

Gas Chromatography-Mass Spectrometry (GC-MS) and ¹H Nuclear Magnetic Resonance (NMR)-based profiling and phytochemical screening revealed industrially important metabolites in wild seabuckthorn (*Hippophae salicifolia* D.Don) berries

Sugandh Singh

Guru Gobind Singh Indraprastha University

Prakash Chand Sharma (✉ prof.pcsharma@gmail.com)

Guru Gobind Singh Indraprastha University <https://orcid.org/0000-0002-7939-4780>

Research Article

Keywords: GC-MS, ¹H NMR, *Hippophae salicifolia*, Metabolome profiling, OPLS-DA, PLS-DA, Seabuckthorn

Posted Date: March 14th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-2659503/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

In recent times, seabuckthorn (*Hippophae* Sp.) has attracted worldwide attention as a source of traditional medicines, nutraceuticals, and cosmeceuticals due to the presence of a wide range of bioactive phytochemicals. These bioactive phytochemicals may vary in different regions of plant collections. The amount of plant materials traded within the State has increased dramatically in recent years as a result of the rising demand for herbal products. Moreover, there are no organized methods of seabuckthorn cultivation in Uttarakhand. Through experience and practice, the locals preserve their customary knowledge, which is passed down orally and without documentation. Currently, this information is in danger of disappearing. The scenario regarding the development of seabuckthorn as a future crop in the Uttarakhand region is very alarming. Most of the metabolomics research has been carried out on a single analytical platform. However, each platform has intrinsic advantages and drawbacks that can affect the analytical coverage of the metabolome. The present study was carried out to identify the entire unrevealed metabolites in *H. salicifolia* berries originating from Uttarakhand regions for the first time, utilizing both GC-MS and ^1H NMR techniques. Phytochemical screening was performed by various qualitative and quantitative methods such as GC-MS and ^1H NMR in *H. salicifolia* berry extract originating from three regions (Badrinath, Yamunotri, and Gangotri) of Uttarakhand (India). Phytochemical determination revealed the presence of industrially important metabolites. In total, 318 metabolites were identified, 305 by GC-MS and 13 by ^1H NMR. These metabolites were classified in 23 sets of metabolites such as alkane hydrocarbons, alkene hydrocarbons, fatty acid methyl esters, etc. in which total metabolomic percentage and its amount was highly detected in the Gangotri region followed by Badrinath and Yamunotri. The present study provides valuable information for the identification of the natural source of metabolites related to the quality of various seabuckthorn-based products useful for applications in the food and pharmaceutical industries as well as for effective germplasm collection.

1. Introduction

The genus *Hippophae* (family Elaeagnaceae) is represented by five species namely, *H. rhamnoides* L., *H. salicifolia* D. Don, *H. neurocarpa* Liu & He, *H. tibetana* Schlecht, and *H. goniocarpa* Lianex Swenson & Bartish (Aras et al. 2007). In India, three species of *Hippophae* are found, including *H. rhamnoides*, *H. salicifolia*, and *H. tibetana*, which grow naturally at high-altitude (2391–4300 m) in the Himalayas (Raina et al. 2011). Among them, seabuckthorn (*Hippophae salicifolia*) is marked as an abundant natural source of metabolites. The high phytochemical content present in the berries and leaves of *H. salicifolia* accounts for its tremendous nutraceutical properties (Manu et al. 2014). In the traditional medical system, berries are commonly used to treat some internal and external ailments such as asthma, skin diseases, lung disorders, etc (Ranjith et al. 2006). Currently, more than 150 nutraceutical and pharmaceutical firms across the world, primarily from China, are using seabuckthorn extracts, particularly from *H. rhamnoides* to make medications, health tonics, food, cosmetics, and other products (Dhyani et al. 2012). Since the beginning of modern civilization, people have used various seabuckthorn plant parts to treat a variety of illnesses. Today, many pharmaceutical medicines are prepared using plant extracts.

There are numerous people practicing folk medicine, in the village who only use plant parts. They used conventional techniques to prepare the medicines from different plant parts. Some bioactive substances or phytochemicals are responsible for the plant's therapeutic effectiveness (Verpoorte 1999). Approximately 95% of the prescribed medications on our Indian subcontinent were founded on the conventional Unani, Homeopathy, and Ayurvedic medical systems (Satyawati and Gupta 1987). The phytochemicals produced in the plant may be a source for production of a crude drug by different pharmaceutical industries. The discovery of compounds that are both medicinally and commercially significant depends on phytochemical screening. In India, *H. salicifolia* grows widely in the Garhwal and Kumaon divisions of Uttarakhand (Yamunotri, Gangotri, and Badrinath regions). In Uttarakhand, the species is locally known as *Brahmaphal*, the fruit of God. Even though seabuckthorn berries from these regions have long been recognized as having significant medicinal and therapeutic potential, no systematized metabolome analysis has been conducted so far. Considering the health advantages of seabuckthorn berry bioactives and for conservation purpose, a dataset is needed to document phytochemical variation in different regions of Uttarakhand, which can vary depending on elevation, temperature, water availability, light intensity, etc. The superior quality of metabolites present in *H. salicifolia* has been established in some studies over other seabuckthorn species. Commercially and ecologically, the plant has been unattended despite all its beneficial disease-curing properties. From earlier times, the other *Hippophae* species of the lower nutrient content of metabolites have extensively been used by the pharmaceutical, nutraceutical, and cosmetic industries in making diverse products. Furthermore, a lack of recognition of the importance of *H. salicifolia* among growers is resulting in the gradual extinction of wild populations in some areas. Metabolomics, a robust top-down system biology approach, can simultaneously measure global metabolic composition in any biosystem (Nicholson et al. 1999). Analyzing and quantifying huge metabolites in a non-hypothetical way is the basic aim of metabolomics research applied to the discovery of biomarkers (Fancy et al. 2008). Comprehensive metabolomic information acquired using distinct analytical techniques allows the characterization of variability prevailing within a species. Different spectroscopic techniques have been exploited for plant metabolome analysis (Saito and Matsuda 2010). Of these, GC-MS and ¹H NMR are the most popular technique to date that can effectively identify the whole metabolome in a single run. Natural product chemistry has frequently used techniques using both NMR and MS data to develop potential compounds but the use of this combinatorial strategy in metabolomics has been limited in metabolomics (Forseth et al. 2011). GC-MS is a prominent and efficient analytical technique for the sensitive identification of a wide range of metabolic classes due to substantial earlier analytical method advancement. It is widely used to identify volatile metabolites. Derivatization of samples is required for the detection of nonpolar metabolites using GC-MS. NMR is another popular analytical method for metabolite profiling because it is essentially quantitative, detects organic substances universally, and easily displays structural information depending on the chemical shifts and coupling patterns. Even without connecting to a separation technique, NMR can typically resolve the component resonances of mixtures of intermediate complexity when a high-field magnet is implemented (Holzgrabe et al. 1998). Publicly available libraries of metabolites and their corresponding NMR and GC-MS spectra offer researchers resources for confident resonance assignments and inter/intra-laboratory transferability (Cui et al. 2008; Wishart et al. 2009).

Some research efforts have aimed in the past towards understanding the interrelationship between plant metabolome with fruit growth, ripening, storage, and geographical origin employing multivariate analysis as a key tool for determining the nutraceutical quality of the products (Oms-Oliu et al. 2013).

The objective of this study was to evaluate the metabolome diversity of seabuckthorn (*H. salicifolia*) berries collected from three major regions in Uttarakhand for industrially important metabolites. For this purpose, the GC-MS and ^1H NMR techniques were used in combination with applicable chemometric analysis tools. We hope that the findings of the study would be helpful for the cosmeceutical, nutraceutical, and pharmaceutical industries, and in selecting natural resources useful in seabuckthorn conservation and breeding programs.

2. Materials And Methods

2.1. Seabuckthorn berry collection

In total, 30 seabuckthorn (*H. salicifolia*) berry samples were collected from three geographical regions in Uttarakhand, India, namely Yamunotri, Gangotri, and Badrinath. The ripened berries were collected randomly from several plants in the first week of October 2019. The ripened berry-bearing branches were sliced from the tip with a 'plant cutter' at a distance of 10 cm. The berries were chosen with care to ensure uniformity in size and maturity. Of the 30 berry samples, 10 samples were collected from Badrinath, 6 samples from Yamunotri, and 14 samples from Gangotri sites. A Global Positioning System (GPS) device (Garmin) was used for tagging distinct geographical collection sites. The berries were harvested and placed in zip-lock polybags, which were then placed in ice packs until transported to the lab. We next cleaned and preserved the samples in an ultra-deep freezer at -80°C until used for processing. The information on the collection sites is presented in Supplementary Table 1.

2.2. Sample preparation for GC-MS

The sample preparation methodology was largely based on an earlier published method (Datta et al. 2012). Berry seeds were extracted using forceps before they were crushed in liquid nitrogen for metabolite extraction. After homogenization, the pulp powder was lyophilized for 24 hours and then preserved at -80°C till further processing. 200 mg of homogenized pulp powder was mixed with 7500 μl of chloroform: methanol (CHCl_3 : CH_3OH) (1:2), followed by rapid vortexing. The mixture was vortexed again after adding 2500 μl of CHCl_3 . The bottom organic phase was transferred to a new vial, and the solvent was evaporated completely at 35°C , separating the mixture into two layers. The dried residue was treated with 1000 μl of 6% methanolic potassium hydroxide (methanolic KOH) (w/v) and incubated for 1 hour at 37°C . Half volume, i.e. 500 μl of dH_2O , was added to the mixture, followed by the addition of proportionate volume, i.e. 1500 μl of n-heptane, and vortexed well. The mixture was left to stand for a few minutes to allow the separation of the two layers. The upper layer was transferred to a new vial. The process was repeated once. The residual solvent was allowed to evaporate completely by keeping tubes open for 24 hours at room temperature. Later, using a one-step process, 200 μl of derivatization reagent mixture [160

μl N, O-bistrifluoroacetamide (BFSTA) + 40 μl Trimethylchlorosilane (TMCS)] was added and incubated for 1 h at 37°C to extract all the polar metabolites. All metabolites were extracted at the optimal temperature, including thermolabiles, which could be damaged if extracted at a higher temperature (Egon 2005). The resulting derivatized extract was subjected to GC-MS analysis.

2.2.1. GC-MS analysis

The berry extract was run on GC-MS-QP2010 model integrated with Rxi-5sil MS silica column (Singh and Sharma 2022). Different column temperatures were programmed into the GC protocol. For sample injection, a 10:1 split ratio was used. Helium was used as the carrier gas at a flow rate of 1.21 ml/min. The temperature of the injector was set to 250°C. Ion source and interface temperatures of 220°C and 270°C, respectively, were other important MS operating parameters.

The National Institute of Standards and Technology libraries, namely NIST14, NIST14s, and Wiley8 were considered for the identification of resolved metabolites. The peak intensity, retention time, and molecular mass of every individual metabolite were evaluated and used as source information for chemometric analysis using MetaboAnalyst4.0 software (Chong et al. 2018). The GC-MS spectrum verified the existence of numerous compounds with specific retention times in samples from (a) Badrinath, (b) Gangotri, and (c) Yamunotri, as illustrated in Fig. 1 and its associated metabolite in Table 1.

Table 1

List of crucial metabolites in differential solvent derivatized extracts of seabuckthorn (*H. salcifolia*) berry pulp

Metabolites	Molecular formula	Molecular mass (g/mol)	Retention time (minutes)	Mass fragmentation (m/z)	VIP values
2,3,3,4-Tetramethylpentane	C ₉ H ₂₀	128.2	10.17	43.05	0.5873
Tetracontane	C ₄₀ H ₈₂	563.1	31.52	57.1	0.1567
3,4,5-Trimethylheptane	C ₁₀ H ₂₂	142.2	9.73	43	0.2387
Butanedioic acid, 2TMS derivative	C ₁₀ H ₂₂ O ₄ Si ₂	262.4	13.72	147.1	0.0499
2,6-Bis (tert-butyl)phenol, TMS derivative	C ₁₇ H ₃₀ OSi	278.5	17.31	263.1	0.8725
11-Octadecenoic acid, methyl ester	C ₁₉ H ₃₆ O ₂	296.5	24.86	55.05	0.4062
Hexadecanoic acid, methyl ester	C ₁₇ H ₃₄ O ₂	270.4	22.87	74.05	0.038
9,12-Octadecadienoic acid (Z, Z)-, methyl ester	C ₁₉ H ₃₄ O ₂	294.4	25.29	67.05	1.024
Heneicosanoic acid, methyl ester	C ₂₂ H ₄₄ O ₂	340.6	28.31	74	1.272
4-Methoxymandelic acid, ethyl ester, TMS	C ₁₄ H ₂₂ O ₄ Si	282.4	35.37	209	1.497
Malic acid, 3TMS derivative	C ₁₃ H ₃₀ O ₅ Si ₃	350.6	16.54	73	0.3567
Oleic acid, trimethylsilyl ester	C ₂₁ H ₄₂ O ₂ Si	354.6	26.06	73.05	0.2943
Hexadecadienoic acid (Z, Z)-, methyl ester	C ₁₇ H ₃₀ O ₂	266.4	24.59	70.05	0.3595
Phytol, TMS derivative	C ₂₃ H ₄₈ OSi	368.7	25.55	143.1	1.594
Docosanol, TMS derivative	C ₂₅ H ₅₄ OSi	398.7	29.73	383.3	0.1274
1-Heptacosanol, TMS derivative	C ₃₀ H ₆₄ OSi	468.9	35.04	103.05	1.224
beta-Sitosterol, TMS derivative	C ₃₂ H ₅₈ OSi	486.9	40.03	129.1	1.612
Lupeol trimethylsilyl ether	C ₃₃ H ₅₈ OSi	498.8	43.71	189.1	1.815

Metabolites	Molecular formula	Molecular mass (g/mol)	Retention time (minutes)	Mass fragmentation (m/z)	VIP values
3,3-Dimethylhexanal	C ₈ H ₁₆ O	128.2	13.21	85.05	0.1022
Heptacosanal	C ₂₇ H ₅₄ O	394.7	35.05	82.05	0.8568
Ursolic aldehyde TMS ether	C ₃₃ H ₅₆ O ₂ Si	512.8	48.07	203.1	0.1053
Cyclohexanol, 5-methyl-2-(1-methylethyl)-, (1 α , 2 β , 5 α)	C ₁₀ H ₂₀ O	156.2	34.77	71.05	1.006
Silane, Trimethyl(9-octadecenyloxy)-, (Z)-	C ₂₁ H ₄₄ OSi	340.7	31.67	75.05	0.4304
1-Heptacosanol	C ₂₇ H ₅₆ O	396.7	31.35	97.1	1.296
9,19-Cyclolanostan-3-ol, 24-methylene-, (3. β .)-	C ₃₁ H ₅₂ O	440.7	43.06	95.05	0.1012
Docosyl heptafluorobutyrate	C ₂₆ H ₄₅ F ₇ O	522.6	30.39	57.05	0.693
β -Amyrin, TMS derivative	C ₃₃ H ₅₈ OSi	498.8	40.43	218.15	1.961
α -Amyrin, TMS derivative	C ₃₃ H ₅₈ OSi	498.9	41.24	218.1	2.009
Eicosanoic acid, TMS	C ₂₃ H ₄₈ O ₂ Si	384.7	28.37	117.05	1.055
1-Phenylethanone	C ₈ H ₈ O	120.1	9.22	105.05	0.7418
Squalene	C ₃₀ H ₅₀	410.7	32.48	69.05	0.2951

<Insert Fig. 1>

<Insert Table 1>

2.3. Sample preparation for ¹H NMR

The extraction process was modified slightly from the earlier described method (Liu et al. 2017). Two hundred mg of the powdered sample was vortexed in 1.0 ml of deuterated methanol (CD₃OD), and 0.3 ml pH buffer constituting sodium dihydrogen phosphate (NaH₂PO₄) in deuterium oxide (D₂O) (containing 0.04% 3-(trimethylsilyl)propionic-2,2,3,3-d₄ acid, sodium salt (TSMP) as internal chemical shift standard), which was then extracted by ultrasonication for 30 min at room temperature. The sample was centrifuged at 16000xg for 10 minutes after extraction, and the supernatant was passed to a 2ml Eppendorf tube. For ¹H NMR analysis, 1.5 ml of supernatant was transferred to a standard 5mm NMR tube.

2.3.1. ¹H NMR analysis

The deuterated plant extract was loaded in a 5-mm NMR tube and exposed to ¹H NMR (Bruker Advanced Spectrophotometer, frequency 500 MHz) with Z-gradient; BTO probe for a series of scans (16 scans). Temperature (318.9K), relaxation delay (4 s), water signal pre-saturation (25C), and pulse width (6.9 s) were also optimized to achieve a comparable signal-to-noise ratio. Using processing software, the collected spectra were manually phased and appropriately normalized to the internal standard (TMS 4.79 ppm) (TopSpin). FIDs (free induction decay) with a spectral width of 7,500 Hz were collected into 32 K data points. All spectra were treated to a 0.3 Hz line broadening function prior to Fourier treatment. The ¹H NMR spectra revealed the existence of numerous metabolites with different ppm values and chemical shifts of these representative berry samples from Badrinath, Yamunotri, and Gangotri regions, as illustrated in Fig. 2, and its associated metabolites are presented in Table 2.

Table 2
Metabolites obtained in seabuckthorn berries from Uttarakhand using ¹H NMR (500 MHz) spectroscopy

Metabolites	Shift	Shift range	Class
Glucose	3.89	3.90 .. 3.87	Doublet (d)
Malic acid	2.73	2.74 .. 2.71	Double doublet (d)
Methyl alcohol	3.09	3.11 .. 3.07	Singlet (s)
Choline	3.24	3.33 .. 3.11	Singlet (s)
Glycine	3.43	3.43 .. 3.35	Singlet (s)
Quinic acid	3.58	3.62 .. 3.58	Double doublet (dd)
Malate	4.3	4.41 .. 4.35	Double doublet (dd)
Dihydroxyacetone	2.1	2.12 .. 2.08	Singlet (s)
Fructose	4.02	4.04 .. 4.01	Multiplet (m)
Succinic acid	2.56	2.62 .. 2.58	Singlet (s)
L-quebrachitol	3.45, 3.63	3.40 .. 3.62	Singlet (s), Multiplet (m)
Dihydroxyacetone	2.1	2.11 .. 2.08	Singlet (s)
Valine	1.05	1.10 .. 1.03	Doublet (d)

<Insert Fig. 2>

<Insert Table 2>

Peak chemical shift range is often ascribed based on the peak position (ppm) and splits (singlet, doublet, etc.) corresponding to established standards' ^1H NMR signals. Integrating the intensities yield the relative abundances of each peak. The entire spectrum was used for ^1H NMR fingerprinting of seabuckthorn berry samples to visualize the differentiation pattern of seabuckthorn berries from the Uttarakhand region by integrating with multivariate statistical analysis. The raw spectrum was typically separated into small bins for this purpose, and a window size of equal width was assigned to the entire integrated region. The spectral regions that correspond with the deuterated solvent and internal standard signals were removed from the analysis. The retained bins and their cumulative signal intensities are often adjusted before being subjected to multivariate statistical analysis using MetaboAnalyst.

2.4. Data processing of GC-MS and ^1H NMR dataset

Before statistical analysis, the customized analytical data must be validated to ensure that the synonym variable in each sample has the same integrity. For this, the raw data undergoes different steps of peak alignment, feature detection, and normalization. The MetaboAnalyst4.0 software included in the R platform was employed for this purpose (Chong et al. 2019). A single zip file of GC-MS and ^1H NMR data sets (GCMS: peak intensity, retention time, and a peak mass fragmentation value (m/z) and ^1H NMR: peak position (ppm) and peak intensity table was uploaded separately. Each group has its subfolder. One peak list file per spectrum was included in each folder. A total of 30 samples were subjected to the MetaboAnalystR 4.0 software.

2.4.1. Normalization of GC-MS and ^1H NMR dataset

The GC-MS dataset, including peak intensity and peak mass fragmentation (m/z) and for ^1H NMR dataset including peak frequency (ppm), and intensity was loaded separately into the MetaboAnalyst software, and the submitted data were normalized with auto-scaling parameters before any descriptive statistics. All undesirable signals were removed after normalization, and pre-processed data were used for multivariate analysis.

2.5. Statistical analysis

The metabolome compositional dataset was determined and represented as mean $\pm$ standard deviation. The Principal Component Analysis (PCA) was performed on the metabolomic dataset to obtain an overview of the distribution pattern of different variables and their outliers (Eriksson et al. 2006). To maximize group separation, a supervised method called partial least squares-discriminant analysis (PLS-DA) was used (Antti et al. 2004). Here, for understanding the role of each principal component on the different groups of variable datasets, PCA and PCA scree plots were prepared. PLS-DA was conducted with a mean-centered pre-process through MetaboAnalyst4.0 (Chong et al. 2019). The distinct R^2 and Q^2 values reflect the model's quality by analyzing the model's class segregation and predictive ability, respectively. These parameters have a range of 0 to 1, and use cross-validation to interpret the variance in a test set (Q^2). Finally, the importance of the variable in the projection model (VIP) was documented to

identify the most significant variable for the discrimination of metabolites from different regions of Uttarakhand. The variable effect on projection (VIP) parameter was used to separate the most significant metabolites. For clinical assessment, VIP values > 1 and $p < 0.05$ were considered significant. In addition, a comparison of different group parameters was conducted by one-way analysis of variance (ANOVA) using the MetaboAnalyst4.0 software.

3. Results And Discussion

3.1. GC-MS and ^1H NMR-based metabolome profiling

In the present study, metabolome profiling of seabuckthorn (*H. salicifolia*) berries was performed using GC-MS and ^1H NMR techniques, and data obtained were subjected to multivariate statistical analysis to discriminate berries originating from Badrinath, Gangotri, and Yamunotri regions. The quantification of significant metabolites in GC-MS analysis was accomplished involving peak retention time, mass fragmentation value (m/z), and intensity, whereas peak ppm and intensity was used for qualitative analysis of ^1H NMR spectra.

The GC-MS spectra of 30 samples of plant extracts revealed the presence of a total of 305 metabolites with a matching retention time of respective spectra with NIST 14, NIST 14s, and Willey 8 library (Supplementary Table S2), whereas 15 unique metabolites signals were revealed by ^1H NMR spectral analysis (Table 2). The combined GC-MS and ^1H NMR spectra revealed the various metabolites present in the seabuckthorn berry samples such as amino acids, sugars, organic acids, secondary metabolites, etc. from Uttarakhand regions. The individual segments of the ^1H NMR spectra correlate to different characteristic signals of the protons of the metabolites, for example, the amino acids are represented by the 0 to 3 ppm region, while the sugar region is represented by 3 to 5.5 ppm region.

3.2. Metabolomic differentiation of berries

To study differences among berries of *H. salicifolia* collected from Badrinath, Gangotri, and Yamunotri, the peak retention time between 7.0 and 50 minutes was separately evaluated. The dataset was also exported into MetaboAnalyst4.0. On comparing with metabolites present in different libraries, including NIST14, NIST14S, and Willey8, a total of 305 compounds were recognized using GC-MS technique, including 135 in Badrinath, 68 in Yamunotri, and 173 in Gangotri samples. These metabolites could be placed in 23 different groups of compounds that varied in their area%, including fatty acid methyl esters, alkane and alkene-containing hydrocarbons, etc. as presented in Tables 3 and 4. Different groups of metabolomic groups such as alkane hydrocarbons, alkene hydrocarbons, fatty acid methyl esters, ketones, amides, alkyl ethers, alkyl esters, aldehydes, alcohols, fatty acids, triterpenes, sugar esters, nitriles, heterocyclic compounds, polyphenols, glycoside, chloride compounds, and thiols were highly identified in area percentage from Gangotri region whereas other compounds such as fluorinated compound, aldehydes, alcohols, sugars, and carbonyl compounds were major in Badrinath region. Group of compound such as fluorinated esters were major only in Badrinath region. Overall the total

metabolome area% in samples from different locations remained $22.9 \pm 10.4\%$ in Badrinath, $7.94 \pm 1.29\%$ in Yamunotri, and $45.9 \pm 3.98\%$ in Gangotri samples. The metabolome mass area% from different locations was in the order of Gangotri > Badrinath > Yamunotri (Table 3). The amount of these metabolites discovered in various locations followed a similar pattern (Table 4). In berries from the Gangotri region, the highest number of phytochemicals and their corresponding area% were found, followed by Badrinath and Yamunotri. The metabolomic information of 23 groups of metabolite data set is given in Table 3 and Table 4.

Table 3

List of groups of metabolites (mean $\pm$ standard deviation) area% identified in the differential solvent derivatized extract of seabuckthorn (*H. salicifolia*) berry samples from different locations in Uttarakhand

Uttarakhand			
Metabolites	Badrinath	Yamunotri	Gangotri
Alkane hydrocarbons	1.22 $\pm$ 0.0848	0.290	3.77 $\pm$ 1.072
Alkene hydrocarbons	0.180	0.175 $\pm$ 0.176	0.241 $\pm$ 0.178
Fatty acid methyl esters	3.35 $\pm$ 2.47	1.32 $\pm$ 0.241	4.94 $\pm$ 0.672
Fluorinated compound	0.0500 $\pm$ 0.0141	ND	ND
Fluorinated esters	0.645 $\pm$ 0.0353	0.880	0.510 $\pm$ 0.494
Ketones	0.220	0.100	1.28 $\pm$ 0.0424
Amides	0.0500	0.0600	0.290
Alkyl ethers	1.055 $\pm$ 0.507	0.180	14.4 $\pm$ 0.0932
Alkyl esters	9.85 $\pm$ 3.20	3.087 $\pm$ 0.736	14.5 $\pm$ 0.0932
Aldehydes	3.33 $\pm$ 3.83	0.490	0.465 $\pm$ 0.509
Alcohols	2.16 $\pm$ 0.236	1.31 $\pm$ 0.141	1.80 $\pm$ 0.644
Fatty acids	0.070	ND	2.77
Triterpenes	0.330	ND	0.685 $\pm$ 0.0212
Sugars	0.0900	ND	ND
Sugar esters	ND	ND	0.0100
Nitriles	ND	ND	0.0100
Nitrogen-containing compounds	0.0100	ND	0.0100
Heterocyclic compound	ND	ND	0.0700
Polyphenol	ND	ND	0.0100
Glycoside	ND	ND	0.0633 $\pm$ 0.0282
Chloride compounds	0.0650 $\pm$ 0.0212	0.0500	0.220 $\pm$ 0.141
Carbonyl compounds	0.0100	ND	ND
Thiols	0.0100	ND	0.0500

*ND represents metabolite not detected

Uttarakhand			
Total metabolome area %	22.9 ± 10.4	7.94 ± 1.29	45.9 ± 3.98
*ND represents metabolite not detected			

Table 4
Number of different groups of metabolites revealed in berry collections
from various regions in Uttarakhand

Uttarakhand			
Metabolites	Badrinath	Yamunotri	Gangotri
Alkane hydrocarbons	10	4	13
Alkene hydrocarbons	6	4	5
Fatty acid methyl esters	17	12	32
Fluorinated compound	1	ND	ND
Fluorinated esters	1	2	4
Ketones	5	1	13
Amides	1	1	1
Alkyl ethers	15	6	9
Alkyl esters	47	26	52
Aldehydes	7	1	4
Alcohols	15	10	22
Fatty acids	3	ND	2
Triterpenes	2	ND	2
Triterpene ethers	ND	ND	ND
Sugars	1	ND	ND
Sugar esters	ND	ND	1
Nitriles	ND	ND	1
Nitrogen-containing compounds	1	ND	2
Heterocyclic compound	ND	ND	1
Polyphenol	ND	ND	1
Glycoside	ND	ND	1
Chloride compounds	1	1	4
Carbonyl compounds	1	ND	ND
Thiols	1	ND	3

*ND represents metabolite not detected

Uttarakhand			
Organosilicons	ND	ND	ND
Lactone	ND	ND	ND
Carbazole	ND	ND	ND
Silicate anion	ND	ND	ND
Purine	ND	ND	ND
Total metabolites	135	68	173
*ND represents metabolite not detected			

<Insert Table 3>

<Insert Table 4>

In ^1H NMR analysis, the signals were assessed preliminary based on literature and using the Chenomx database library. The NMR peak list (ppm) and intensity data were used to designate the name of the metabolite. The signal strength in the ^1H NMR spectrum is directly proportional to the molar concentration of the metabolite (Fan et al. 2013; Chen et al. 2016). The individual peak intensity of the metabolite was compared at different regions in Uttarakhand. The characteristic signal of glycine (3.43 ppm), Choline (3.43 ppm) and methyl alcohol (3.09 ppm) were highly intense at the Badrinath region, whereas the characteristic signals of malate (4.30 ppm), fructose (4.02 ppm), glucose (3.89 ppm), quinic acid (3.58 ppm), malic acid (2.73 ppm), and succinic acid (2.56 ppm) showed lower intensity and almost similar profile from Badrinath, Yamunotri, and Gangotri region. Other metabolites such as dihydroxyacetone (2.10 ppm) showed their presence at lower intensity in Badrinath and Yamunotri samples but absence in Gangotri samples. Two metabolites such as valine (1.05 ppm) and L-quebrachitol (3.45 and 3.63 ppm) showed its minute presence in Gangotri samples but the signal was absent in Badrinath samples (Fig. 2). The compound's peak shift, and its class is given in Table 2.

3.3. Multivariate statistical analysis

To understand how berry samples from Badrinath, Gangotri, and Yamunotri regions are distinguishable from each other, principal component analysis (PCA) was used on the preliminary dataset that was the same in berry samples from all three locations. Before any supervised statistical analysis, these first evaluations of the dataset assessed the quality of the data. This unsupervised approach (PCA) allows the detection of outliers as well as the trend of data grouping. The five different latent variables (principal components) were retrieved after PCA was applied to the GC-MS and ^1H NMR dataset, as illustrated in a scree plot (Supplementary Fig. S1a and Supplementary Fig. S1b). The PC1 and PC2 of the PCA scree plot obtained from both GC-MS and ^1H NMR dataset exhibited the most discriminating features among berry samples when all five latent variables were considered (Supplementary Fig. S2a and Supplementary Fig.

S3a). PCA was used to analyze the wide range of metabolites found in the 30 seabuckthorn berry collections from various regions, with two significant principal components (PC1/PC2) accounting for a higher percentage of variance (GC-MS (PC1 = 36%) and ^1H NMR (PC1 = 53%) of the overall variation from individual GC-MS and ^1H NMR data analysis (Supplementary Fig. S2a and Supplementary Fig. S3a). The overlaps in samples from three regions indicated that the PCA model applied was not suitable for separating berry samples representing three different locations (Supplementary Fig. S2b and Supplementary Fig. S3b). However, we also performed a one-way ANOVA test followed by posthoc analysis to identify differences amongst the berry samples. The results exhibited no significant difference among the peaks (mass) at the p-value threshold of 0.05.

To establish an effective classification model, multiple multivariate analytical approaches were performed. For each naturally developing population of *H. salicifolia* from different regions, a two-component model (PLS-DA) was used to determine the interclass differentiation and examine putative distinctive markers. The mean-centered GC-MS and ^1H NMR dataset's PLS-DA pairwise score plot (GC-MS: PC1 = 18.6% and PC2 = 9.7%) and (^1H NMR: PC1 = 22.6% and PC2 = 21.6%) showed the most discriminating ability in comparison to other PCs, as shown in Fig. 3a and Fig. 3c. PC1 and PC2 are the loading plots of the first two latent variables (LVs). Since the first component (PC1) defines maximum data variation and the second component (PC2), it may be concluded reasonably that the first component (PC1) explains the majority of the variation. We identified interesting trends in the sample distribution by examining the score plot (Fig. 3b). Even though there is some overlap when represented on a 2D score plot, the two principal components together contributed a significant percentage of the variation. Berry samples from Badrinath and Yamunotri could be discriminated from each other, whereas berry samples from Gangotri showed some overlap with samples from the other two regions (Fig. 3b). All the samples from Badrinath have a positive score on PC2 and negative scores on PC1. Samples from Gangotri showed positive and negative scores on both LVs (PC1 and PC2), whereas samples from Yamunotri showed positive scores on PC1 and PC2 plots. Similarly, when the PLS-DA 2D score plot using NMR data was generated, no apparent differentiation was evident in samples from Badrinath, Gangotri, and Yamunotri, and the data was mostly overlapping (Fig. 3d).

<Insert Fig. 3>

On the other hand, looking at the score plot and assessing the PLS-DA loading plot of GC-MS dataset revealed a reasonable command of variables that lead to class discrimination. On the loading plot, both latent variables (PC1 and PC2) showed the lower and higher quantitative composition of the different classes of metabolites. The negative loadings on the first and second PC plots representing beta-sitosterol, TMS derivative, eicosanoic acid, TMS, and 1-heptacosanol provide a piece of evidence that these parameters have negative logarithmic scale values (Fig. 4) due to low abundance and belong to the Gangotri area. Other metabolites included hexadecanoic acid, methyl ester, 9,12-octadecadienoic acid (Z, Z)-, methyl ester, tetracontane, 11-octadecenoic acid, methyl ester have negative loading on the first latent variable (PC1) plot and positive loading on second latent variable (PC2) plot representing the higher concentration of metabolites from the Gangotri region. Berry samples from Yamunotri accommodated

the higher concentration of 1-phenylethanone, cyclohexanol, 5-methyl-2-(1-methylethyl)-, (1 α , 2 β , 5 α)-, 3, 3-dimethylhexanal, 3,4,5-trimethylheptane, 1-heptacosanol, TMS derivative, and lupeol trimethylsilyl ether, which were located on the positive axis of loading plot (PC1 and PC2). The metabolites from Badrinath showed positive values on PC1 plot and negative values on PC2 plot bysilane, trimethyl(9-octadecenyl)-, (Z)-, heneicosanoic acid, methyl ester, α -amyrin, TMS derivative, phytol, TMS derivative, phytol, TMS derivative, 9,19 cyclolanostan-3-ol, 24 methylene-, (3 β)-, 4-methoxymandelic acid ethyl ester, TMS, squalene, hexadecadienoic acid, methyl ester, urosolic aldehyde, TMS ether, docosanol, TMS derivative, 2,3,3,4-tetramethylpentane, heptacosanol, malic acid, 3TMS derivative, butanedioic acid, 2TMS derivative, 2,6-bis(tert-butyl) phenol TMS derivative, butanedioic acid, bis(trimethylsilyl)ester, docosyl heptafluorobutyrate. The chromatogram from three locations in Uttarakhand revealed the presence of these valuable metabolites at varying retention times (Fig. 1). The associated metabolites' information at different retention times and molecular mass of the respective spectra (Table 1) confirmed the presence of different metabolites retrieved from the PLS-DA loading plot.

<Insert Fig. 4>

PLS model has important features also for the effective quantitative measure of discriminating ability of each metabolite due to variable importance for the projection (VIP) parameter. To identify the essential metabolites amongst the samples from Badrinath, Gangotri, and Yamunotri, we performed partial least square discriminant analysis (PLS-DA) with the variable importance of projection (VIP) parameter. Based on VIP values, PLS-DA revealed the most essential metabolite (Fig. 5). A variable with a VIP value greater than 1 is considered essential, and it is used to find metabolites that help to distinguish different locations of seabuckthorn berry samples. Here, fifteen metabolites were considered to be distinguishing metabolites (VIP > 1) among different berry samples. Furthermore, boxplots represent the mean values with standard deviation intervals for these selected metabolite concentrations (Fig. 6) to highlight metabolomic differences found among berry samples. Although statistically significant, the concentration of these metabolites changes with the growing locations of Uttarakhand.

<Insert Fig. 5>

<Insert Fig. 6>

The examination of these 15 metabolites shown in the VIP plot and Box plot demonstrated that β -amyrin, TMS derivative, β -sitosterol, TMS derivative, 1-heptacosanol, eicosanoic acid, TMS were expressed in higher concentrations in Badrinath berry samples comparative to samples from Gangotri and Yamunotri regions. Other essential metabolites such as heneicosanoic acid, methyl ester, 9,12 octadecadienoic acids (Z, Z)-, methyl ester, 3,3-dimethylhexanal were significantly higher in berry samples from Gangotri, whereas metabolites such as α -amyrin, TMS derivative, lupeol trimethylsilyl ether, phytol, TMS derivative, 4-methoxymandelic acid, ethyl ester, TMS, 1-heptacosanol, TMS derivative, cyclohexanol, 5-methyl-2-(1-methylethyl)-, (1 α , 2 β , 5 α), urosolic aldehyde, TMS ether, 2, 6-bis(tertbutyl)phenol TMS derivative were higher in samples from Yamunotri region (Fig. 5 and Fig. 6). The varied concentration of these metabolites from each location was represented as Log (concentration/200 mg) of berry samples

(Fig. 6). The higher concentration of metabolites from these three locations was expressed as the smallest upper area of error bars.

For a few key metabolites such as α -amyrin, TMS derivative, lupeol trimethylsilyl ether, phytol, TMS derivative, 4-methoxymandelic acid, ethyl ester, TMS, 1-heptacosanol, TMS derivative, cyclohexanol, 5-methyl-2-(1-methylethyl)-, (1 α , 2 β , 5 α), urosolic aldehyde, TMS ether, a dynamic variation in signal intensity was detected. Furthermore, the PLS-DA loading plot as well as the PLS-DA combined with the VIP plot and boxplot demonstrated that these metabolites were crucial discriminators of *H. salicifolia* berries from different locations of Uttarakhand. Metabolites that revealed values of more than 1 in PLS-DA biplot and during VIP plot analysis were considered as discriminators of seabuckthorn berries.

Seabuckthorn berries have long been recognized for their medicinal and nutritional properties but have also recently gained international interest for being a metabolite repository for researchers in biotechnology and pharmacology (Stobdan et al. 2013). As a result, the growing appeal for seabuckthorn berries for industrial reasons necessitates improvement in source plant identification and quality control. The PLS-DA loading plot and VIP plot revealed a higher concentration of 15 metabolites, including four from Badrinath, three from Yamunotri, and eight from Gangotri, whereas ^1H NMR revealed 15 unique metabolites not resolved in GC-MS spectral analysis. The overall 318 metabolites (GC-MS: 305, and ^1H NMR: 13) were identified by combining GC-MS and ^1H NMR data. Our findings imply that metabolites isolated in PLS-DA-VIP plot using GC-MS data can be crucial discriminators for the quality evaluation of seabuckthorn berries from various geographic regions. Hence, the information on metabolite variation could be useful for seabuckthorn cultivation and conservation programs.

According to the available evidence, seabuckthorn berries contain a large number of metabolites (Bal et al. 2011). Our study revealed 15 metabolomic indicators identified in different regions of Uttarakhand. The -amyrin esters (triterpene esters), which are abundant in Yamunotri samples, have been shown to have substantial biological functions, including cytotoxic and anti-inflammatory properties (Akihisa et al. 2010; Barros et al. 2010). Triterpene esters, such as -amyrin and -amyrin esters, exhibit anti-inflammatory properties and have been shown to suppress the action of Epstein-Barr virus early antigen (EBV-EA) in Raji cells (Holanda et al. 2008). Earlier, it has been recorded that terpenoids have an inhibitory effect against bacteria. Still, the mechanism of action of these metabolites remains unknown, however, it is thought to include lipophilic compound disrupting membranes (Oliveira et al. 2010). Another study revealed that esters of α -amyrin and β -amyrin could modulate antibacterial activity (Fernandes et al. 2014). Lupeol ester highly expressed in the Yamunotri region possesses a positive effect on malaria (Fotie et al. 2006) and acetylcholinesterase inhibitory activities (Gurovic et al. 2010). In the past decades, phytosterols and their derivatives have been extensively used in diverse food and cosmetic industries. These analyzed bioactives have recently received appreciable attention as nutraceutical additives. Phytosterol such as β -sitosterol, identified in a significant amount in Badrinath samples, could have the ability to lower the level of cholesterol in the human body and is also widely used as a food additive (Moreau et al. 2002). The higher expression of phytol TMS derivatives exhibited in Yamunotri samples be linked with the ability to inhibit the activity of *Plasmodium* showing antimalarial activity (Saxena et al.

2018). Further, 4-methoxymandelic acid esters derivatives have been reported to alleviate the symptoms of wrinkles (Eugene et al. 1997). The presence of 1-heptacosanol was significantly higher in Badrinath samples, advocating their utility in reducing the cholesterol level in addition to antimicrobial and antioxidant activity (Imada 2005). In food items, it is used as flavour and fragrance agent. Metabolite like 1-heptacosanol has also been reported to have antithrombotic effect (Isoe et al. 2015). Fatty acid methyl esters (heneicosanoic acid methyl ester, 1-heptacosanol, TMS derivative, eicosanoic acid, TMS derivative, 9,12-octadienoic acid (Z, Z)-, methyl ester, and 2,6-Bis(tert butyl), TMS derivative) were highly expressed in berries from the different regions of Uttarakhand and should be explored in detail for antibacterial, antifungal and antioxidant activities (Salem et al. 2014). Recently, urosolic derivatives have been reported to serve as potential candidates for developing effective anti-*T. gondii* drugs with lesser side effects (Luan et al. 2019). Similarly, 3,3-dimethylhexanal, an aldehyde present in significantly higher amounts in Gangotri samples, and earlier reported in beef, is a flavoring agent (Ramarathnam et al. 1991). The other crucial and higher intensity of compounds revealed in ^1H NMR analysis such as glycine could be used in nutraceuticals as a flavouring agent, and as a hepatoprotective agent (<https://pubchem.ncbi.nlm.nih.gov/compound/Glycine>). The other highly intensity of compound choline is required by the human body for the production of acetylcholine, an important neurotransmitter for memory, mood, muscle control, and other brain and nervous system functions. Choline is also involved in gene expression regulation, cell membrane communication, lipid transport and metabolism, and early brain development (Zeisel et al. 2010; Zeisel et al. 2012). These crucial metabolites identified from different locations of Uttarakhand for the first time could create a vast repository of resources important for the formulation of various seabuckthorn-based industrial products.

The OPLS-DA is a powerful supervised statistical technique for dimension reduction and discovery of spectral features that govern group classification, similar to the PLS-DA. When compared to PLS-DA, OPLS-DA produces models that are less complex. We did, however, apply both OPLS-DA and PLS-DA after examining the cluster formation from different regions of Uttarakhand using GC-MS dataset for cross-validation to ensure model fidelity (Worley and Powers 2016). $R^2\text{X}(\text{cum})$, $R^2\text{Y}(\text{cum})$, and $Q^2(\text{cum})$ for the OPLS-DA model were 0.0976, 0.387, and 0.0794, respectively. The $R^2\text{Y}$ had a high goodness of fit and a high $Q^2(\text{cum})$, indicating that it was a good predictor. The validation tool was evaluated using the OPLS-DA model, which showed $R^2\text{Y}$ intercept of 0.791 and a $Q^2\text{Y}$ intercept of -0.073 (Supplementary Fig. S4). These intercept values indicate good model reliability. The PCA plot did not exhibit any good separation, whereas PLS-DA and OPLS-DA score plots showed good separation between samples from different locations of Uttarakhand (Supplementary Fig. S5). The higher R^2 (0.79) value obtained from the OPLS-DA plot demonstrated good model validity.

We also attempted to find any correlation between the impact of altitude and temperature at growing locations on the concentration of these fifteen metabolites in *H. salicifolia* berry samples. However, we could not find any significant correlation suggesting thereby that there would be some other factors also which may affect the concentration of these metabolites at different locations of Uttarakhand.

4. Conclusion

The present study on the GC-MS and ^1H NMR-based metabolomic profiling of berry samples from different regions of Uttarakhand revealed the presence of 318 metabolites. The analysis of metabolome profiles revealed crucial discriminators of seabuckthorn berries by using the supervised PLS-DA-VIP approach. The findings provide a unique method for evaluating the quality of seabuckthorn berry products for nutraceutical and therapeutical uses, as well as a platform for distinguishing seabuckthorn berries originating from different sources.

Declarations

Declaration of Competing Interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments SS acknowledges a short-term research fellowship awarded by Guru Gobind Singh Indraprastha University, New Delhi.

Author contributions Sugandh Singh collected plant material, performed analysis and prepared the first draft of the manuscript. P. C. Sharma conceptualized and designed the study, and finalized the manuscript.

Funding PCS acknowledges a Faculty Research Grant awarded by the Guru Gobind Singh Indraprastha University, New Delhi.

Data availability Data that support the finding of this study are available from the corresponding author upon reasonable request.

References

1. Akihisa T, Kojima N, Kikuchi T, Yasukawa K, Tokuda HT, Masters E, Manosroi A, Manosroi J (2010) Anti-inflammatory and chemopreventive effects of triterpene cinnamates and acetates from shea fat. *J Oleo Sci* 59:273–280
2. Antti H, Ebbels T, Keun HC, Bollard ME, Beckonert O, Lindon JC, Nicholson JK, Holmes E (2004) Statistical experimental design and partial least squares regression analysis of biofluid metabonomic NMR and clinical chemistry data for screening of adverse drug effects. *Chemom Intell Lab Syst* 73:139–149
3. Aras U, Akkemik Z, Kaya (2007) *Hippophae rhamnoides* L: Fruit and seed morphology and its taxonomic problems in Turkey. *Pak J Bot* 39(6):1907–1916
4. Bal LM, Meda V, Naik SN, Satya S (2011) Seabuckthorn berries: a potential source of valuable nutrients for nutraceuticals and cosmeceuticals. *Food Res Int* 44:1718–1727. <https://doi.org/10.1016/j.foodres.2011.03.002>

5. Barros FW, Bandeira PN, Lima DJ, Meira AS, de Farias SS, Albuquerque MR, dos Santos HS, Lemos TL, de Moraes MO, Costa-Lotufo LV, Pessoa C (2010) Amyrin esters induce cell death by apoptosis in HL-60 leukemia cells. *Bioorg Med Chem* 19(3):1268–1276
6. Chen X, Guo Y, Hu Y, Yu B, Jin Q (2016) Quantitative analysis of highly similar salvianolic acids with ^1H qNMR for quality control of traditional Chinese medicinal preparation salvianolate lyophilized injection. *J Pharm Biomed Anal* 24:281–287. <https://doi.org/10.1016/j.jpba.2016.02.016>
7. Chong J, Soufan O, Li G, Garaus I, Li S, Bourque G, Xia J (2018) MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res* 46(W1):W486–W494
8. Chong J, Wishart DS, Xia J (2019) Using MetaboAnalyst 4.0 for comprehensive and integrative metabolomics data analysis. *Curr Protoc Bioinform* 68:e86
9. Cui Q, Lewis IA, Hegeman AD, Anderson ME, Li J, Schulte CF, Westler WM, Eghbalian HR, Sussman MR, Markley JL (2008) Metabolite identification via the Madison Metabolomics Consortium Database. *Nat Biotechnol* 26(2):162–164
10. Datta A, Kamthan A, Kamthan M, Chakraborty N, Chakraborty S (2012) A simple protocol for extraction, derivatization, and analysis of tomato leaf and fruit lipophilic metabolites using GCMS. *Protoc Exch* 10. <http://doi:101038/protex.2012.061>
11. Dhyani D, Maikhuri RK, Dhyani S (2012) Effect of auxin treatments on male and female cuttings of *Hippophae salicifolia*. *Afr J Biotechnol* 11(90):15712–15718
12. Egon S (2005) Thin-layer chromatography: A laboratory handbook, 2nd edn. Springer, New York
13. Eriksson L, Kettaneh-Wold N, Trygg J, Wikström C, Wold S (2006) Multi-and megavariable data analysis part I: Basic principles and applications, 2nd ed. Umetrics Academy
14. Eugene JVS, Pennsylvania A, Ruey JY, Pennsylvania A (1997) Use of hydroxycarboxylic acids and derivatives for the preparation of topical dermatologic compositions for the treatment of wrinkles. Leonhard - Olgemoller - Fricke, Patentanwalt, Josephspitalstrasse 7 80331 München (DE), European patent EP 0 599 819 B1
15. Fan G, Zhang MY, Zhou XD, Lai XR, Yue QH, Tang C, Luo WZ, Zhang Y (2012) Quality evaluation and species differentiation of *Rhizomacoptidis* by using proton nuclear magnetic resonance spectroscopy. *Anal Chim Acta* 747:76–83. <https://doi.org/10.1016/j.aca.2012.08.038>
16. Fancy SA, Rumpel K (2008) GC-MS-Based Metabolomics. In: Wang, F. (Ed.), *Biomarker methods in drug discovery and development*, Humana Press. *Methods Pharmacol Toxicol*. https://doi.org/10.1007/978-1-59745-463-6_15
17. Fernandes CP, Corrêa AL, Lobo JFR, Caramel OP, de Almeida FB, Castro ES, Souza KFCS, Burth P, Amorim LMF, Santos MG, Ferreira JLP, Falcao DQ, Carvalho JCT, Rocha L (2014) Triterpene esters and biological activities from edible fruits of *Manilkara subsericea* (Mart.) Dubard, Sapotaceae. *Biomed Res Int* 2013:1–7
18. Forseth RR, Fox EM, Chung D, Howlett BJ, Keller NP, Schroeder FC (2011) Identification of Cryptic Products of the Gliotoxin Gene Cluster Using NMR-Based Comparative Metabolomics and a Model for Gliotoxin Biosynthesis. *J Am Chem Soc* 133(25):9678–9681

19. Fotie J, Bohle DS, Leimanis ML, Georges E, Rukunga G, Nkengfack AE (2006) Lupeol long-chain fatty acid esters with antimalarial activity from *Holarrhena floribunda*. *J Nat Prod* 69(1):62–67
20. Gurovic MS, Castro MV, Richmond VJ, Faraoni MB, Maier MS, Murray AP (2010) Triterpenoids with acetylcholinesterase inhibition from *Chuquiragaerinacea* D. Don. Subsp. *Erinacea* (Asteraceae). *Planta Med* 76(6):607–610
21. Holanda Pinto SA, Pinto LMS, Cunha GMA, Chaves MH, Santos FA, Rao VS (2008) Anti-inflammatory effect of α , β -Amyrin, a pentacyclic triterpene from *Protiumheptaphyllum* in rat model of acute periodontitis. *Inflammopharmacology* 16(1):48–52
22. Holzgrabe U, Diehl BWK, Wawer I (1998) NMR spectroscopy in pharmacy. *J Pharm Biomed Anal* 17(4–5):557–616
23. Imada C (2005) Enzyme inhibitors and other bioactive compounds from marine *Actinomycetes*. *Antonie Van Leeuwenhoek* 87:59–63
24. Isoe ME, Alagar YS, TPD (2015) Deivamarudachalam, Spectral analyses of the bioactive compounds present in the ethanolic leaf extract of *Strobilanthes kunthiana* (Nees) T. Anderson ex. Benth *Adv Biores* 6(3):65–67
25. Liu Y, Fan G, Zhang Y, Li J, Xiong C, Qi Z, Li X, Lai X (2017) Metabolic discrimination of seabuckthorn from different *Hippophae* species by ^1H NMR based metabolomics. *Sci Rep* 7:1587
26. Luan T, Jin C, Jin CM, Gong GH, Quan ZS (2019) Synthesis and biological evaluation of ursolic acid derivatives bearing triazole moieties as potential anti-Toxoplasma gondii agents. *J Enzyme Inhib Med Chem* 34(1):761–772
27. Manu P, Lal A, Rani A (2014) *Hippophae salicifolia* d don- a plant with multifarious benefits. *Int J Pharm Pharm Sci* 6(11):37–40
28. Moreau RA, Whitaker BD, Hicks KB (2002) Phytosterols, phytostanols, and their conjugates in foods: structural diversity, quantitative analysis, and health-promoting uses. *Prog Lipid Res* 41:457–500
29. Nicholson JK, Lindon JC, Holmes E (1999) Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* 29(11):1181–1189
30. Oliveira AP, Cruz RAS, Botas GS, Gonzalez MS, Santos MG, Teixeira L, Rocha L (2010) Chemical and biological investigations of *Pilocarpus spicatus* essential oils. *Bol Latinoam Caribe Plant Med Aromat* 9(3):206–211
31. Oms-Oliu G, Odriozola-Serrano I, Martin-Belloso O (2013) Metabolomics for assessing safety and quality of plant-derived food. *Food Res Int* 54:1172–1183
32. Raina SN, Jain S, Sehgal D, Kumar A, Dar TH, Bhat V, Pandey V, Vaishnavi S, Bhargav A, Singh V, Rani V, Tandon R, Tewari M, Mahmoudi A (2011) Diversity and relationships of multipurpose seabuckthorn (*Hippophae* L.) germplasm from the Indian Himalayas as assessed by AFLP and SAMPL markers. *Genet Resour Crop Evol* 59:1033–1053

33. Ramarathnam N, Rubin LJ, Diosady LL (1991) Studies on Meat Flavor. 2. A quantitative investigation of the volatile carbonyls and hydrocarbons in uncured and cured beef and chicken. *J Agric Food Chem* 39:1839–1847
34. Ranjith A, Kumar SK, Venugopalan VV, Arumughan C, Sawhney RC, Singh V (2006) Fatty acids, tocopherols and carotenoids in pulp oil of three seabuckthorn species (*Hippophae rhamnoides*, *H. salicifolia*, and *H. tibetana*) grown in the Indian Himalayas. *J Am Oil Chem Soc* 83:359–364
35. Saito K, Matsuda F (2010) Metabolomics for functional genomics, systems biology and biotechnology. *Ann Rev Plant Biol* 61:463–489
36. Salem MZM, Ali HM, Mansour MM (2014) Fatty acid methyl esters from air-dried wood, bark, and leaves of brachychiton diversifolius R. Br: antibacterial, antifungal, and antioxidant activities. *BioResources* 9(3):3835–3845. <http://doi:10.15376/biores.9.3.3835-3845>
37. Saxena A, Upadhyay HC, Cheema HS, Srivastava HK, Darokar MP, Bawankule DU (2018) Antimalarial activity of phytol derivatives: in vitro and in vivo study. *Med Chem Res* 27:1345–1354
38. Satyawati GV, Gupta AK (1987) Medicinal plants of India. Indian Council of Medical Research, New Delhi
39. Singh S, Sharma PC (2011) Gas chromatography-mass spectrometry (GC-MS) profiling reveals substantial metabolome diversity in seabuckthorn (*Hippophae rhamnoides* L.) berries originating from different geographical regions in the Indian Himalayas. *Phytochem Anal* 33(2):214–225
40. Stobdan T, Korekar G, Srivastava R (2013) Nutritional attributes and health application of seabuckthorn (*Hippophae rhamnoides* L.) - A Review. *Curr Nutr Food Sci* 9:151–165
41. Verpoorte R (1999) Chemodiversity and the biological role of secondary metabolites, some thoughts for selecting plant material for drug development. In *Bioassay Methods in Natural Product Research and Drug Development*. 11–23. <http://doi.org/10.1007/978-94-011-4810-8-2>
42. Wishart DS, Knox C, Guo AC, Eisner R, Young N, Gautam B, Hau DD, Psychogios N, Dong E, Bouatra S, Mandal R, Sinelnikov I, Xia J, Jia L, Cruz JA, Lim E, Sobsey CA, Shrivastava S, Huang P, Liu P, Fang L, Peng J, Fradette R, Cheng D, Tzur D, Clements M, Lewis A, De Souza A, Zuniga A, Dawe M, Xiong Y, Clive D, Greiner R, Nazyrova A, Shaykhutdinov R, Li L, Vogel HJ, Forsythe I (2009) HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res* 37:D603–D610
43. Worley, Powers R (2016) PCA as a Practical Indicator of OPLS-DA Model Reliability. *Curr Metabolomics* 4(2):97–103. <http://doi.org/10.2174/2213235x04666160613122429>
44. Zeisel SH, Corbin KD (2012) Choline. In: Erdman JW, Macdonald IA, Zeisel SH (eds) *Present Knowledge in Nutrition*, 10th edn. Wiley-Blackwell, Washington, DC, pp 405–418
45. Zeisel SH (2010) Choline. In: Coates PM, Betz JM, Blackman MR et al (eds) *Encyclopedia of Dietary Supplements*, 2nd edn. London and New York, Informa Healthcare, pp 136–143

Figures

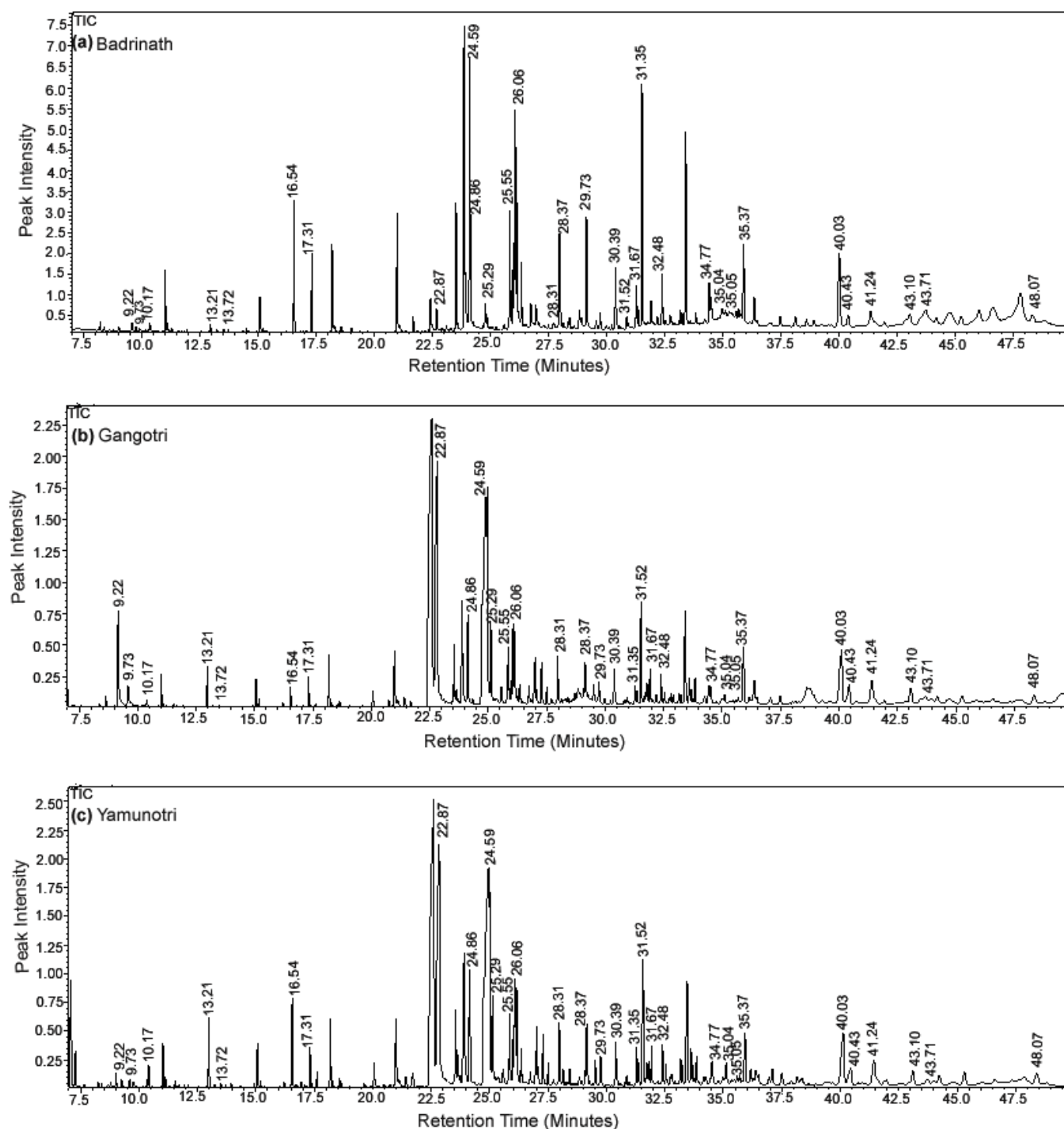


Figure 1

GC-MS chromatogram at different retention times in total ion chromatogram illustrating the presence of essential metabolites in seabuckthorn (*H. salcifolia*) berries from (a) Badrinath, (b) Gangotri, (c) Yamunotri

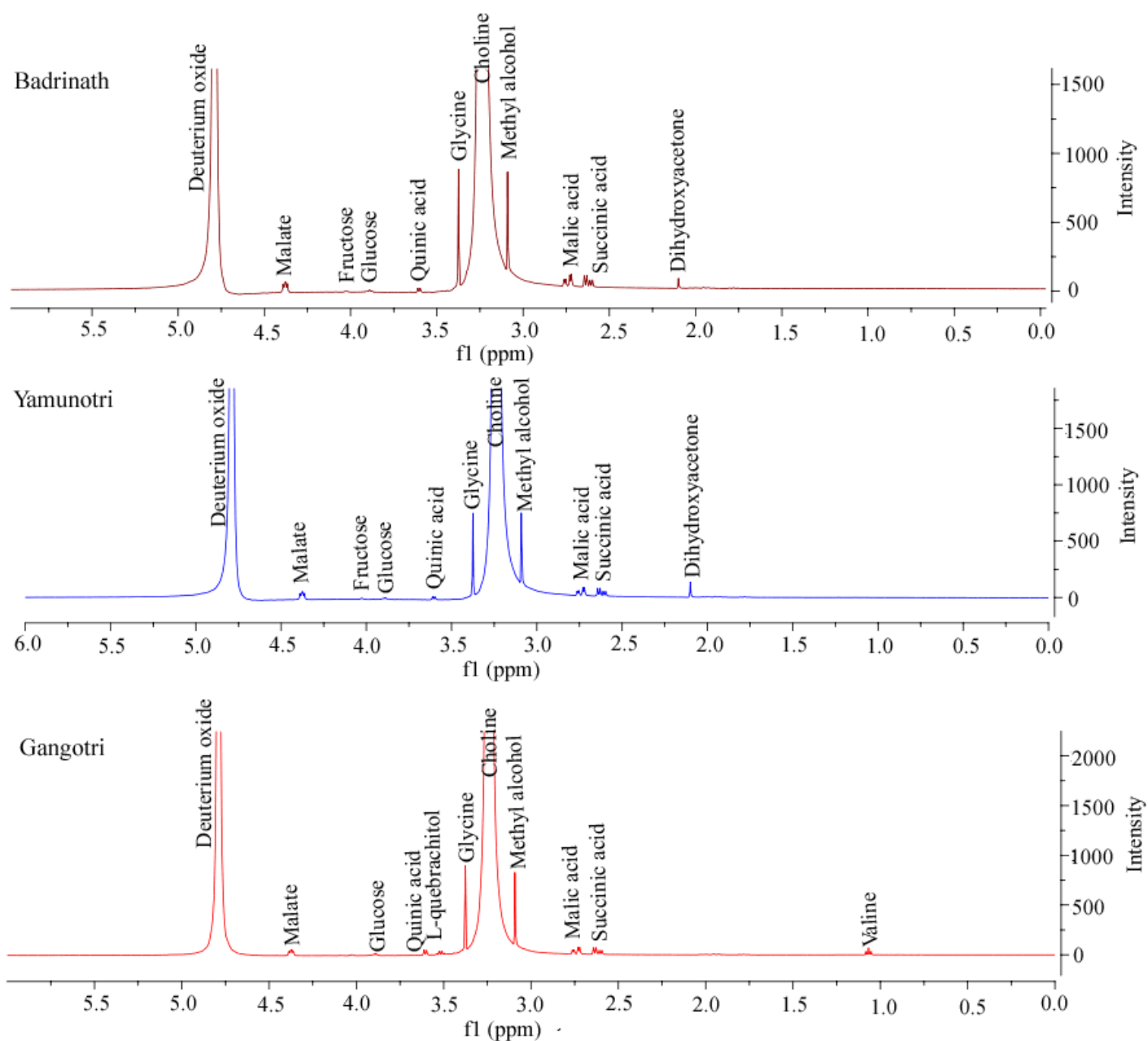


Figure 2

A representative ^1H NMR spectra at different peak position (ppm) and peak intensity illustrating the presence of essential metabolites in seabuckthorn (*H. salicifolia*) berries from (a) Badrinath, (b) Yamunotri, (c) Gangotri

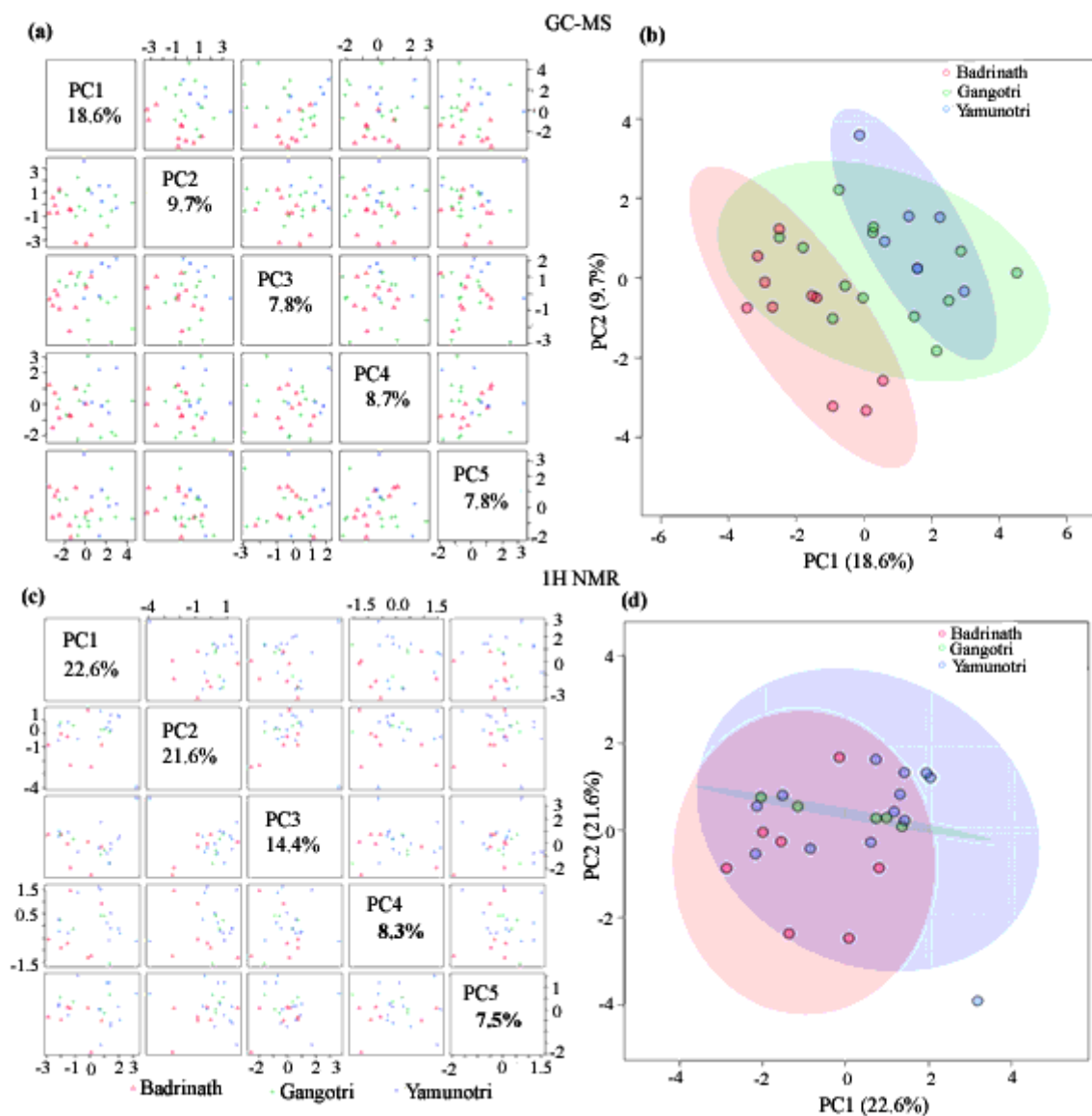


Figure 3

PLS-DA pairwise score plot obtained from GC-MS and ^1H NMR data (a and c) PC1 and PC2 shows a maximum percentage of variance among all PCs. PLS-DA: 2D- PCA score plot (b and d) represents clustering pattern of berry samples from Badrinath, Gangotri and Yamunotri

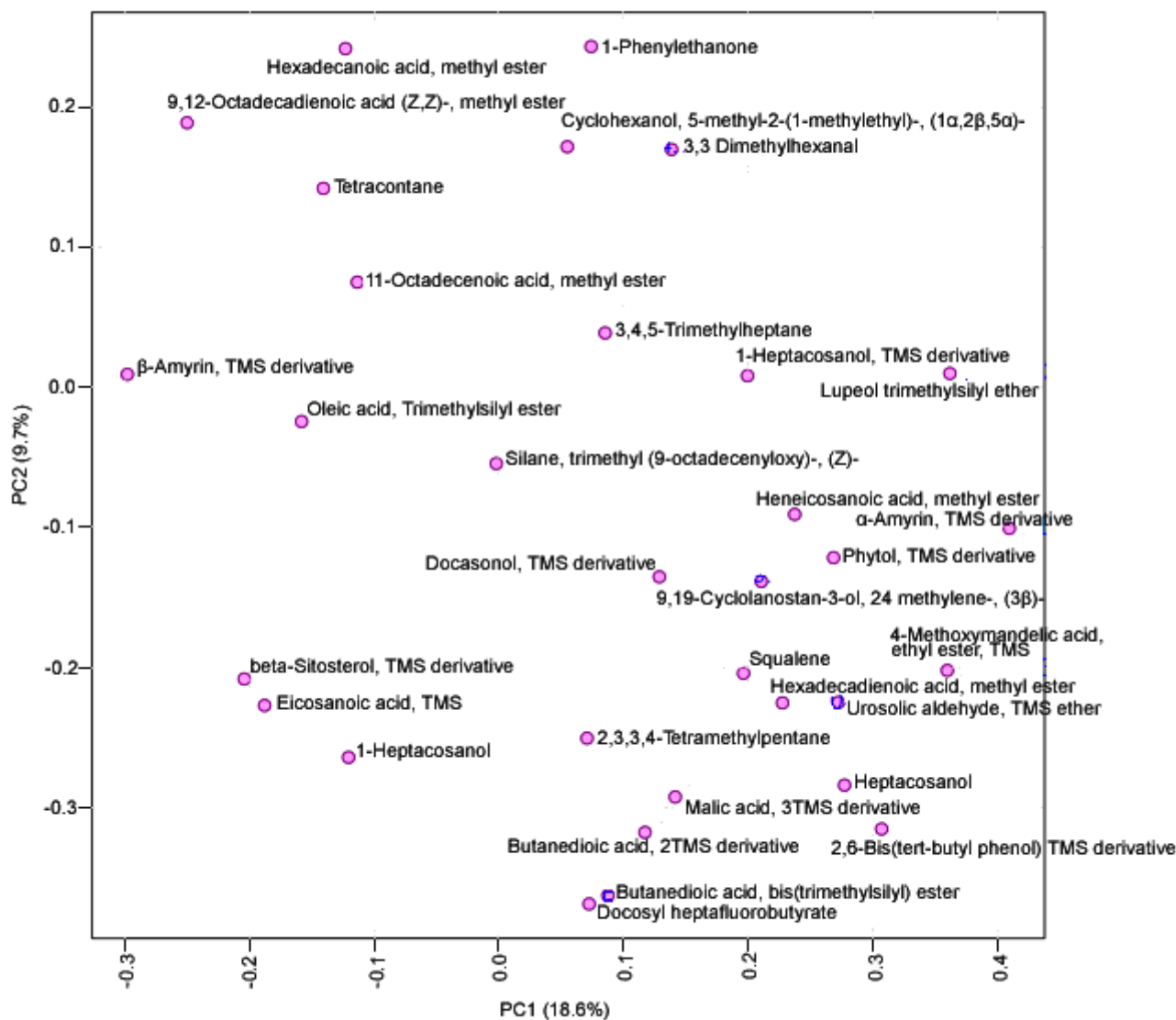


Figure 4

GC-MS based PLS-DA loading plot of the first two latent variables shows different metabolite classes from varying locations of Uttarakhand displayed by score plot

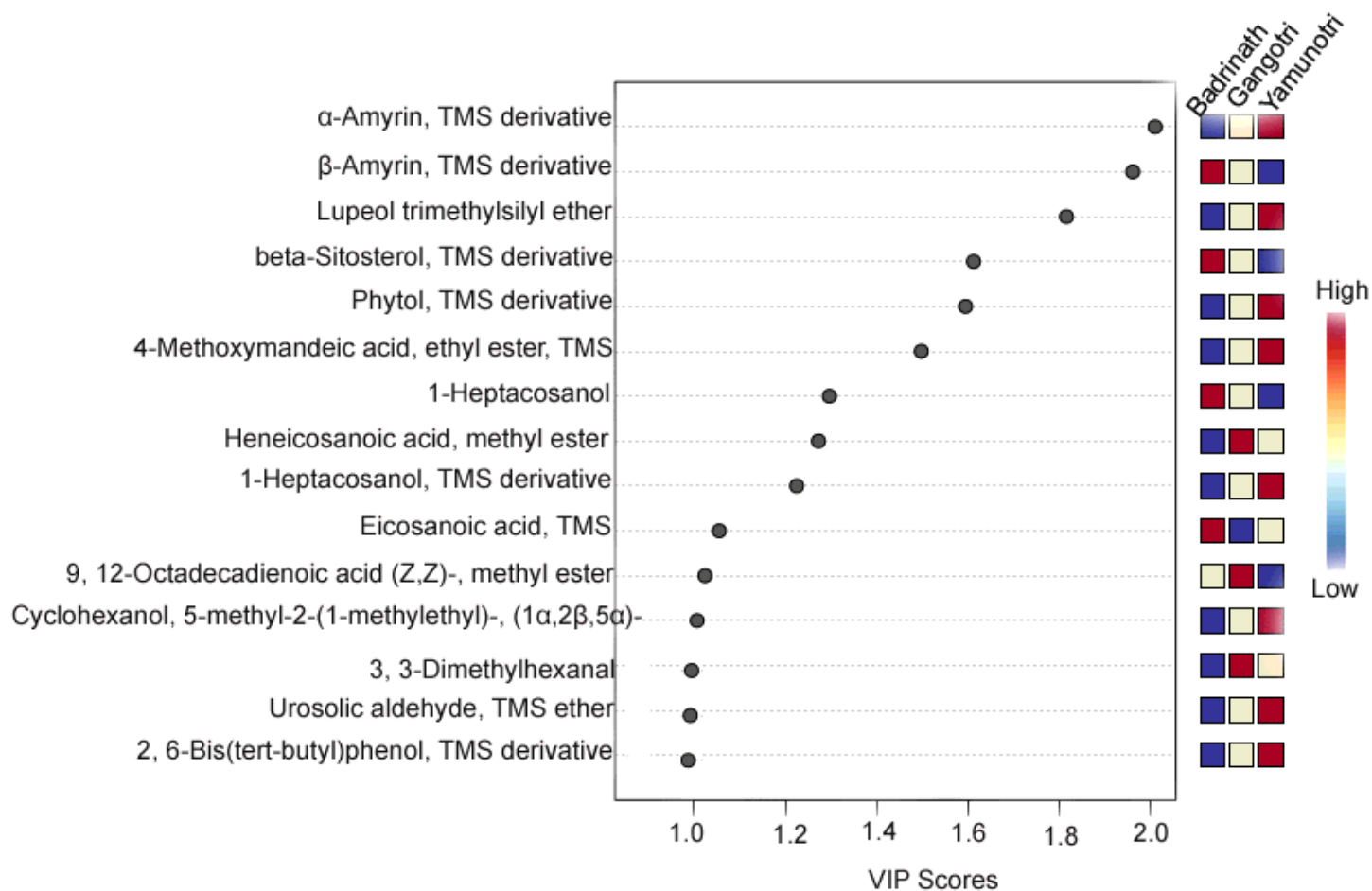


Figure 5

GC-MS based PLS-DA-VIP plot identifies important features. The expression profile of each class of male and female samples' metabolites was given on the right side of the colored box

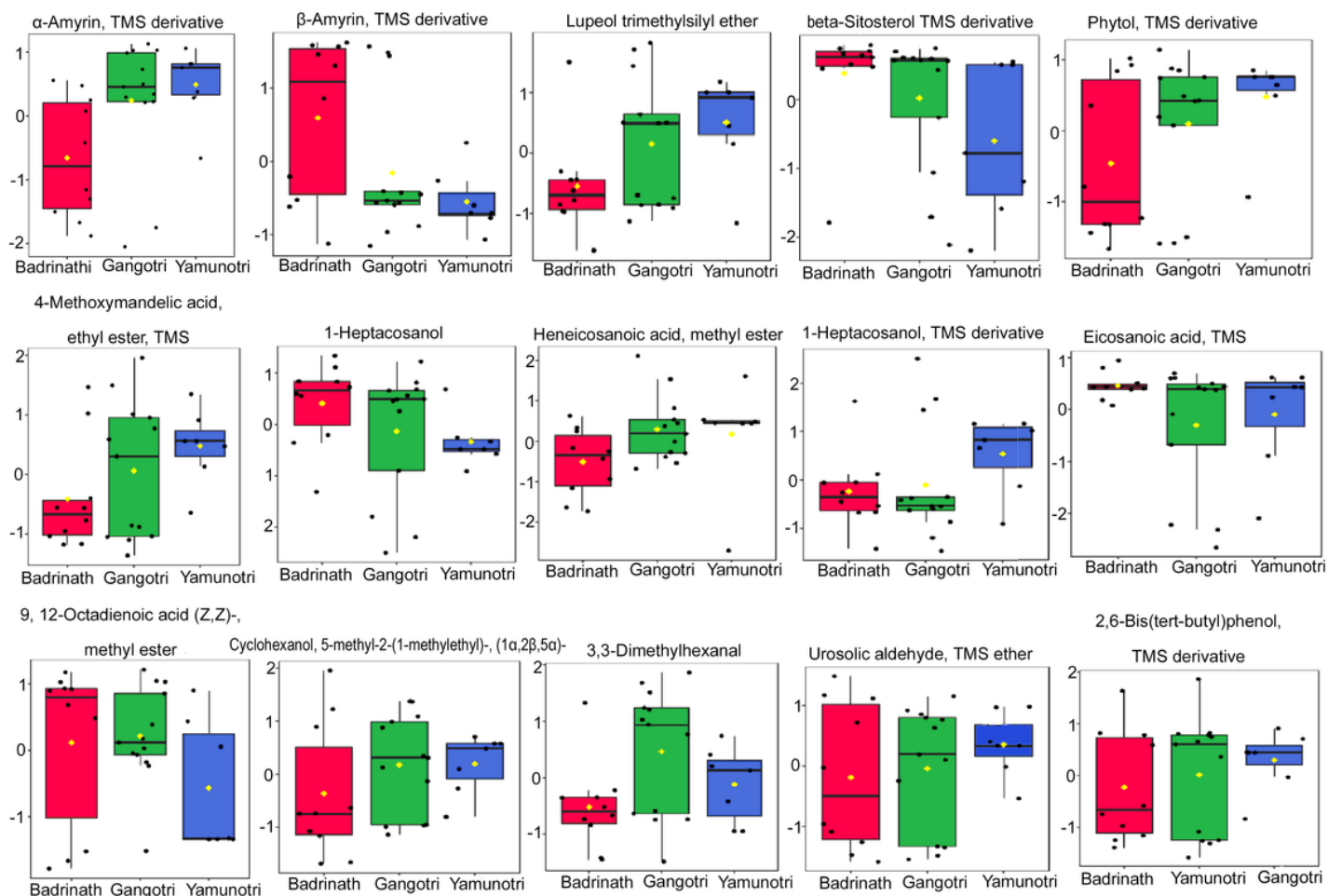


Figure 6

Box plots with whiskers for discriminant metabolites identified by GC-MS profiling for differentiation of seabuckthorn (*H. salicifolia*) berry samples from various locations of Uttarakhand. The name of the berry sample was represented in horizontal axis, whereas the logarithmic scale (concentration/200 mg) of different metabolites were represented in the vertical axis

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementarymaterialGenetResourCropEvol.docx](#)