powdered, and stored in a fridge for further water extract, HPLC-QTOF chemical analysis, and ATR-FTIR analysis.

2. 3. Water extraction

The fresh leaves of *A. racemosus* are washed, sliced, and dried for 5 days in an oven at 40 °C and ground into powder. Dried powder of the leaves of *A. racemosus* was extracted with water (3 times) at room temperature in an ultrasonic 20 kHz and 1 kW extractor. The combined extract was concentrated to dryness under reduced pressure, yielding a residue (brown solid). The residue of water extract was flushed with nitrogen gas and stored at -20 °C for future use [5].

2. 4. UHPLC-Q-TOF-MS analysis

Sample analysis by the UHPLC-Q-TOF-MS instrument was investigated with a procedure described previously [5, 23]. 100.0 (mg) of the extract was accurately weighed into a tube with a cover, and 2.0 (mL) methanol–water (8:2, v/v) solvent was added. The sample was ultrasonicated for 10 min. The sample was filtered through a 0.45 µm filter membrane before being injected for UHPLC-Q-TOF analysis. Sample analysis was performed on an Exion LCTM UHPLC system (AB SCIEX, USA) consisting of Exion LC degasser, AC pumps, AC autosampler, controller, and AC column oven. Samples were analyzed on a Hypersil GOLD C18 column (150 × 2.1 mm, 3µ) (Thermo Fisher Scientific, USA). The mobile phase, water containing 0.1% formic acid (A) and acetonitrile containing 0.1% formic acid (B), was run at a flow rate of 0.4 mL/min at room temperature. The gradient programming was as follows: 0–4 min, 2-20% B; 4-30 min, 20-68% B; 30-32 min, 68-98% B; 32-40 min, 98% B. Sample injection volume was 5.0 (µL).

An AB SCIEX X500R QTOF mass spectrometer (AB SCIEX, USA) with a Turbo V ion source was coupled with the UHPLC system. Mass data were acquired in both negative and positive Electrospray Ionization (ESI) modes. The ESI-MS conditions were set as follows: the ion source temperature, 500 °C; curtain gas, 30 psi; nebulizer gas (GS 1), 45 psi; heater gas (GS 2), 45 psi. For the TOF MS scan, the mass range was set at m/z 70–2000. For the TOF MS/MS scan, the mass range was set at m/z 50–1500.

For the negative mode, ion spray voltage was set at -4.5 kV, the declustering potential (DP) was -70 V, the collision energy (CE) was performed at -20 eV, and the collision energy spread (CES) was 10 eV. For the positive mode, ion spray voltage was set at 5.5 kV, the DP was 80 V, the CE was 20 eV and the CES was 10 eV. All the obtained data were processed by SCIEX OS software version 1.2.0.4122 (AB SCIEX, USA).

2. 5. ATR-FTIR characterization

The dried *A. racemosus* leaves were pulverized again and then sieved using mesh sizes ranging from no. 100 to no. 120 to isolate particles with diameters lying within the 0.125-0.150 mm range before ATR-FTIR analysis. The ATR-FTIR measurements were conducted using an FTIR Spectrometer (Nicolet Apex, Thermo Scientific), covering a wavelength spectrum between 4000⁻¹ and 400 cm⁻¹ as described elsewhere with minor modification [12]. A total of 128 scans with a spectral resolution of 4 cm⁻¹ were performed on each sample.

2. 6. ATR-FTIR spectra data pre-processing

The ATR-FTIR spectra of forty samples were randomly divided into a training set (32 samples, estimated at 80%) and a test set (08 samples, 20%). The construction of the unsupervised learning model PCA and the supervised learning model SVM was carried out on the training dataset, while classification evaluations were conducted on the test dataset. Before model building, data preprocessing techniques, including auto-scaling and mean-centering, were applied to the entire dataset. The second derivative and smoothing using the Savitzky-Golay algorithm were utilized to enhance the resolution of the spectral bands and the quality of the spectra. The PCA and SVM algorithms, along with the data processing, were executed using Python 3.11.

2. 7. Principal Components Analysis (PCA)

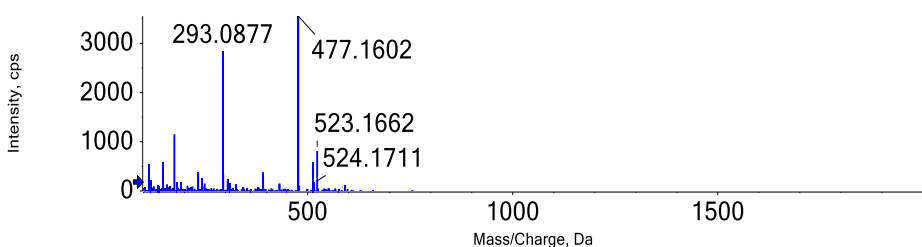
The use of the PCA algorithm in this study, with 32 training samples and 08 test samples, aims to reduce the dimensionality of a large number of features (902 features) to a few features (corresponding to a few initial PCs). The SVM model is constructed based on 10 training samples that have been reduced to a 2-dimensional space using PCA, effectively creating a decision boundary that best distinguishes these 10 samples. This decision boundary must have a uniform margin to the nearest support vectors and should maximize this margin.

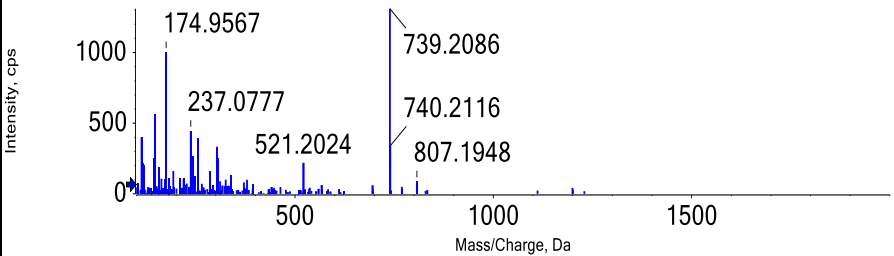
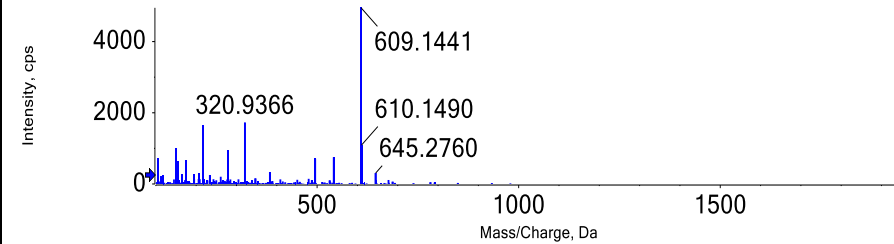
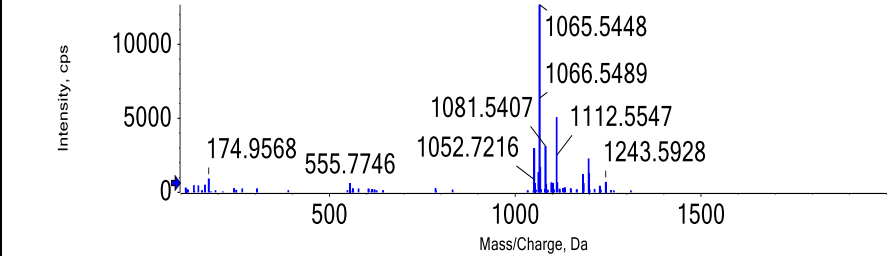
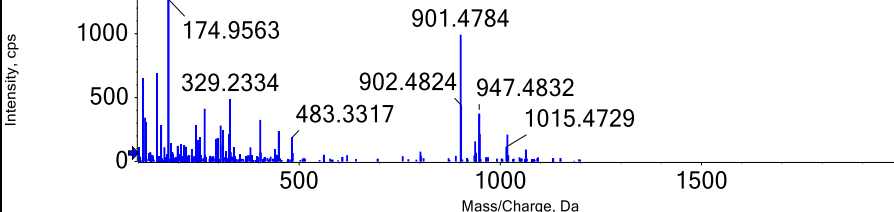
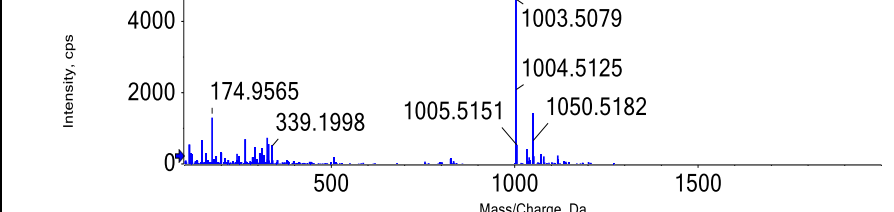
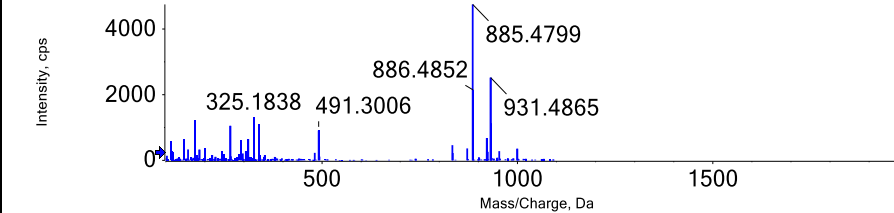
The dimensionality-reduced data will be displayed to check for linear discriminative ability, thereby constructing a binary SVM classification model. The use of PCA with the derived and smoothed dataset does not select the number of PCs based on the total % variance explained (since the differences in the signal have been amplified after derivation compared to the original) but rather on the suitability and optimal classification ability of the subsequent classification algorithm. Therefore, it is expected that reducing the dimensionality to 2 PCs will allow the classification model (SVM) to achieve the best classification performance, corresponding to the ability to recognize linear discrimination based on the 2 PCs score plot.

3. RESULTS AND DISCUSSION

3. 1. UHPLC-Q-TOF-MS qualitative analysis of *A. racemosus* leaves

Table 1(A). HPLC-ESI-QTOF NEG assignments of compounds in *A. racemosus* leaves.

RT	M/z and compound	MS spectrum
7.32	477.1616 Quercetin-3-glucuronide	Spectrum from TMC_KC_MY_L_23-2-22_NEG.wiff2 (sample... multiplier = 1.5), Gaussian smoothed (0.5 points)  The mass spectrum displays relative intensity on the y-axis (0 to 3000 cps) against mass-to-charge ratio on the x-axis (0 to 1500 Da). The base peak is at m/z 293.0877. Other significant peaks are labeled at m/z 477.1602, 523.1662, and 524.1711.

8.27	739.2095 Asparanin A	Spectrum from TMC_KC_MY_L_23-2-22_NEG.wiff2 (sample... multiplier = 1.5), Gaussian smoothed (0.5 points)  Intensity, cps Mass/Charge, Da
8.47	609.1457 Rutin	Spectrum from TMC_KC_MY_L_23-2-22_NEG.wiff2 (sample... multiplier = 1.5), Gaussian smoothed (0.5 points)  Intensity, cps Mass/Charge, Da
9.75	1065.5461 Shatavarin I	Spectrum from TMC_KC_MY_L_23-2-22_NEG.wiff2 (sample... multiplier = 1.5), Gaussian smoothed (0.5 points)  Intensity, cps Mass/Charge, Da
11.14	901.4792 Shatavarin IX	Spectrum from TMC_KC_MY_L_23-2-22_NEG.wiff2 (sample... multiplier = 1.5), Gaussian smoothed (0.5 points)  Intensity, cps Mass/Charge, Da
12.79	1003.5094 Asparacoside	Spectrum from TMC_KC_MY_L_23-2-22_NEG.wiff2 (sample... multiplier = 1.5), Gaussian smoothed (0.5 points)  Intensity, cps Mass/Charge, Da
14.06	885.4821 Shatavarin IV	Spectrum from TMC_KC_MY_L_23-2-22_NEG.wiff2 (sample... multiplier = 1.5), Gaussian smoothed (0.5 points)  Intensity, cps Mass/Charge, Da

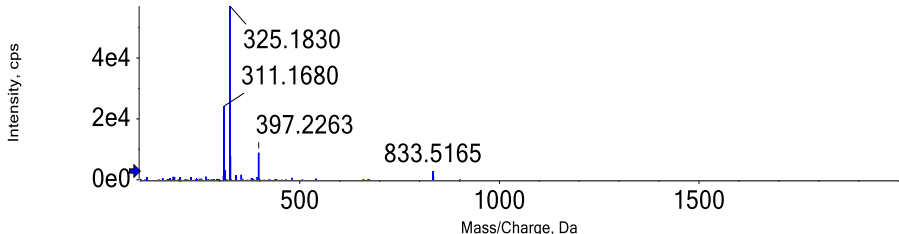
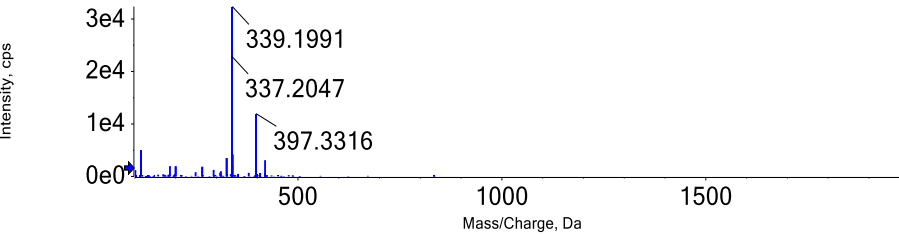
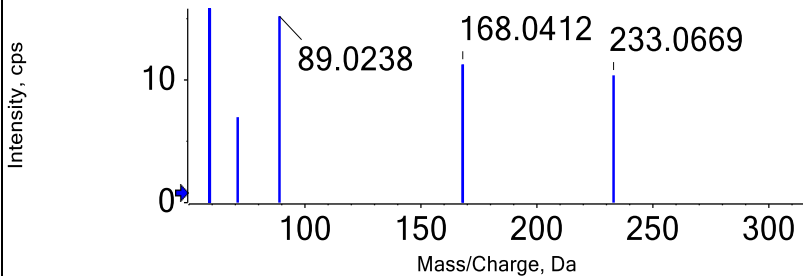
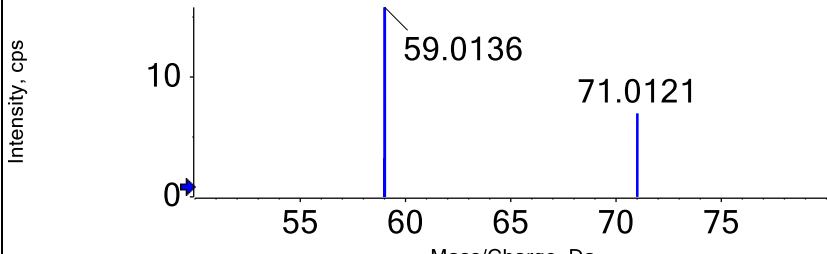
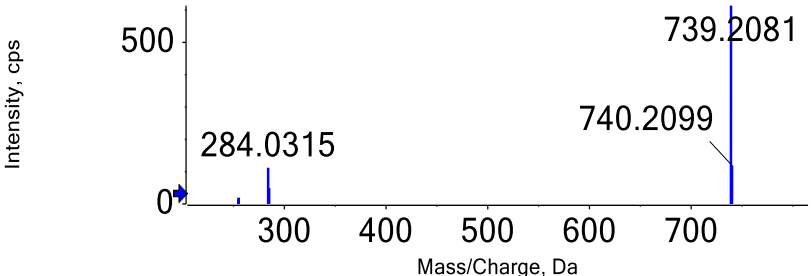
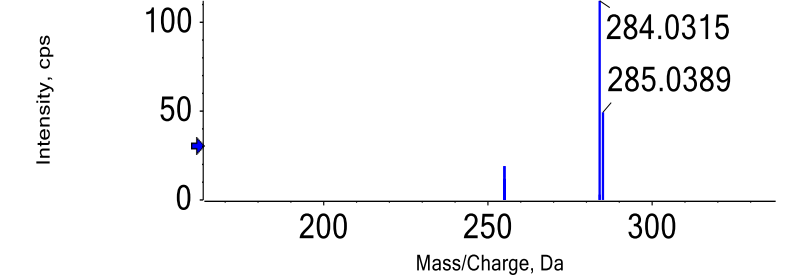
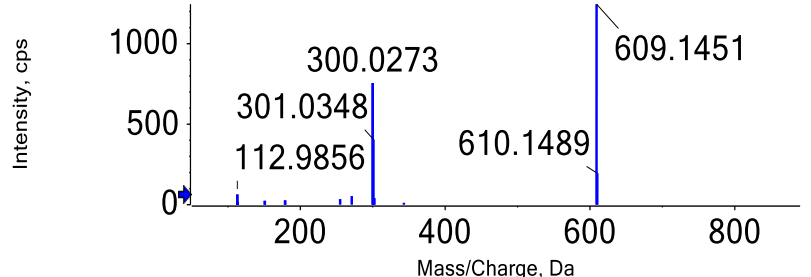
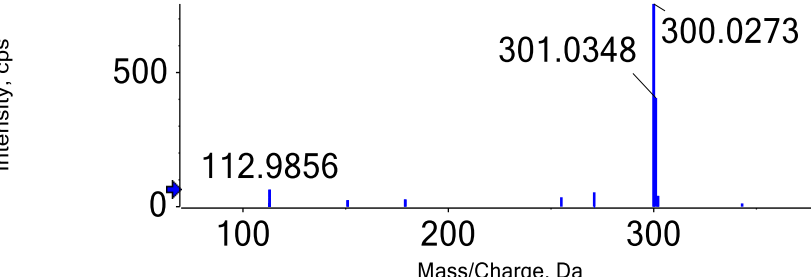
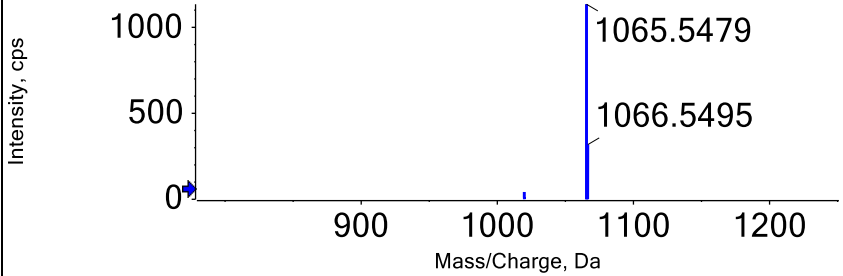
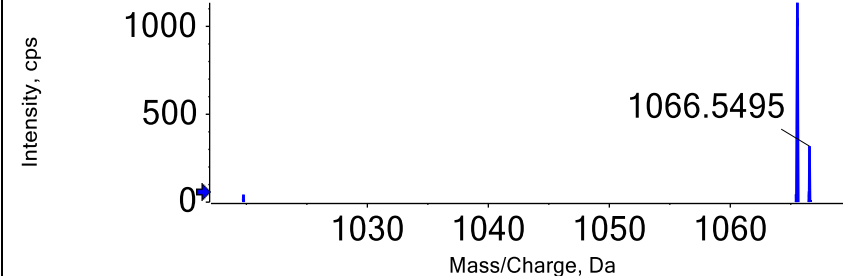
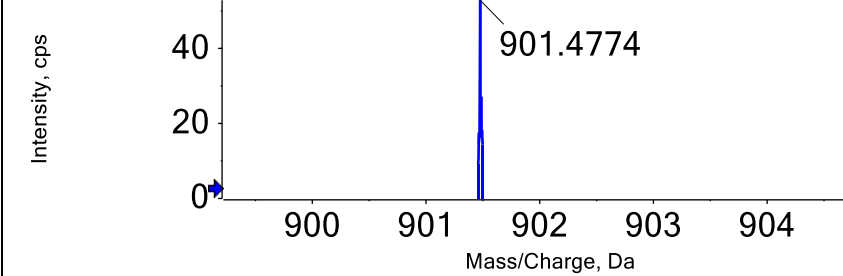
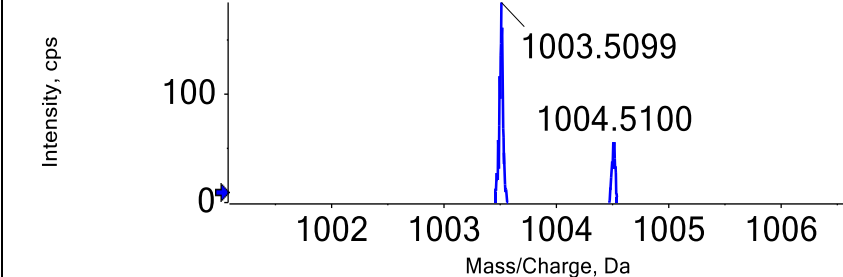
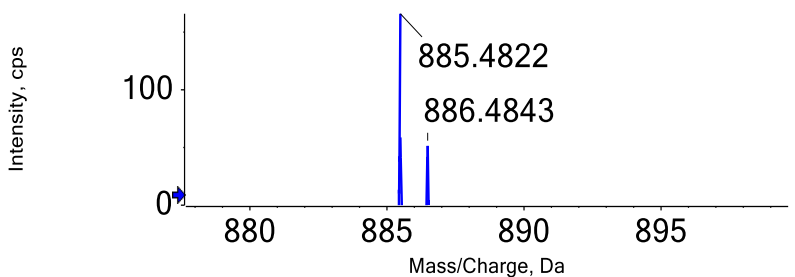
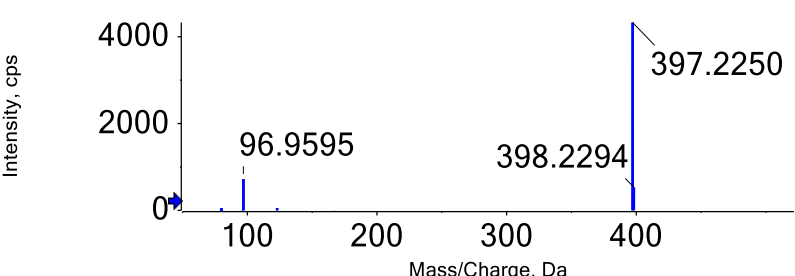
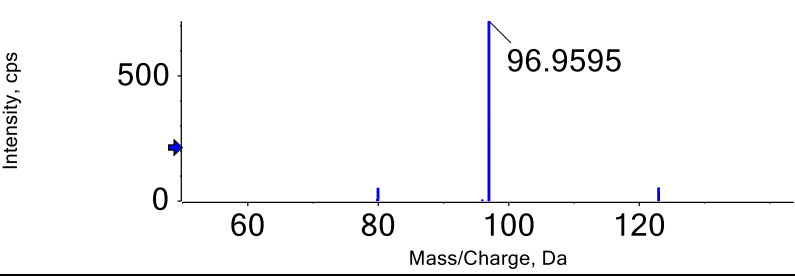
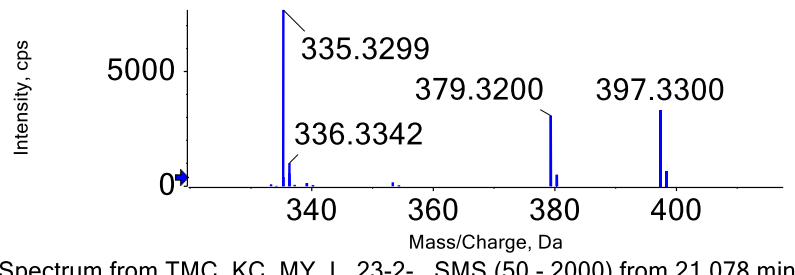
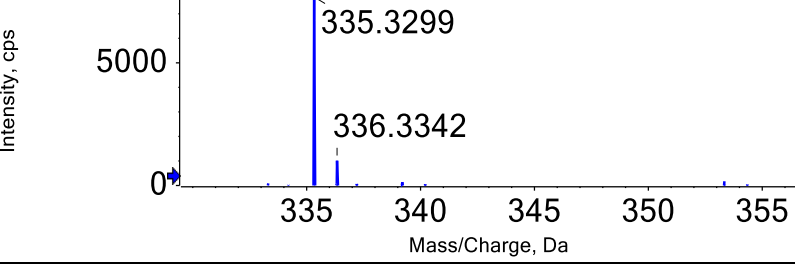
18.93	397.2275 β -sitosterol	Spectrum from TMC_KC_MY_L_23-2-22_NEG.wiff2 (sample... multiplier = 1.5), Gaussian smoothed (0.5 points)  Intensity, cps Mass/Charge, Da
21.08	397.3324 Daucosterol	Spectrum from TMC_KC_MY_L_23-2-22_NEG.wiff2 (sample... multiplier = 1.5), Gaussian smoothed (0.5 points)  Intensity, cps Mass/Charge, Da

Table 1(B). HPLC-ESI-QTOF NEG assignments of compounds in *A. racemosus* leaves.

RT	M/z and compound	MS/MS spectrum
7.32	477.1616 Quercetin-3-glucuronide	Spectrum from TMC_KC_MY_L_23-2-...SMS (50 - 2000) from 7.317 min Precursor: 477.2 Da, +1, noise filte...1.5), Gaussian smoothed (0.5 points)  Intensity, cps Mass/Charge, Da Spectrum from TMC_KC_MY_L_23-2-...SMS (50 - 2000) from 7.317 min Precursor: 477.2 Da, +1, noise filte...1.5), Gaussian smoothed (0.5 points)  Intensity, cps Mass/Charge, Da

8.27	739.2095 Asparanin A	Spectrum from TMC_KC_MY_L_23-2-...SMS (50 - 2000) from 8.276 min Precursor: 739.2 Da, +1, noise filte...1.5), Gaussian smoothed (0.5 points)  Spectrum from TMC_KC_MY_L_23-2-...SMS (50 - 2000) from 8.276 min Precursor: 739.2 Da, +1, noise filte...1.5), Gaussian smoothed (0.5 points) 
8.47	609.1457 Rutin	Spectrum from TMC_KC_MY_L_23-2-...SMS (50 - 2000) from 8.463 min Precursor: 609.1 Da, +1, noise filte...1.5), Gaussian smoothed (0.5 points)  Spectrum from TMC_KC_MY_L_23-2-...SMS (50 - 2000) from 8.463 min Precursor: 609.1 Da, +1, noise filte...1.5), Gaussian smoothed (0.5 points) 

9.75	1065.5461 Shatavarin I	Spectrum from TMC_KC_MY_L_23-2-...SMS (50 - 2000) from 9.751 min Precursor: 1065.5 Da, +1, noise filt...1.5), Gaussian smoothed (0.5 points)  Intensity, cps Mass/Charge, Da Spectrum from TMC_KC_MY_L_23-2-...SMS (50 - 2000) from 9.751 min Precursor: 1065.5 Da, +1, noise filt...1.5), Gaussian smoothed (0.5 points)  Intensity, cps Mass/Charge, Da
11.14	901.4792 Shatavarin IX	Spectrum from TMC_KC_MY_L_23-2-...SMS (50 - 2000) from 11.163 min Precursor: 901.5 Da, +1, noise filt...1.5), Gaussian smoothed (0.5 points)  Intensity, cps Mass/Charge, Da
12.79	1003.5094 Asparacotide	Spectrum from TMC_KC_MY_L_23-2-...SMS (50 - 2000) from 12.819 min Precursor: 1003.5 Da, +1, noise filt...1.5), Gaussian smoothed (0.5 points)  Intensity, cps Mass/Charge, Da

14.06	885.4821 Shatavarin IV	Spectrum from TMC_KC_MY_L_23-2-...SMS (50 - 2000) from 14.075 min Precursor: 885.5 Da, +1, noise filte...1.5), Gaussian smoothed (0.5 points)  Intensity, cps Mass/Charge, Da
18.93	397.2275 β -sitosterol	Spectrum from TMC_KC_MY_L_23-2-...SMS (50 - 2000) from 18.928 min Precursor: 397.2 Da, +1, noise filte...1.5), Gaussian smoothed (0.5 points)  Intensity, cps Mass/Charge, Da Spectrum from TMC_KC_MY_L_23-2-...SMS (50 - 2000) from 18.928 min Precursor: 397.2 Da, +1, noise filte...1.5), Gaussian smoothed (0.5 points)  Intensity, cps Mass/Charge, Da
21.08	397.3324 Daucosterol	Spectrum from TMC_KC_MY_L_23-2-...SMS (50 - 2000) from 21.078 min Precursor: 397.3 Da, +1, noise filte...1.5), Gaussian smoothed (0.5 points)  Intensity, cps Mass/Charge, Da Spectrum from TMC_KC_MY_L_23-2-...SMS (50 - 2000) from 21.078 min Precursor: 397.3 Da, +1, noise filte...1.5), Gaussian smoothed (0.5 points)  Intensity, cps Mass/Charge, Da

The total ion chromatograms (TICs) of *A. racemosus* leaves' water extract in HPLC-ESI-Q-TOF-MS negative mode were shown in Fig. 1.

Some parent and fragment ions were observed in the mass spectra and summarized in Table 1(A,B).

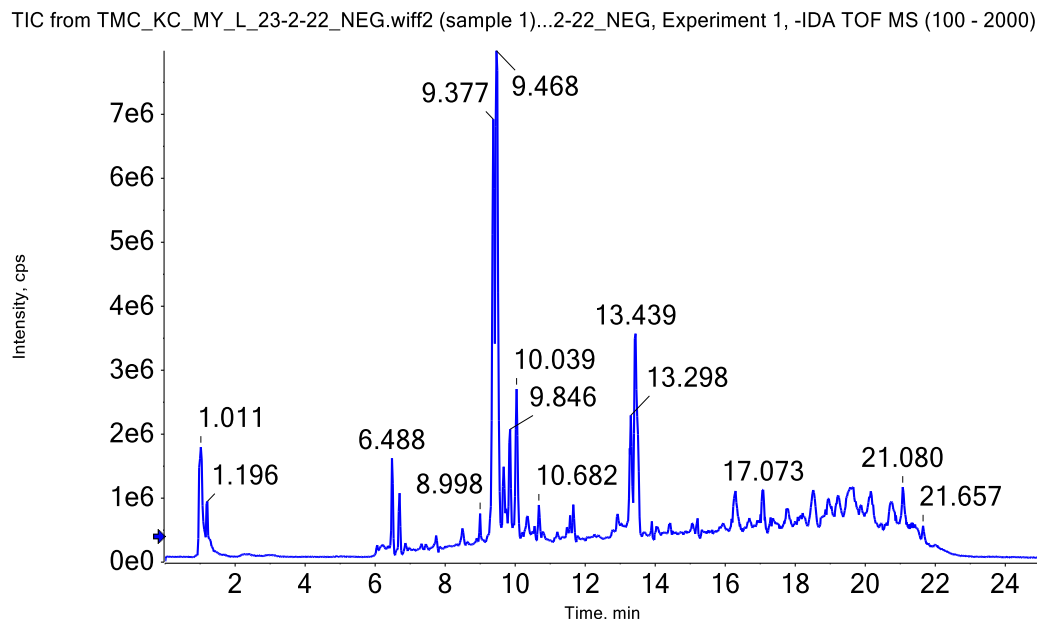


Fig. 1. TIC of HPLC-ESI-Q-TOF-MS NEG mode of water extract of wild *A. racemosus* leaves.

Compound (1), at $T_R = 7.32$, in the ESI-QTOF negative mode, yielded a parent ion $[M-H]^-$ at m/z 477.1616, and provided fragment ions at m/z 59.0136, m/z 71.0121, m/z 89.0238, m/z 168.0412, and m/z 233.0669. Xue et al. (2021, 2022) found the same compound in the leaves of *A. racemosus* and determined its structure as Quercetin-3-glucuronide [9, 24]. Thus, compound (1) was tentatively defined as Quercetin-3-glucuronide. Compound (2), at $T_R = 8.27$, in the ESI-QTOF negative mode, yielded a parent ion $[M-H]^-$ at m/z 739.2095, and provided major fragment ions at m/z 284.0315. Its structure is determined as Asparanin A [25]. At $T_R = 8.47$, in the ESI-QTOF negative mode, compound (3) yielded a parent ion at m/z 609.1457 and provided two fragment ions at m/z 112.9856 and m/z 301.0348. Saxena et al. (2001) found the same compound in *A. racemosus* and determined its structure as Rutin [26]. The MS spectra of compound (4) ($T_R = 9.75$ min), yielded a parent ion $[M-H]^-$ at m/z 1065.5461 in the negative mode, without any fragmented ions. Compound (4) was determined as Shatavarin I [27].

Compound 5 ($T_R = 11.14$ min) exhibited a parent ion $[M-H]^-$ at m/z 901.4792 without any fragmented ions in the ESI-QTOF negative mode. It was assigned to Shatavarin IX [27, 28]. At $T_R = 12.79$, in the ESI-QTOF negative mode, compound (6) yielded a parent ion at m/z 1003.5094, without any fragmented ions. Onlom et al. (2017) found the same compound in *A. racemosus* and determined its structure as Asparacoside [28]. Compound (7), at $T_R = 14.06$, in the ESI-QTOF negative mode, yielded a parent ion $[M-H]^-$ at m/z 885.4821, without any fragmented ions. Its structure is determined as Shatavarin IV [28, 29]. At $T_R = 18.93$, in the

ESI-QTOF negative mode, compound (8) yielded a parent ion at m/z 397.2275 and provided one fragment ion at m/z 96.9595. Ariful (2014) found the same compound in *A. racemosus* and determined its structure as β -sitosterol [30]. The MS spectra of compound (9) ($T_R = 21.08$ min), yielded a parent ion $[M-H]^-$ at m/z 397.3324 in the negative mode, and provided two fragment ions at m/z 335.3299, m/z 336.3342, and m/z 378.3200. Compound (10) was determined as Daucosterol [30].

In this study, ultrahigh resolution liquid chromatography (UHPLC) was used to effectively separate the phytoconstituents. High resolution mass spectroscopy (Q-TOF-MS/MS) detector was used to determine the parent ion and its fragments which can be compared with databases and literatures for structure of phytoconstituents without the use of their expensive standard compounds. With the use of standard compounds, respective constituents could be quantified by less expensive UHPLC-DAD equipment.

3. 2. ATR-FTIR spectra of *A. racemosus* leaves

Chemical compositions of forty powdering samples of *A. racemosus* leaves were investigated by ATR- FTIR spectroscopy. After IR analysis, spectral band positions were identified, and then assigned to the major functional groups by comparing with the literature. ATR- FTIR spectra of the samples are shown in Fig. 2.

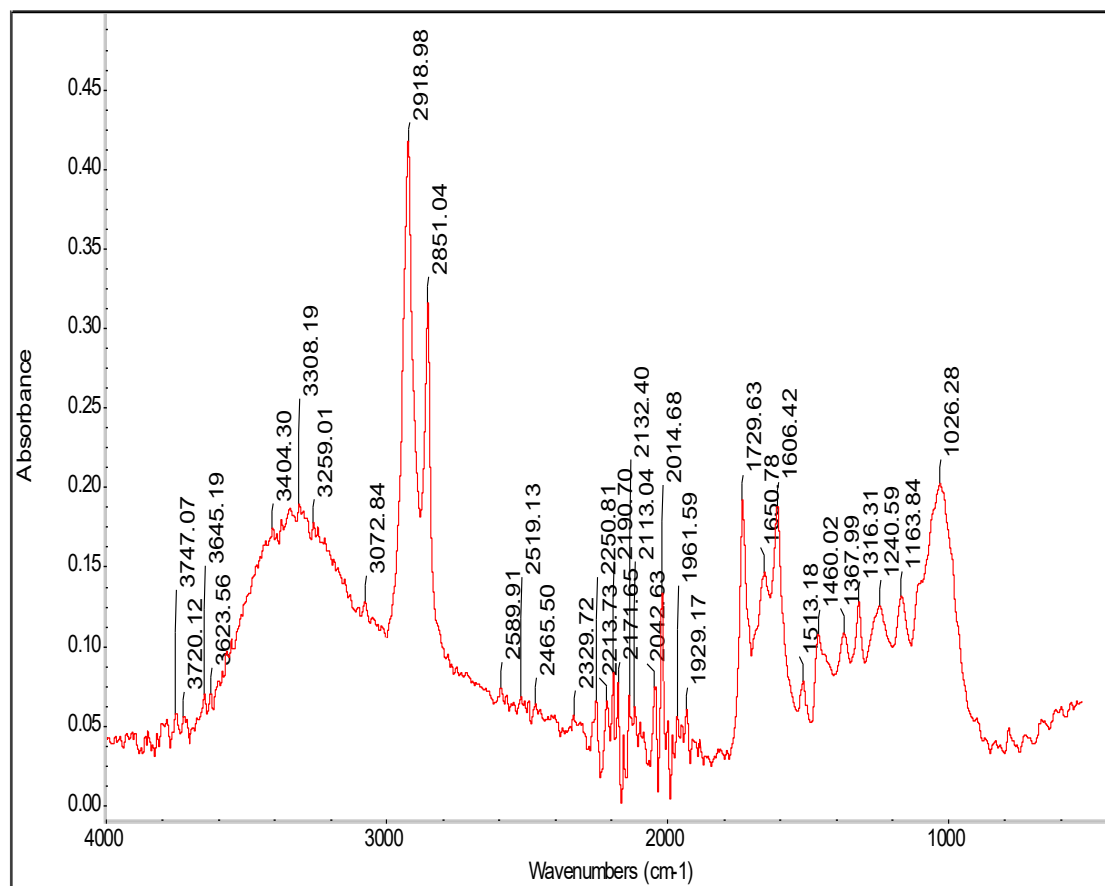


Fig. 2. ATR-FTIR spectra of wild *A. racemosus* leaves' powder sample
TMC_Hatam_DP_GL_Duoitan_L_BT_27-4-23

Thirty two ATR-FTIR spectra of *A. racemosus* leaves were measured and the peaks were assigned as follows: The spectra showed the O-H (H-bonded) stretching of the phenolic compounds such as tannins and flavonoids at 3404.30 cm^{-1} [13], and other O-H groups of the alcohol and phenol compounds at both vibrations of 2519.13 cm^{-1} and 2465.50 cm^{-1} [15].

The C-H stretching vibrations of the methyl/ methylene groups and alkane/ aldehyde compounds were assigned by both wavenumbers at 2918.95 cm^{-1} and 2851.04 cm^{-1} , respectively [15]. At lower vibrations of 1460.02 cm^{-1} and 1367.99 cm^{-1} , the C-H bending modes presented the lignin / carboxylic acid compounds [13] and alkane compounds [15], respectively. The double bonds stretching vibrations of the allene compounds' -C=C functional groups and the amide I constituents' -C=O functional group were presented at 1929.17 cm^{-1} and 1650.78 cm^{-1} , respectively [15]. The triple bonds stretching vibrations of -C $\equiv$ C functional groups of the alkyne constituents were presented in both wavenumbers at 2213.76 cm^{-1} and 2132.40 cm^{-1} [15]. The absorption band of polysaccharides was observed at 1163.94 cm^{-1} [13]. The ATR-FTIR's C-OH bending and C-OH stretching of constituents of alcohol, ethers, anhydrides, and carboxylic acids compounds were shown at 1026.29 cm^{-1} vibration [15].

3.3. Principal Components Analysis model

The labeled spectral data from the Excel data files were processed. The entire dataset, consisting of forty ATR-FTIR spectra measured at wavelengths from 400 cm^{-1} to 4000 cm^{-1} , forms a matrix of dimensions 64×902. As the measured signals from the leaf samples are relatively low and exhibit numerous noisy areas, especially in the range of 1900-2300 cm^{-1} .

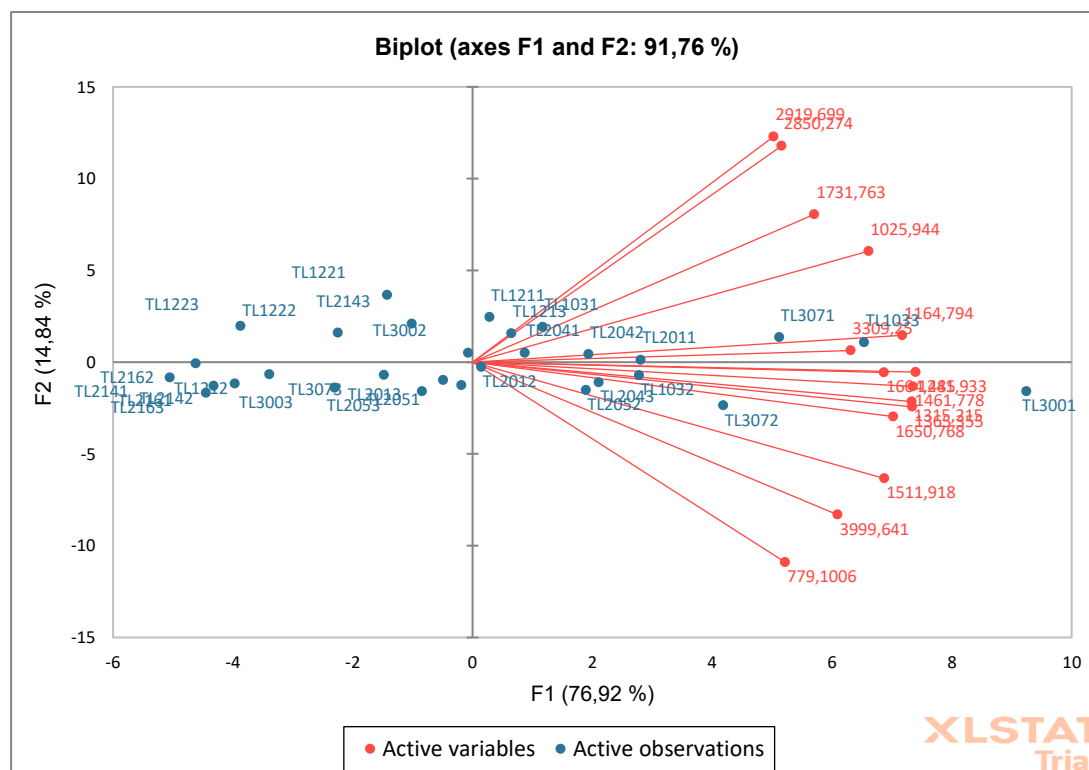


Fig. 3. Two PCs scores (F1 and F2) by PCA classification of ATR-FTIR spectra of wild *A. racemosus* leaves' powder samples.

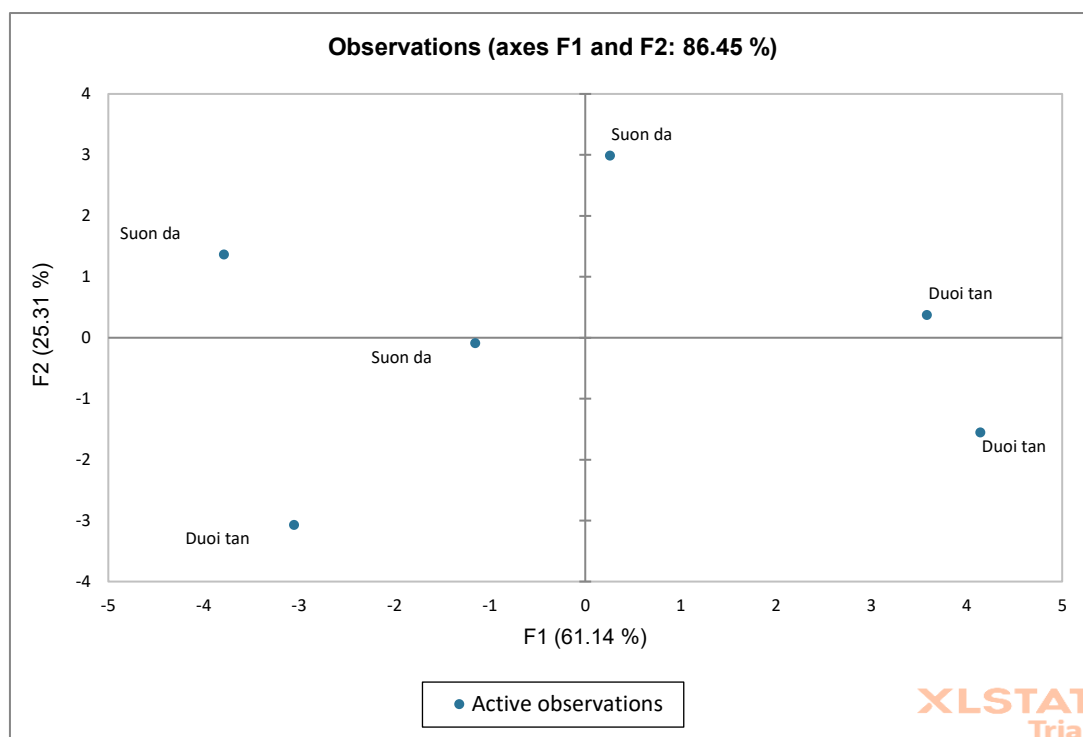


Fig. 4. Two PCs scores (F1 and F2) by PCA classification of ATR-FTIR spectra of wild *A. racemosus* leaves' powder samples collected differently from “rocky places” (left) and “under the canopy” (right).

Therefore, to facilitate easier evaluation, it is necessary to perform spectral derivation to amplify the features and then smooth the spectra to reduce noise. After applying differentiation and smoothing using the Savitzky-Golay algorithm, significant reduction in signal overlaps among the samples. This is particularly evident with second-order differentiation, which reveals distinct signal differences in the 1350-1850 cm^{-1} region. Previously, this region exhibited overlapping spectral signals, despite some initial signs of differentiation.

Although noise values in the 1900-2300 cm^{-1} region persist, the classification differentiation has improved, allowing this issue to be disregarded. To conveniently explore linear separability, PCA was applied to reduce the dataset's dimensionality. The PCA results indicate that using 2 PCs scores is most promising, so these reduced dataset scores are employed to construct the model. Test data points are classified based on their domains by two PCs scores (F1 and F2): blue points left of the boundary are labelled as the leaves collected mainly from “rocky places” (synonym in Vietnamese as “suon da”), while those on the right are the leaves planted mainly “under the canopy” (synonym in Vietnamese as “duoi tan”) as illustrated in Figs 3 and 4. This underscores PCA's suitability for classifying leaves planted in different places using ATR-FTIR spectral signals.

In conclusion, from the water extract of *A. racemosus* leaves, we have tentitatively identified two flavonoids (Quercetin-3-glucuronide, Rutin), five steroidal saponins (Shatavarin I, Shatavarin IV, Shatavarin IX, Asparacocide, Asparanin A), and two steroids (β -sitosterol, Daucosterol) in the leaves of *A. racemosus* by UHPLC-Q-TOF-MS spectroscopy. ATR-FTIR

spectroscopy revealed the presence of various functional groups by of *Asparagus racemosus*'s leaves viz. phenol, alkyl group, methyl groups, alcohols, ethers, carboxylic acid which may present due to several biomolecules; flavonoids, alcohols, phenols, alkanes, carboxylic acids, alkenes, aromatics, allenes, amides, lignin, cellulose, hemicelluloses and polysaccharides bearing these groups. With clearly identified collected places of the *Asparagus racemosus*'s leaves, the use of ATR-FTIR spectroscopy combined with supervised learning algorithms (PCA) recognized the chemical characteristic differences of all powdered samples from different places. This method shows the potential to increase the number of samples and apply algorithms and data processing methods for the classification of other medicinal samples, as well as aiming at the classification of geographic origin, species, and harvest season.

4. CONCLUSION

This study showed the chemical investigation of *A. racemosus*'s leaves analyzed by UHPLC-Q-TOF-MS spectroscopy with identification of nine known compounds. And, quantitative analysis of the assigned compounds might be carried out to assess quality of *A. racemosus* species. ATR-FTIR spectroscopy has been found extremely advantageous and easy to identify the IR active phytoconstituents of *Asparagus racemosus*'s leaves in powder forms. By using ATR-FTIR spectra data and the second derivative of these spectra, the chemical characteristic differences between the leaves planted in different places were classified. It might be of interest to further investigate the *A. racemosus* species on isolation of its chemical constituents and pharmaceutical activities.

ACKNOWLEDGEMENTS

The project was supported by the grant KHGL-01-22.

References

- [1] Goyal RK, Singh J, Lal H. *Asparagus racemosus* - an update. *Indian J Med Sci.* 2003; 57: 408-414
- [2] Nguyen Van Vu, Nguyen Danh, Tran Minh Đức. Morphology and molecular structural analysis of *Asparagus racemosus* wild. from Gia Lai province, Vietnam. *Hue University Journal of Science Agriculture and Rural Development.* 2019,128(3A): 83. DOI:10.26459/hueuni-jard.v128i3A.5050
- [3] Guo Y, Liu Z, Wan Y, Zhang Y, Abdu HI, Yang M, Pei J, Yue T, Zhang X, Hacimuftuoglu A, Abd El-Aty AM. Literature analysis on *Asparagus* roots and review of its functional characterizations. *Front Nutr.* 2023, 9: 1024190. doi: 10.3389/fnut.2022.1024190
- [4] Zhang H, Birch J, Pei J, Ma ZF, Bekhit AED. Phytochemical compounds and biological activity in *Asparagus* roots: a review. *Int J Food Sci Technol.* 2019, 54: 966-977. doi: 10.1111/ijfs.13993

- [5] Huong NT, Thuy DN, van Trung P, Hung LN, van Nam M. DNA Genetics and UHPLC-Q-TOF-MS Analysis of Phytochemicals for *Asparagus racemosus* Roots. *IgMin Res.* 2024; 2(7): 558-563. DOI: 10.61927/igmin217
- [6] Sumitra Nain, Ruchi Singh, Samriti Faujdar and Sarvesh Paliwal. Phytochemical Evaluation of *Asparagus racemosus* and *Chlorophytum borivilianum* Leaves. *Der Pharma Chemica* 2022; 14(1): 10-21
- [7] Battu, G. S. and B.V.V. Ravi Kumar. Phytochemical and antimicrobial activity of leaf extract of *Asparagus racemosus* Willd. *Pharmacognosy Journal* 2 (2010) 456-463
- [8] Khan KM, Nahar L, Mannan A, Arfan M, Khan GA, Al-Groshi A, Evans A, Dempster NM, Ismail FMD, Sarker SD. Liquid Chromatography Mass Spectrometry Analysis and Cytotoxicity of *Asparagus adscendens* Roots against Human Cancer Cell Lines. *Pharmacogn Mag.* 2018, 13(Suppl 4): S890-S894. doi: 10.4103/pm.pm_136_17
- [9] Xue X, Jin R, Jiao Q, Li X, Li P, Shen G, Shi S, Huang Z, Dai Y, Zhang S (2022). Differentiation of three *Asparagus* species by UHPLC-MS/MS based molecular networking identification and chemical profile analysis. *J Pharm Biomed Anal.* 219: 114863. doi: 10.1016/j.jpba.2022.114863
- [10] Gyoung S. Na, Yecheol Rho. Explainable Machine Learning for Characterizing Unknown Molecular Structures in Infrared Spectra. *Analytical Chemistry* 2025, 97(38), 20869-20878. <https://doi.org/10.1021/acs.analchem.5c03126>
- [11] Ana M. Jiménez-Carvelo, María Teresa Osorio, Anastasios Koidis, Antonio González-Casado, Luis Cuadros-Rodríguez. Chemometric classification and quantification of olive oil in blends with any edible vegetable oils using FTIR-ATR and Raman spectroscopy. *LWT*, 86 (2017) 174-184, <https://doi.org/10.1016/j.lwt.2017.07.050>
- [12] Pei Y, Zhang Q, Zuo ZT, Wang Y (2018). Comparison and Identification for Rhizomes and Leaves of *Paris yunnanensis* Based on Fourier Transform Mid-Infrared Spectroscopy Combined with Chemometrics. *Molecules*, 23(12): 3343-3343. Doi: 10.3390/molecules23123343
- [13] Mishra, A.K., Mishra, V.K., Dwivedi, D., Dwivedi, K.N. (2015). A FT-IR spectroscopic study of phytoconstituents of *Asparagus racemosus* willd root tuber. *World Journal of Pharmaceutical Research*, 4(8), 2754-2763
- [14] Jyothi, A., & Satyavati, D. (2016). FT-IR studies of ethanolic extract of *Asparagus racemosus*. *WJPPS*, 5(4), 1388-1393
- [15] Lalit Samant (2022). Evaluation of Phytochemicals Profile and Antioxidant Activities of *Asparagus racemosus* using in vitro and in silico approach. *Der Pharma Chemica*, 14(8): 01-09
- [16] Mario Koppen (2000). The Curse of Dimensionality. (held on the internet), September 4-18 2000. 5th Online World Conference on Soft Computing in Industrial Applications (WSC5)
- [17] Jolliffe I.T (2002). Principal Component Analysis, 2nd edition. Springer Nature. <https://doi.org/10.1007/b98835>

- [18] Medin DL, Schwanenflugel PJ (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning & Memory*, 7(5): 355-368. Doi: 10.1037/0278-7393.7.5.355
- [19] Suthaharan, S. (2016). Support Vector Machine. In: Machine Learning Models and Algorithms for Big Data Classification. Integrated Series in Information Systems, vol 36. Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7641-3_9
- [20] Zhang YC, Deng J, Lin XL, Li YM, Sheng HX, Xia BH, Lin LM (2022). Use of ATR-FTIR Spectroscopy and Chemometrics for the Variation of Active Components in Different Harvesting Periods of *Lonicera japonica*. *Int J Anal Chem.* 2022, 2022: 8850914. doi: 10.1155/2022/8850914
- [21] Rodel E. Quero, Kayla Lucas, Jessica Higgins, Elmer-Rico E. Mojica. ATR-FTIR characterization and multivariate analysis classification of different commercial propolis extracts. *Measurement: Food*, 18 (2025) 100224, <https://doi.org/10.1016/j.meafao.2025.100224>
- [22] Hasan, M., & Panda, B.P. (2020). Chemometric analysis of selective polyphenolic groups in *Asparagus racemosus* (Shatavar) root extracts by traditional and supercritical fluid (CO₂) based extractions. *Separation Science and Technology*, 55, 1339-1355
- [23] Pham HN, Tran CA, Trinh TD, Nguyen Thi NL, Tran Phan HN, Le VN, Le NH, Phung VT. (2022). UHPLC-Q-TOF-MS/MS Dereplication to Identify Chemical Constituents of *Hedera helix* Leaves in Vietnam. *J Anal Methods Chem.* 2022, 2022: 1167265. doi: 10.1155/2022/1167265
- [24] Xue, X., Jiao, Q., Jin, R. et al. The combination of UHPLC-HRMS and molecular networking improving discovery efficiency of chemical components in Chinese Classical Formula. *Chin Med.* 2021, 16, 50. doi: 10.1186/s13020-021-00459-6
- [25] Ahmad Viqar Uddin and Basha Anwer. Asparanin A - Spectroscopic Data of Steroid Glycosides: Stigmastanes, Furostanes, Spirtostanes: Volume 2. Springer US. 2006; pages 1286-1287. ISBN 978-0-387-39572-2. DOI 10.1007/978-0-387-39572-2_570
- [26] Saxena VK, Chaurasia S. A new isoflavone from the roots of *Asparagus racemosus*. *Fitoterapia*, 2001; 72: 307-309
- [27] Hayes, P. Y., Jahidin, A. H., Lehmann, R., Penman, K., Kitching, W., and De Voss, J. J. Structural revision of shatavarins I and IV, the major components from the roots of *Asparagus racemosus*. *Tetrahedron Letters* 2006, 47 (39), 6965-6969 <https://doi.org/10.1016/j.tetlet.2006.07.121>
- [28] Onlom C, Nuengchamnong N, Phrompittayarat W, Putalun W, Waranuch N, Ingkaninan K (2017). Quantification of Saponins in *Asparagus racemosus* by HPLC-Q-TOF-MS/MS. *Nat Prod Commun.* 12(1): 7-10
- [29] Saran PL, Singh S, Solanki VH, Devi G, Kansara RV, Manivel P. Identification of potential accessions of *Asparagus racemosus* for root yield and shatavarin IV content. *Heliyon* 2020, 6(12): e05674. doi: 10.1016/j.heliyon.2020.e05674
- [30] Ariful Hoque (2014). Phytochemical and Biological Investigation on the Roots of *Asparagus rasemosus*. Bangladesh University of Engineering and Technology