Phytochemistry 173 (2020) 112292

Contents lists available at ScienceDirect

Phytochemistry

journal homepage: www.elsevier.com/locate/phytochem

Assessing specialized metabolite diversity of *Alnus* species by a digitized LC–MS/MS data analysis workflow

Kyo Bin Kang[a,b,*], Sunmin Woo[c], Madeleine Ernst[b,d], Justin J.J. van der Hooft[b,e], Louis-Félix Nothias[b], Ricardo R. da Silva[f], Pieter C. Dorrestein[b], Sang Hyun Sung[c,1], Mina Lee[g,**]

[a] Research Institute of Pharmaceutical Sciences, College of Pharmacy, Sookmyung Women's University, Seoul, Republic of Korea
[b] Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA, USA
[c] College of Pharmacy and Research Institute of Pharmaceutical Sciences, Seoul National University, Seoul, Republic of Korea
[d] Center for Newborn Screening, Department of Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark
[e] Bioinformatics Group, Department of Plant Sciences, Wageningen University, Wageningen, the Netherlands
[f] NPPNS, Physic and Chemistry Department, School of Pharmaceutical Sciences of Ribeirão Preto, University of São Paulo, Ribeirão Preto, Brazil
[g] College of Pharmacy, Sunchon National University, Sunchon, Republic of Korea

ARTICLE INFO

Keywords:
Alnus spp.
Betulaceae
Chemical diversity
Specialized metabolites
LC–MS/MS
Data analysis

ABSTRACT

Alnus spp. (Betulaceae) have been used for treatments of hemorrhage, burn injuries, antipyretic fever, diarrhea, and alcoholism in traditional medicines. In this study, a digitized LC–MS/MS data analysis workflow was applied to provide an overview on chemical diversity of 15 *Alnus* extracts prepared from bark, twigs, leaves, and fruits of *A. japonica*, *A. firma*, *A. hirsuta*, and *A. hirsuta* var. *sibirica*. Most of the MS/MS spectra could be putatively annotated based on library matching, *in silico* fragmentation, and substructural topic modeling. The putative annotation allowed us to discriminate the extracts into three chemotypes based on dominant chemical scaffolds: diarylheptanoids, flavonoids or tannins. This high-throughput chemical annotation was correlated with α-glucosidase inhibition data of extracts, and it allowed us to identify gallic acid as the major active compound of *A. firma*.

1. Introduction

The genus *Alnus* (Betulaceae), also known as alder, comprises about 40 species of monoecious trees and shrubs which are mainly distributed throughout the north temperate zone. In East Asian traditional medicines, *Alnus* species have been used for treatments of hemorrhage, burn injuries, antipyretic fever, diarrhea, and alcoholism (Park et al., 2010). Recent pharmacological studies reported anti-inflammatory, anti-tumor, anti-obesity, and antioxidative effects of *Alnus* species extracts (Sati et al., 2011). These pharmacological activities of *Alnus* species are attributed to various specialized metabolites of these plants. Different types of specialized metabolites, including flavonoids, triterpenoids, tannins, phenolics, and diarylheptanoids have been isolated from *Alnus* species (Ren et al., 2017). Many isolated molecules are reported to have various biological activities such as anti-inflammation (, Lee et al., 2017a, b), antioxidative effect (Kim et al., 2016), anti-adipogenesis (Sung and Lee, 2015), cytotoxicity (Novakovic et al., 2017), and DNA methylation (Krasilnikova et al., 2018), which makes *Alnus* species possible sources for natural drug discovery. Despite the pharmacological potentials of specialized metabolites of this genus, interspecific chemical diversity between different *Alnus* species has not yet been systematically investigated. Here, we aimed to establish a liquid chromatography coupled to tandem mass spectrometry (LC–MS/MS)-based analytical method for a comprehensive characterization of different classes of phytochemical constituents biosynthesized by *Alnus* species.

LC–MS/MS is the key technology of modern natural product chemistry and metabolomic studies to explore chemical space (Allard et al., 2018; Wolfender et al., 2019). Many untargeted metabolomics studies using LC–MS/MS, especially in the automated Data Dependent Acquisition (DDA) method, provide information on intra- and inter-specific differences on chemical composition of medicinal plants (Kang et al., 2018; Negrin et al., 2019; Robertson et al., 2018; Zhou et al., 2019). However, in most of these studies, interpretation of MS/MS spectra were performed manually, based on experts' knowledge about

* Corresponding author. (K. B. Kang).
** Corresponding author. (M. Lee).
E-mail addresses: kbkang@sookmyung.ac.kr (K.B. Kang), minalee@sunchon.ac.kr (M. Lee).
[1] Deceased on July 24th, 2018.

https://doi.org/10.1016/j.phytochem.2020.112292
Received 23 September 2019; Received in revised form 31 January 2020; Accepted 1 February 2020

phytochemical distribution and fragmentation patterns of certain specialized metabolite classes. This way of data curation is low-throughput, laborious, and time consuming, and furthermore has the limitation that only a few selected spectra (generally biased by ion intensity in chromatograms) can be annotated among thousands of collected MS/MS spectra. Recently, several computational data analysis methods have been developed for MS/MS spectral annotation to overcome this limitation. Public and proprietary MS/MS spectral repositories and databases such as MassBank (Horai et al., 2010), ReSpect (Sawada et al., 2012), NIST (The National Institute of Standards and Technology; http://www.nist.gov/srd/nist1a.cfm) and Global Natural Product Social Molecular Networking (GNPS) (Wang et al., 2016) enable users to compare their own spectra with reference spectra for spectral annotation. MS/MS molecular networking clusters and organizes similar spectra as a network, which allows propagation of molecular annotation to unknown molecules (Wang et al., 2016; Watrous et al., 2012). *In silico* fragmentation predictors and combinatorial fragmentators (Allen et al., 2014; da Silva et al., 2018; Dührkop et al., 2015; Ruttkies et al., 2016) expand the primary annotation to molecules without any reference spectra deposited in public databases.

Previously we demonstrated that a workflow combining several computational MS/MS data analysis tools can be a powerful method for phenotyping plant specialized metabolites (Ernst et al., 2019a, 2019b; Kang et al., 2019). In this study, we applied this digitized MS/MS-based molecular annotation method to analyze the specialized metabolites of four *Alnus* taxa occurring in Korea, *A. japonica*, *A. firma*, *A. hirsuta*, and *A. hirsuta* var. *sibirica*.[2] Phytochemical profiles of bark, twigs, leaves, and fruits of four *Alnus* species were analyzed to reveal not only interspecific diversity of specialized metabolites, but also metabolite localization patterns. In addition, we integrated the digitized chemical space information of *Alnus* extracts with its biological activity, demonstrating that this workflow is an efficient dereplication method for annotation of bioactive compounds from complex phytochemical mixtures.

2. Results and discussion

2.1. MS/MS analysis and annotation of Alnus specialized metabolites

The LC–MS/MS analysis revealed that the 15 *Alnus* extracts are different in their specialized metabolite contents, in both a qualitative and quantitative manner (Fig. 1). As a first step of the digitized data analysis, 531 mass features were extracted from the entire dataset by MZmine2-based preprocessing, then the feature table, metadata, and extracted MS/MS spectral.mgf files were uploaded to the GNPS MS/MS molecular networking workflow (Nothias et al., 2019). MS/MS molecular networking organized them into a network consisting of 33 molecular families (two or more connected nodes of a graph (Nguyen et al., 2013)) and 268 singletons (nodes not having any molecular relatives). Two types of metadata related to spectral sources, plant species and plant parts, were visualized upon the molecular network. Both species-mapping (Fig. S1, Supplementary Data) and plant parts-mapping (Fig. 2) visualized that many metabolites in certain molecular families are constrained to specific species plant parts. For example, spectral nodes in molecular family **B** are mainly found in fruits, while other metabolites in molecular family **D** are from bark and twigs and spectra in molecular family **E** are predominantly observed in leaves. Metabolites in molecular family **F** was only found in *A. firma*. The largest molecular family, **A**, could be divided into two subclusters based on distribution (Fig. 2). This finding reveals the localization of closely related yet different chemical structures in different plant parts.

The MS/MS spectral library search through GNPS resulted in 47 hits to reference MS/MS spectra, which are level 2 annotations according to the 2007 Metabolomics Standards Initiative (MSI) (Sumner et al., 2007). Nine previously isolated and purified diarylheptanoids, platyphylloside (**1**; numbers mean spectra indices in the MS/MS molecular network), aceroside VII (**63**), aceroside VIII (**195**), oregonin (**202**), rubranoside B (**206**), rubranoside A (**214**), rubranoside D (**240**), (5*S*)–*O*-methylhirstanonol (**241**), and oregonoyl A (**246**) were used to confirm annotations and give level 1 annotations by matching retention time and MS/MS spectra (Lee et al., 2013, 2010; Sung and Lee, 2015). Retention time and spectra of these compounds can be found in Result S1, Supplementary Data.

To maximize the annotation coverage upon the entire dataset, we applied MolNetEnhancer, a recently developed computational workflow for MS/MS-based untargeted metabolomics (Ernst et al., 2019a). Based on the GNPS library matching and *in silico* annotation derived from Network Annotation Propagation (NAP) (da Silva et al., 2018), most of the molecular families could be annotated for their unique chemical classes, which are MSI level 3 annotations. Computational class annotations made by MolNetEnhancer was double-checked manually in order to prevent false annotations. Most *in silico* predicted annotations were reasonable, but some molecular families showed incorrect annotations. For example, the whole molecular family **A** was annotated as triterpenoids by MolNetEnhancer, but manual inspection on molecular annotations of each spectral node suggested that two subclusters within **A** should be annotated as diarylheptanoids and triterpenoids, respectively (Fig. S2, Supplementary Data). Within the molecular network, class annotation elucidated the localization patterns of different classes. For example, molecular families **A**, **D**, and **I**, which are present in bark and twigs, were annotated as diarylheptanoids and their glycosides. Fruits of *Alnus* species showed high contents of ellagitannins (**B**), while the leaves were abundant in flavonoid glycosides (**E**).

As described in previous studies (Ernst et al., 2019a; Kang et al., 2019), the MolNetEnhacer workflow takes advantage of MS2LDA, which provides information about substructural diversity within same classes of metabolites (van der Hooft et al., 2016). MS2LDA extracts patterns of fragment ions or neutral losses which are observed together in multiple spectra, called Mass2Motifs, which can show how the compounds in same chemical classes are different in their substructure. For example, Mass2Motifs 41, 49, 72, and 81 were extracted from MS/MS spectra which are clustered as molecular families **A**, **D**, and **I**, which are annotated as diarylheptanoids. These Mass2Motifs were annotated to represent diarylheptanoid scaffolds which are different in the pattern of unsaturation and hydroxylation. As shown in Fig. 3, Mass2Motif 72 represents the presence of a fragment ion *m/z* 331.1525 ($[C_{19}H_{23}O_5]^-$) while Mass2Motif 81 represents the presence of a fragment ion *m/z* 299.1625 ($[C_{19}H_{23}O_3]^-$), and these two Mass2Motifs are observed in MS/MS spectra **214** and **63**, respectively. Mass2Motif 3 represents the neural loss of *m/z* 162.0525, which is caused by the loss of a hexose moiety, often related to glucose, and commonly observed in plant metabolomics data. GNPS spectral library matching annotated these spectra as rubranoside A (**214**) and aceroside VII (**63**). Based on these annotations, Mass2Motifs 72 and 81 could be annotated as rubranol- and centrolobol-related motifs, respectively. Similarly, Mass2Motifs 41 and 49 were annotated as hirsutanonol- and platyphyllonol-related motifs, mainly based on spectral annotation of oregonin (**202**) and platyphylloside (**1**) (Fig. 3). These Mass2Motifs also represent a few fragment ions with smaller *m/z* values; and we confirmed that these fragment ions are identical to the previously reported characteristic fragment ions generated from each diarylheptanoid aglycone. MS/MS

[2] *A. hirsuta* var. *sibirica* is currently classified as a synonym of *A. hirsuta* in the Plant List (http://www.theplantlist.org). However, many local botanists are considering *A. hirsuta* var. *sibirica* as a variety based on their different morphology (Chang et al., 2005), and our study suggests that these two plants differ in their metabolite composition. Unfortunately there is no reported genome sequence of *A. hirsuta* var. *sibirica* which is required to confirm its taxonomy; so we concluded that this is unconcluded and kept it to be described as a genetic variety.

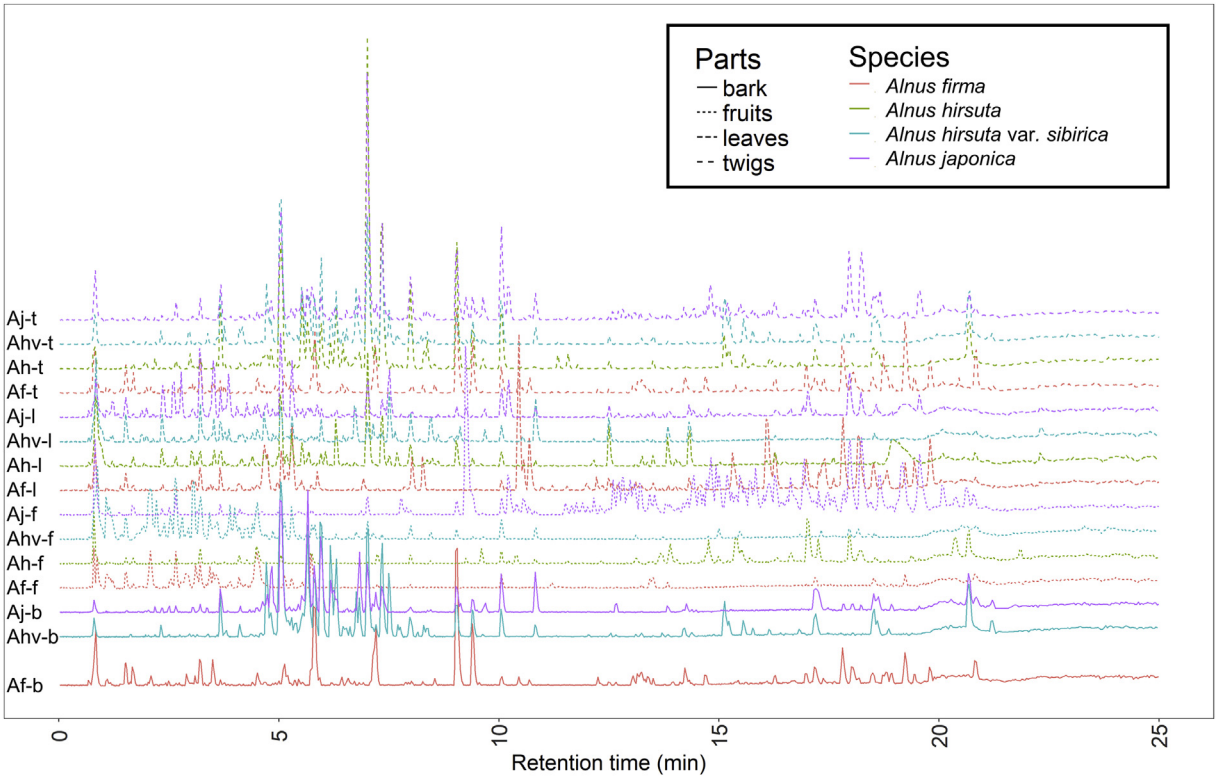

Fig. 1. LC–MS base peak ion (BPI) chromatograms of 15 *Alnus* extracts. Gaps between chromatogram were added to visualize their difference, so y-axis values do not equal to the absolute intensities.

fragmentation pathways of diarylheptanoids under ESI negative ion mode were well established by previous studies (Riethmüller et al., 2015, 2013), and fragment ions represented by Mass2Motifs 41 and 49 agree with the fragmentation patterns described in these studies. Thus, we could demonstrate the potential of MS2LDA, which facilitates the structural annotation of substructures and the storing of these annotations within MS2LDA experiments. These annotated Mass2Motifs can now also be stored in MotifDB (Rogers et al., 2019). In this way, expert knowledge is made available for future substructure annotations based on MS/MS data. By mapping the distribution of Mass2Motifs on to the molecular network, we could visualize the substructural differences within molecular families (Fig. 3), which is helpful during annotation and identification of MS/MS spectra within a molecular network.

2.2. Chemotype discrimination of Alnus extracts

Based on the molecular network and structural annotation, chemical diversity between *Alnus* extracts were systematically analyzed. As a first step, the dissimilarity between samples was calculated. In most metabolomics studies, the diversity between samples has been analyzed by using multivariate analysis techniques such as principal component analysis (PCA) (Worley and Powers, 2012), which is based on the Euclidean distance metric or principal coordinate analysis (PCoA) which can be based on different distance metrics such as for example the Bray-Curtis dissimilarity (Brückner and Heethoff, 2017). However, these conventional methods consider each feature as independent entities, ignoring the structural relationship between molecules. To reflect the chemical similarity at the scaffold level shared by samples from the same plant parts, we applied the Chemical Structural and Compositional Similarity (CSCS) metric (Brejnrod et al., 2019; Sedio et al., 2017), which accounts for the chemical structural similarity across metabolites by integrating the MS/MS spectral similarity calculated in the process of GNPS molecular networking.

Compared to the PCA score plot (Fig. 4C) or hierarchical cluster analysis using the Euclidean distance (Fig. 4D), the weighted (by intensity of MS1 ion intensities) CSCS metric showed clearer discriminative patterns both in a Principle Coordinates Analysis (PCoA) (Fig. 4A) as well as the chemical dendrogram (Fig. 4B); especially the chemodendrogram revealed that the samples can be discriminated as three chemotypes. By mapping these chemotype-classes on the molecular network, the major differences in their metabolites were easily visualized (Fig. 4E). Chemogroup 3 (Ah-L, Ahv-L, Af-L, Af-T, and Af-B) showed high contents of the molecular family **E**, which was annotated as flavonoids, while Chemogroup 2 (Aj-L, Ah-F, Af-F, and Ahv-F) was rich in tannins. On the other hand, diarylheptanoids were majorly represented in the remaining samples (Chemogroup 1). This pattern partially agreed with the plant parts-based distribution (Fig. 2), but it could be revealed that there are some exceptional cases such as Aj-F (rich in diarylheptanoids while the other fruits are abundant in tannins), Aj-L (rich in tannins while the other leaves are abundant in flavonoids), and Af-T and Af-B (rich in flavonoids while the other bark and twigs are abundant in diarylheptanoids). It is well known that plant specialized metabolite profiles vary based on a number of biotic as well as abiotic factors, such as diurnal changes, presence of herbivory or plant symbionts, nutrient availability and exposure to sunlight (Bednarek and Osbourn, 2009; Wink, 2010). We were not able to explain the biological context of these exceptional localizations in *A. firma* and *A. japonica* as metadata collected did not allow for a precise evaluation of contributing factors to the differences of the metabolic profiles observed here; however, differences in chemical profiles revealed here will provide a helpful guidance for further phytochemical studies on *Alnus* species.

2.3. Identification of α-glucosidase inhibitory compounds from Alnus firma

A previous study reported that H_2O or MeOH extracts of *A. firma* leaves exhibited inhibitory activity against α-glucosidase (Yu et al., 2007). However, none of the compounds isolated in this study showed

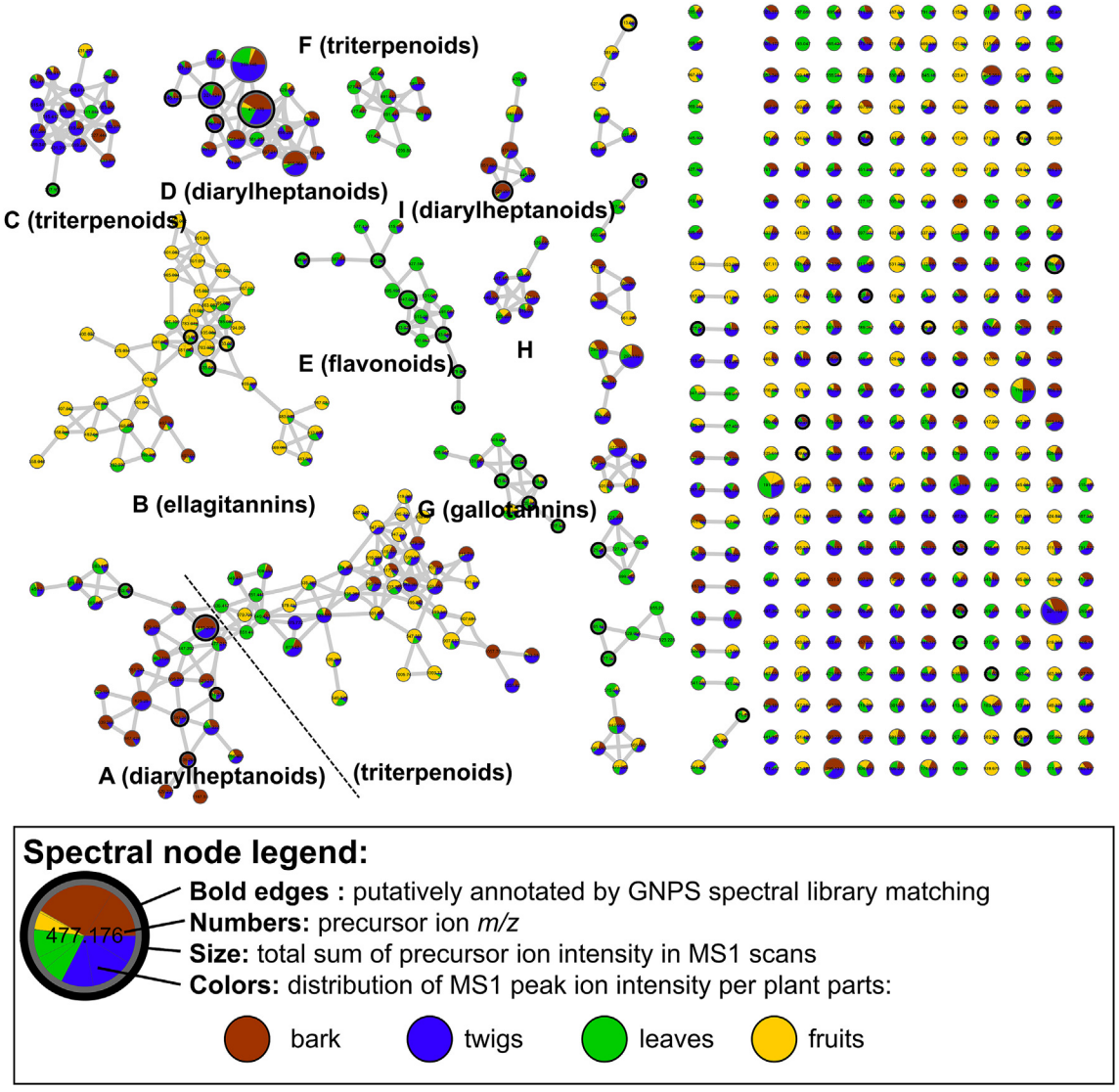

Fig. 2. The MS/MS spectral network of specialized metabolites contained in the bark, twigs, leaves, and fruits of *A. japonica*, *A. firma*, *A. hirsuta*, and *A. hirsuta* var. *sibirica*. Spectral nodes are colored according to the mean precursor ion intensity per different plant parts: bark, twigs, leaves, and fruits. Molecular families **A–I** are highlighted.

α-glucosidase inhibitory activity, and other metabolites from *Alnus* have not been reported to exhibit α-glucosidase inhibition, except three cyclic diarylheptanoids isolated from *A. sieboldiana* (Chiba et al., 2013). We hypothesized that the chemical information derived from MS/MS-based dereplication in this study could be used to reveal which compounds contribute to the α-glucosidase inhibitory activity of *A. firma*, and tried to associate the annotations to bioactive extracts. 15 *Alnus* extracts were evaluated for their α-glucosidase inhibitory activity as shown in Table 1. As previously reported, the leaf extract of *A. firma* exhibited potent inhibitory activity showing IC_{50} values of 12.29 μg/mL. Extracts of *A. firma* bark, twigs, and fruits, and *A. hirsuta* var. *sibirica* fruits also showed potent activity with IC_{50} ranging from 6.80 to 8.48 μg/mL. Extract of *A. japonica* leaves and *A. hirsuta* var. *sibirica* leaves exhibited moderate inhibitory activity with IC_{50} values of 23.36 and 29.84 μg/mL, while other extracts showed very weak or no inhibitory activity against α-glucosidase.

To bridge the bioactivity and chemical information, we applied the workflow named "bioactive molecular networking", which was developed in our previous study (Nothias et al., 2018). In this workflow, the Pearson correlation coefficients (r) and their significances (p values) between the semi-quantitative intensities of MS/MS features and the bioactivity value are calculated. By plotting r and p values on the molecular network, it was easily visualized that MS/MS spectra showing negative r values (negative correlation with IC_{50}; meaning possibly inhibitory against α-glucosidase) with significance ($p < 0.05$) are mainly clustered within the molecular families **B** (ellagitannins) and **G** (gallotannins) (Fig. S3, Supplementary Data). In order to narrow down the candidates from chemical classes to single compounds, we filtered the candidate list using the Bonferroni correction for multiple hypothesis testing (Nothias et al., 2018). As a result, only one MS/MS feature (**10**) was left as a possible contributor to the bioactivity showing an r value of -0.8625 (IC_{50} values were applied as bioactivity index, so negative r values mean higher contribution to bioactivity) and a p value of 3.52×10^{-5}. The MS/MS feature **10** was annotated as gallic acid by spectral matching, and the identification was confirmed by standard compound injection. We inspected the MS/MS feature table and confirmed that active extracts (Af-B, Af-T, Af-L, Af-F, Aj-L, Ahv-L, and Ahv-F; ion intensity for **10** ranging from 49,854 to 96,211) show a higher intensity of the gallic acid ion than inactive extracts (Aj-B, Aj-T, Aj-F, Ah-L, Ah-F, Ah-B, and Ahv-T; ion intensity for **10** ranging from 520 to 11,358) as shown in Fig. 5. A previous study reported gallic acid as a potent α-glucosidase inhibitor (Wansi et al., 2007), so it could be

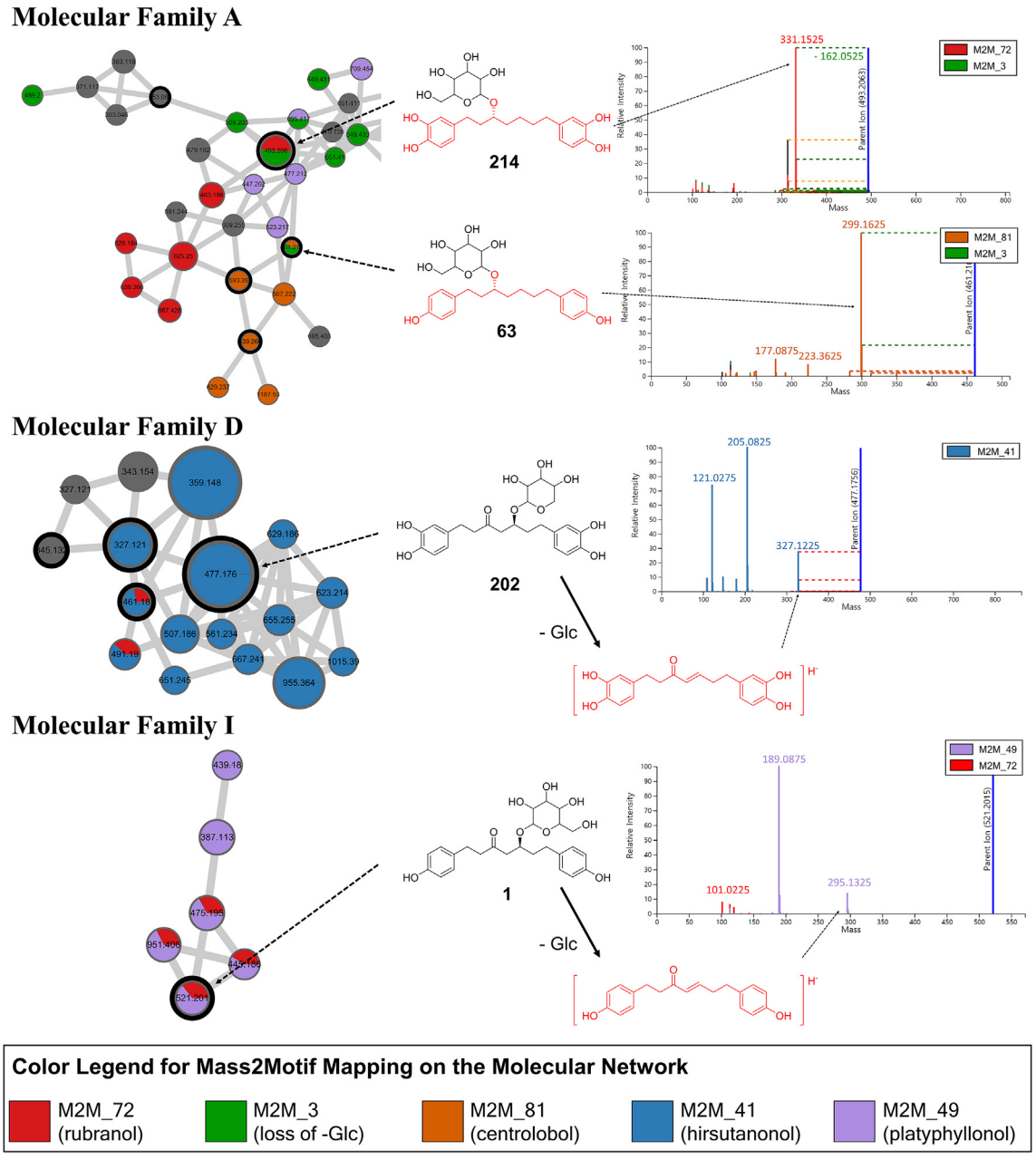

Fig. 3. MS2LDA-driven substructural annotation of diarylheptanoids of *Alnus* species. Integrated with GNPS library matching and NAP *in silico* annotation, diarylheptanoid-related Mass2Motifs 41, 49, 72, and 81 could be characterized and correlated with specific substructures of diarylheptanoid aglycones. Scaffold diversity within diarylheptanoid molecular families **A**, **D**, and **I** were revealed by mapping these Mass2Motifs on the molecular network with different colors. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

hypothesized that gallic acid is the main contributor of α-glucosidase inhibition of these *Alnus* extracts. Although this result did not provoke a discovery of a previously unknown bioactive compound, it demonstrated that the digitized MS/MS-based dereplication strategy can reveal bioactive components from complex phytochemical extracts and reduce the unnecessary effort spent in re-isolation of previously known bioactive compounds.

3. Conclusions

The present study shows that the digitized MS/MS data analysis workflow strongly enhances phytochemical investigation. Commonly in Nature, conservation of specialized metabolism-related genes leads to chemical similarity within specific taxa, and this has been an essential background of chemotaxonomy in the plant kingdom. Our workflow is especially advantageous in analyzing chemical similarity of specialized metabolites; MS/MS molecular networking allows the visualization of the similarity pattern for all metabolites detected in a sample. The CSCS metric takes advantage of spectral similarity, and our result shows that it is especially useful to reveal the chemotypes of different plant species/organs based on the distributional pattern of different molecular scaffolds.

It was previously known that diarylheptanoids, flavonoids, and tannins are abundant in *Alnus* species, but their localization among different species and plant parts has not been investigated before. Here, we found that diarylheptanoids are abundant in bark and twigs, while tannins are accumulated in fruits and flavonoids in leaves. Interestingly there were also some exceptional cases shown in *A. japonica* and *A. firma*. However, we could not conclude here that this is a very general phenomenon in *Alnus* species, because not enough biologically

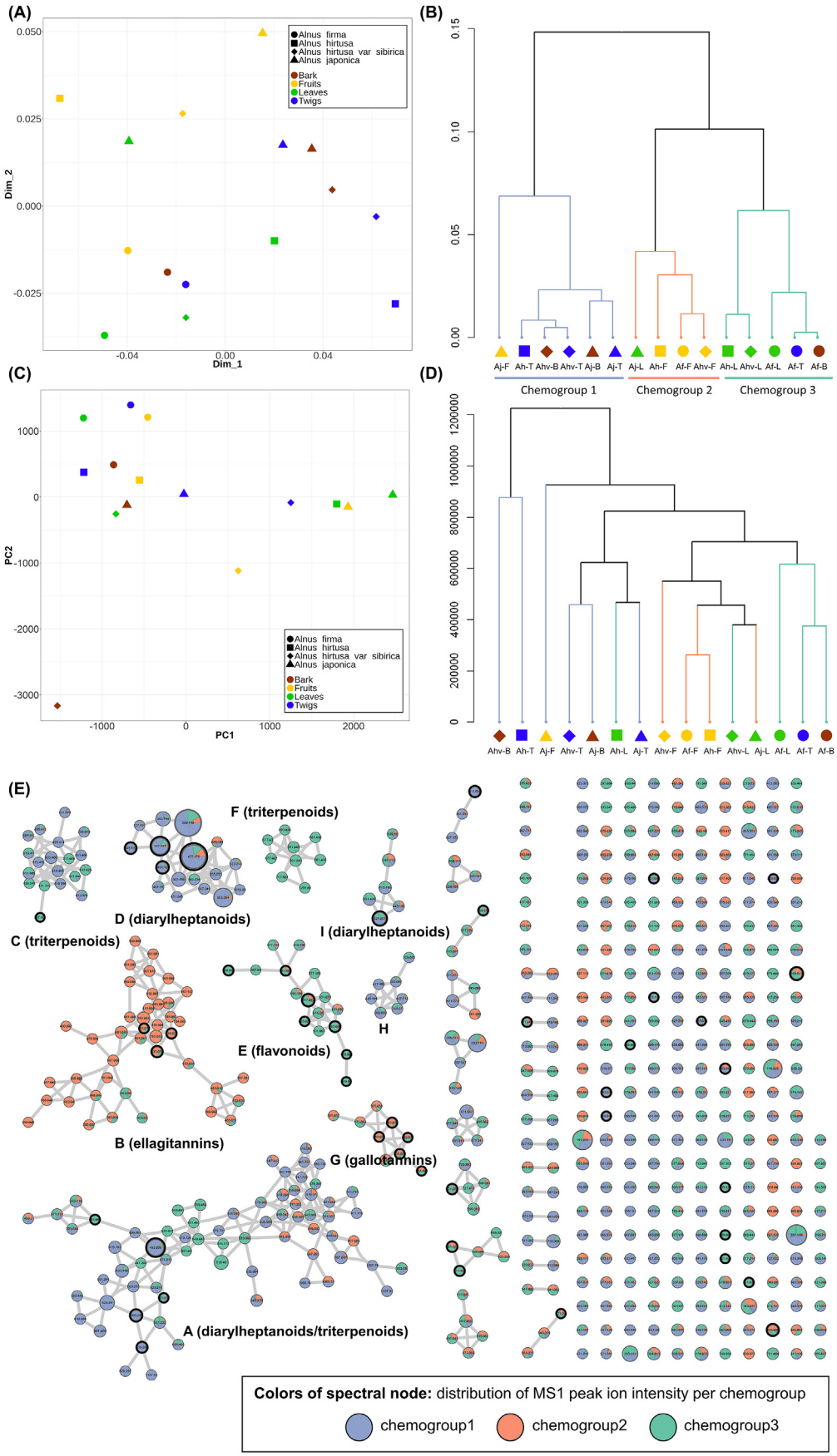

Fig. 4. Discrimination of the analyzed Alnus extracts into chemogroups. The analyzed extracts can be discriminated into three chemogroups by visualizing the CSCS distance metric between samples as PCoA plot (A) and chemical dendrogram (B). On the other hand, conventional methods such as PCA score plot (C) or hierarchical clustering analysis (HCA) using the Euclidean distance (D; chemogroups 1–3 are visualized with same colors used in B to make it easy to be compared) could not discriminate the samples into the same chemotypes. By mapping the chemogrouping of samples on the molecular network, it could be visualized that the three chemogroups were rich in diarylheptanoid, flavonoid, and tannins, respectively (E). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1
α-Glucosidase inhibitory activity of tested *Alnus* extracts.

Samples		IC_{50} (μg/mL)	95% confidence intervals
A. firma	barks	8.48	6.07–11.84
	twigs	7.47	5.57–10.01
	leaves	12.29	7.32–20.65
	fruits	6.80	4.61–10.02
A. japonica	barks	> 100	14.37–37.96
	twigs	> 100	
	leaves	23.36	
	fruits	n.d.	
A. hirsuta	barks	69.55	30.36–159.30
	twigs	> 100	
	leaves	> 100	
	fruits	> 100	
A. hirsuta var. *sibirica*	twigs	> 100	
	leaves	29.84	10.86–82.00
	fruits	7.33	3.68–14.60
acarbose[a]		513.50 μM	432.3–746.1 μM

[a] Positive control.

replicated samples were available for this study. We also could not provide any hypothetical explanation, which was an illustration of the lack of our knowledge on the metabolism of non-model species. Nevertheless, the discrimination of chemotypes and estimation on the bioactive compounds shows the potential of high-throughput metabolite annotation tools for natural product chemistry researchers. We expect that the chemical overview on *Alnus* species made in this study will provide valuable insights for further exploration and research on these and other medicinal plants.

4. Experimental

4.1. Chemicals and reagents

HPLC grade water and acetonitrile (MeCN) were purchased from Avantor Performance Materials, Inc. (Center Valley, PA, USA). Formic acid and leucine-enkephalin were purchased from Sigma-Aldrich (St. Louis, MO, USA). Ultrapure water was triple deionized (Millipore, Bedford, MA, USA). MeOH was purchased from Daejung Chemicals Co., Ltd. (Siheung, Korea).

4.2. Plant material

The bark, twigs, leaves, and fruits of *Alnus japonica* (Thunb.) Steud., *Alnus hirsuta* (Spach) Rupr., and *Alnus hirsuta* var. *sibirica* (Spach) C.K.Schneid. (Betulaceae) were collected in the Nambu forest of Seoul National University, Baegwoon Mountain, Gwangyang, Korea (GPS N35°06′, E127°37′) in March 2008. The bark, twigs, leaves, and fruits of *Alnus firma* Siebold & Zucc. (Betulaceae) were obtained from SK forest, Chungju, Korea (GPS N37°03′, E128°00′) in March 2008. These samples were authenticated by Prof. Jong Hee Park (Pusan National University, Pusan, Korea). The voucher specimens (SNU 751–765) were deposited at the Herbarium of the Medicinal Plant Garden, College of Pharmacy, Seoul National University, Koyang, Korea.

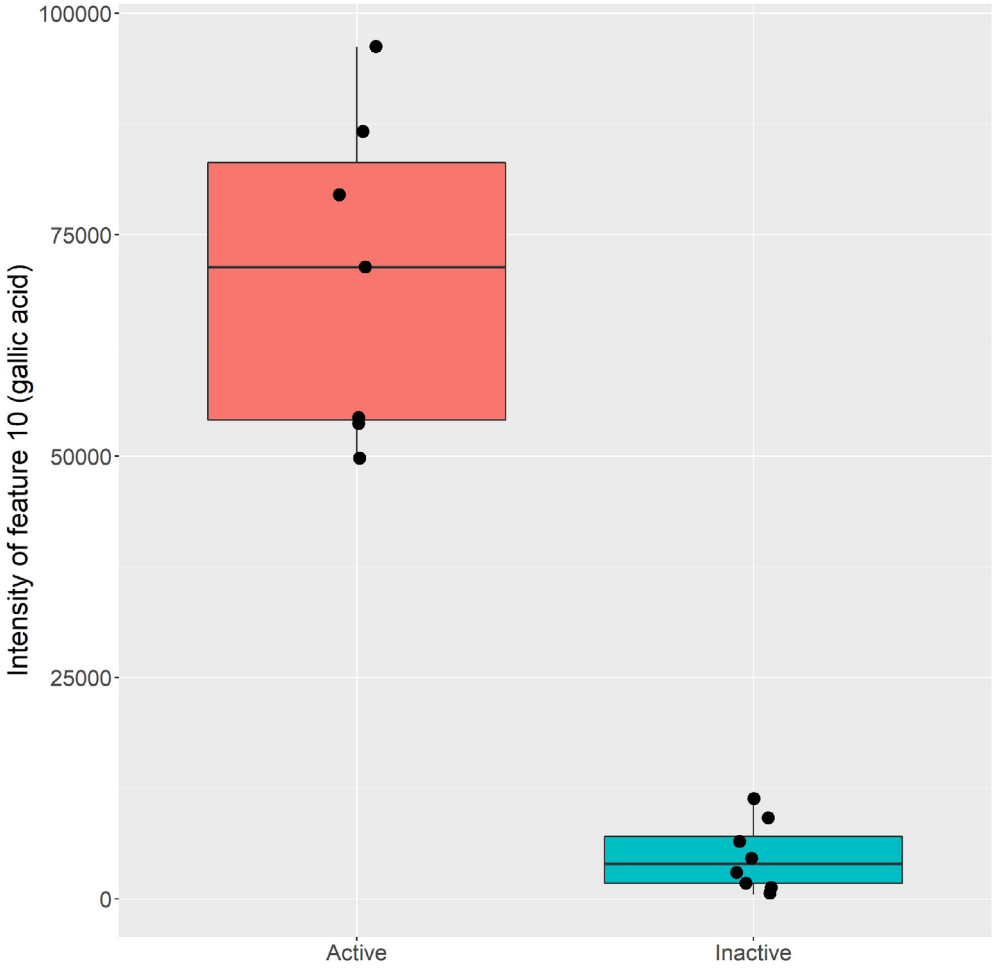

Fig. 5. A boxplot showing the ion intensities of MS/MS feature **10** (gallic acid) in extracts which were active (IC_{50} < 30 μg/mL) and inactive against α-glucosidase.

4.3. Sample preparation

The accurately weighed 100.0 mg of powdered samples were extracted with 1.0 mL MeOH, sonicated at room temperature for 2 h, centrifuged at 10,000 g for 5 min, and 0.5 mL supernatant was collected. The collected supernatant was filtered with a PVDF filter for LC–MS/MS analysis. A Branson 8510 ultrasonic bath (Branson Ultrasonics Corporation, Danbury, CT, USA) was used for extraction. Centrifugation was performed using HANIL micro centrifuge (Micro17TR, Hanil scientific industrial, Seoul, Korea).

4.4. LC–MS/MS data acquisition

LC–MS/MS analyses were performed on a Waters Acquity UPLC system (Waters Co., Milford, MA, USA) coupled to a Waters Xevo G2 QTOF mass spectrometer (Waters MS Technologies, Manchester, UK) which was equipped with an electrospray ionization interface (ESI). Chromatographic separations were performed on a Waters Acquity BEH C_{18} (100 × 2.1 mm 1.7 μm) column. The mobile phase comprised H_2O (A) and MeCN (B), both of which were acidified with 0.1% formic acid. The column temperature and sample organizer were maintained at 40 °C and 15 °C, respectively. A stepwise gradient method at constant flow rate of 0.3 mL/min was used to elute the column with the following conditions: 10–100% B (0–27 min), followed by 3 min of washing and 5 min of reconditioning. Analyses of the samples (1.0 μL injected) were performed in an optimized DDA mode consisting of a full MS survey scan in the *m/z* 100–1500 Da range (scan time: 100 ms) followed by MS/MS scans for the three most intense ions. The collision energy was applied at a gradient from 20 to 80 V. The ESI conditions were set as follows: capillary voltage 3.0 kV, cone voltage 25 V, source temperature 120 °C, desolvation temperature 450 °C, cone gas flow 50 L/h, and desolvation gas flow 800 L/h. High-purity nitrogen was used as the nebulizer and auxiliary gas, and argon was used as the collision gas. Considering the abundance of phenolic compounds in *Alnus* species (Sati et al., 2011; Liigand et al., 2017), negative ion mode (ESI−) was selected for analyses. The $[M - H]^-$ ion of leucine enkephalin at *m/z* 554.2615 was used as the lock mass to ensure mass accuracy and reproducibility.

4.5. LC–MS/MS data analyses

4.5.1. Chromatographic and spectral preprocessing using MZmine

Raw LC−MS/MS data files were imported into MZmine 2.33 (Pluskal et al., 2010). The mass detection was performed with the noise level at 200 (for MS scans) and 20 (for MS/MS scans). MS chromatograms were built with ions showing a minimum time span of 0.02 min, minimum height of 5000, and *m/z* tolerance of 0.002 (or 5.0 ppm), then missing data points were filled by using the peak extender module with minimum height of 1000. The chromatographic deconvolution was achieved by the baseline cutoff algorithm, with the following parameters: minimum peak height of 1000, peak duration range of 0.02–5.00 min, and baseline level of 330. Chromatograms were deisotoped using the isotopic peaks grouper algorithm with an *m/z* tolerance of 0.002 (or 10.0 ppm) and a t_R tolerance of 0.1 min, and then were aligned together into a peak table by the join aligner module [*m/z* tolerance at 0.006 (or 10.0 ppm), absolute t_R tolerance at 0.2 min, weight for *m/z* of 70, weight for t_R of 30]. Peaks without any MS/MS scan were removed by the GNPS filter module, then gap-filled with the peak finder module [intensity tolerance at 30.0%, *m/z* tolerance at 0.001 Da (or 5.0 ppm), absolute t_R tolerance of 0.2 min].

4.5.2. GNPS molecular networking

MS/MS molecular networking was performed using the GNPS web platform (https://gnps.ucsd.edu) (Wang et al., 2016). MS/MS spectra were window filtered by choosing only the top six peaks in the ± 50 Da window throughout the spectrum. A network was then created where edges were filtered to have a cosine score above 0.70 and more than four matched peaks. The spectra in the network were then searched against the spectral library of GNPS. The library spectra were filtered in the same manner as the input data. The molecular network was visualized using Cytoscape 3.7.1 (Shannon et al., 2003).

4.5.3. In silico annotation using NAP

The molecular network was used to compute the propagation of *in silico* annotation with Network Annotation Propagation (NAP (da Silva et al., 2018); https://proteomics2.ucsd.edu/ProteoSAFe/?params={%22workflow%22:%22NAP_CCMS2%22}) using parameters as follows: 10 first candidates, exact mass searches within 5 ppm, and structural databases of DNP (an in house version of Dictionary of Natural Products) and SuperNatural.

4.5.4. MS2LDA substructural topic modeling

Mass2Motifs were extracted from the pre-processed.mgf spectra file using the MS2LDA web platform (http://MS2LDA.org) (Wandy et al., 2018). Parameters for the MS2LDA experiment were set as follows: *m/z* tolerance 5.0 ppm, t_R tolerance 10.0 s, minimum MS1 intensity 0 au, minimum MS2 intensity 20.0 au, no duplicate filtering, number of iterations 1000, number of Mass2Motifs 100.

4.5.5. Integration of annotation data using MolNetEnhancer

In silico annotations by NAP and Mass2Motifs extracted by MS2LDA were integrated into the molecular network using MolNetEnhancer (Ernst et al., 2019a). The MolNetEnhancer workflow embedded in GNPS (https://gnps.ucsd.edu/ProteoSAFe/index.jsp?params=%7B%22workflow%22:%22MOLNETENHANCER%22%7D) was used to retrieve chemical class-level annotation for each molecular family, while the R package named rMolNetEnhancer (https://github.com/madeleineernst/RMolNetEnhancer) was used to integrate the distribution pattern of Mass2Motifs into the network.

4.5.6. Chemotyping of extracts based on chemical dissimilarity

The weighted CSCS distance matrix was calculated using the Python package pyCSCS (https://github.com/askerdb/pyCSCS), and Principal Coordinates Analysis (PCoA) was performed using the *cmdscale* function of the R stats package (R Core Development Team, 2013). Hierarchical clustering analysis (HCA) was performed using the complete-linkage clustering method and was visualized as a dendrogram.

4.6. α-Glucosidase inhibitory assay and data integration with MS/MS analysis

α-Glucosidase inhibitory activity of *Alnus* extracts were exhibited as described in our previous study (Lee et al., 2017a, b). Briefly, 90 mL of 100 mU/mL α-glucosidase (from *Saccharomyces cerevisiae*) solution in 0.05 M PBS buffer (pH 7.0) and 20 mL of *Alnus* extracts diluted with the PBS buffer were mixed and pre-incubated at 36 °C for 10 min in a 96-well microplate. 90 mL of 1.2 mM *p*-nitrophenyl-α-D-glucopyranoside (PNPG) solution was added to initiate the enzymatic reaction. The reaction mixture was incubated at 36 °C and monitored using a microplate reader (SpectraMax M5, Molecular Devices, CA, USA) at a wavelength of 405 nm. Absorbance of the mixture was measured every 2.5 min up to 5 min, which indicates the formation of *p*-nitrophenol. The reaction mixture with the test solution replaced by equivalent PBS buffer was used as control, and acarbose was used as the positive control. All samples were measured in triplicate. IC_{50} values were calculated using GraphPad Prism 5 (Graphpad Software, CA, USA) with a dose-inhibition curve with at least six appropriate concentrations.

For integration of bioassay data with MS/MS analysis data, an R-based Jupyter notebook (Kluyver et al., 2016) named Bioactive Molecular Networks, was downloaded from GitHub, https://github.com/DorresteinLaboratory/Bioactive_Molecular_Networks (Nothias et al., 2018). The spectral feature table generated from MZmine preprocessing

and IC_{50} values calculated in the α-glucosidase inhibitory assay were used in the data analysis workflow.

Acknowledgements

This research was supported by the National Research Foundation of Korea (NRF) grants funded by the Ministry of Science, ICT and Future Planning (NRF-2019R1F1A1058068 and NRF-2017R1C1B2005934). KBK was also supported by Sookmyung Women's University Research Grants (1-1803-2019) and Specialization Program Funding (SP1-201809-6). JJJvdH was supported by an Accelerating Scientific Discoveries grant from the Netherlands eScience Center [NLeSC] (ASDI. 2017.030). RRdS was supported by the São Paulo Research Foundation (Awards FAPESP 2017/18922–2 and 2019/05026–4).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.phytochem.2020.112292.

Data and software availability

All data are deposited in MassIVE (https://massive.ucsd.edu) accession number MSV000084202. The MS/MS molecular network, NAP annotation, MolNetEnhancer, and MS2LDA experiments can be browsed and downloaded with the following links: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=8c6f75c33ce449189236ffdf929e5589.

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task = 675a7abda69b498b883ffce2d2b076dc.

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task = e2c1125c2fa54d1aadac619492cbba24.

http://ms2lda.org/basicviz/summary/930/

Jupyter notebooks and data files used in this study are available on GitHub: https://github.com/KyobinKang/supplementary-Alnus.

References

Allard, P.M., Bisson, J., Azzollini, A., Pauli, G.F., Cordell, G.A., Wolfender, J.L., 2018. Pharmacognosy in the digital era: shifting to contextualized metabolomics. Curr. Opin. Biotechnol. 54, 57–64. https://doi.org/10.1016/j.copbio.2018.02.010.

Allen, F., Pon, A., Wilson, M., Greiner, R., Wishart, D., 2014. CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra. Nucleic Acids Res. 42, W94–W99. https://doi.org/10.1093/nar/gku436.

Bednarek, P., Osbourn, A., 2009. Plant-microbe interactions: chemical diversity in plant defense. Science 324, 746–748. https://doi.org/10.1126/science.1171661.

Brejnrod, A.D., Ernst, M., Dworzynski, P., Rasmussen, L.B., Dorrestein, P., van der Hooft, J., Arumugam, M., 2019. Implementations of the chemical structural and compositional similarity metric in R and Python. BioRxiv 546150. https://doi.org/10.1101/546150.

Brückner, A., Heethoff, M., 2017. A chemo-ecologists' practical guide to compositional data analysis. Chemoecology 27, 33–46. https://doi.org/10.1007/s00049-016-0227-8.

Chang, K.S., Chang, C.S., Park, J.S., 2005. Taxonomic reconsideration of *Alnus hirsuta* var. *hirsuta* and *A. hirsuta* var. *sibirica* in Korea. Bull. Seoul Natl. Univ. Arbor. 25, 82–88.

Chiba, K., Ichizawa, H., Kawai, S., Nishida, T., 2013. α-glucosidase inhibition activity by cyclic diarylheptanoids from *Alnus sieboldiana*. J. Wood Chem. Technol. 33, 44–51. https://doi.org/10.1080/02773813.2012.723778.

da Silva, R.R., Wang, M., Nothias, L.F., van der Hooft, J.J.J., Caraballo-Rodríguez, A.M., Fox, E., Balunas, M.J., Klassen, J.L., Lopes, N.P., Dorrestein, P.C., 2018. Propagating annotations of molecular networks using *in silico* fragmentation. PLoS Comput. Biol. 14, 1–26. https://doi.org/10.1371/journal.pcbi.1006089.

Dührkop, K., Shen, H., Meusel, M., Rousu, J., Böcker, S., 2015. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. Proc. Natl. Acad. Sci. U.S.A. 112, 12580–12585. https://doi.org/10.1073/pnas.1509788112.

Ernst, M., Kang, K.B., Caraballo-Rodríguez, A.M., Nothias, L.-F., Wandy, J., Chen, C., Wang, M., Rogers, S., Medema, M.H., Dorrestein, P.C., van der Hooft, J.J.J., 2019a. MolNetEnhancer: enhanced molecular networks by integrating metabolome mining and annotation tools. Metabolites 9, 144. https://doi.org/10.3390/metabo9070144.

Ernst, M., Nothias, L.F., van der Hooft, J.J.J., Silva, R.R., Saslis-Lagoudakis, C.H., Grace, O.M., Martinez-Swatson, K., Hassemer, G., Funez, L.A., Simonsen, H.T., Medema, M.H., Staerk, D., Nilsson, N., Lovato, P., Dorrestein, P.C., Rønsted, N., 2019b. Assessing specialized metabolite diversity in the cosmopolitan plant genus *Euphorbia* L. Front. Plant Sci. 10, 846. https://doi.org/10.3389/fpls.2019.00846.

Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., Ojima, Y., Tanaka, K., Tanaka, S., Aoshima, K., Oda, Y., Kakazu, Y., Kusano, M., Tohge, T., Matsuda, F., Sawada, Y., Hirai, M.Y., Nakanishi, H., Ikeda, K., Akimoto, N., Maoka, T., Takahashi, H., Ara, T., Sakurai, N., Suzuki, H., Shibata, D., Neumann, S., Iida, T., Tanaka, K., Funatsu, K., Matsuura, F., Soga, T., Taguchi, R., Saito, K., Nishioka, T., 2010. MassBank: a public repository for sharing mass spectral data for life sciences. J. Mass Spectrom. 45, 703–714. https://doi.org/10.1002/jms.1777.

Kang, K.B., Ernst, M., van der Hooft, J.J.J., da Silva, R.R., Park, J., Medema, M.H., Sung, S.H., Dorrestein, P.C., 2019. Comprehensive mass spectrometry-guided phenotyping of plant specialized metabolites reveals metabolic diversity in the cosmopolitan plant family Rhamnaceae. Plant J. 98, 1134–1144. https://doi.org/10.1111/tpj.14292.

Kang, K.B., Kang, S.J., Kim, M.S., Lee, D.Y., Han, S. Il, Kim, T.B., Park, J.Y., Kim, J., Yang, T.J., Sung, S.H., 2018. Chemical and genomic diversity of six *Lonicera* species occurring in Korea. Phytochemistry 155, 126–135. https://doi.org/10.1016/j.phytochem.2018.07.012.

Kim, M.H., Park, K.H., Kim, S.R., Park, K.J., Oh, M.H., Heo, J.H., Yoon, K.H., Yin, J., Yoon, K.H., Lee, M.W., 2016. Two new phenolic compounds from the leaves of *Alnus sibirica* Fisch. ex Turcz. Nat. Prod. Res. 30, 206–213. https://doi.org/10.1080/14786419.2015.1053087.

Kluyver, T., Ragan-kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C., Jupyter development team, 2016. Jupyter Notebooks – a publishing format for reproducible computational workflows. In: Loizides, F., Scmidt, B. (Eds.), Positioning and Power in Academic Publishing: Players, Agents and Agendas. IOS Press, pp. 87–90. https://doi.org/10.3233/978-1-61499-649-1-87.

Krasilnikova, J., Lauberte, L., Stoyanova, E., Abadjieva, D., Chervenkov, M., Mori, M., De Paolis, E., Mladenova, V., Telysheva, G., Botta, B., Kistanova, E., 2018. Oregonin from *Alnus incana* bark affects DNA methyltransferases expression and mitochondrial DNA copies in mouse embryonic fibroblasts. J. Enzyme Inhib. Med. Chem. 33, 1055–1063. https://doi.org/10.1080/14756366.2018.1476504.

Lee, M., Shim, S.Y., Sung, S.H., 2017b. Triterpenoids isolated from *Alnus japonica* inhibited LPS-induced inflammatory mediators in HT-29 cells and RAW264.7 cells. Biol. Pharm. Bull. 40, 1544–1550. https://doi.org/10.1248/bpb.b16-00895.

Lee, M., Song, J.Y., Chin, Y.W., Sung, S.H., 2013. Anti-adipogenic diarylheptanoids from *Alnus hirsuta* f. *sibirica* on 3T3-L1 cells. Bioorg. Med. Chem. Lett. 23, 2069–2073. https://doi.org/10.1016/j.bmcl.2013.01.127.

Lee, D.Y., Kim, H.W., Yang, H., Sung, S.H., 2017a. Hydrolyzable tannins from the fruits of *Terminalia chebula* Retz and their α-glucosidase inhibitory activities. Phytochemistry 137, 109–116. https://doi.org/10.1016/j.phytochem.2017.02.006.

Lee, M.A., Lee, H.K., Kim, S.H., Kim, Y.C., Sung, S.H., 2010. Chemical constituents of *alnus firma* and their inhibitory activity on lipopolysaccharide-induced nitric oxide production in BV2 microglia. Planta Med. 76, 1007–1010. https://doi.org/10.1055/s-0029-1240911.

Liigand, P., Kaupmees, K., Haav, K., Liigand, J., Leito, I., Girod, M., Antoine, R., Kruve, A., 2017. Think Negative: Finding the Best Electrospray Ionization/MS Mode for Your Analyte. Anal. Chem. 89 (11), 5665–5668. https://doi.org/10.1021/acs.analchem.7b00096.

Negrin, A., Long, C., Motley, T.J., Kennelly, E.J., 2019. LC-MS Metabolomics and chemotaxonomy of caffeine-containing holly (*Ilex*) species and related taxa in the Aquifoliaceae. J. Agric. Food Chem. 67, 5687–5699. https://doi.org/10.1021/acs.jafc.8b07168.

Nguyen, D.D., Wu, C.H., Moree, W.J., Lamsa, A., Medema, M.H., Zhao, X., Gavilan, R.G., Aparicio, M., Atencio, L., Jackson, C., Ballesteros, J., Sanchez, J., Watrous, J.D., Phelan, V.V., Van De Wiel, C., Kersten, R.D., Mehnaz, S., De Mot, R., Shank, E.A., Charusanti, P., Nagarajan, H., Duggan, B.M., Moore, B.S., Bandeira, N., Palsson, B., Pogliano, K., Gutieŕrez, M., Dorrestein, P.C., 2013. MS/MS networking guided analysis of molecule and gene cluster families. Proc. Natl. Acad. Sci. U.S.A. 110, E2611–E2620. https://doi.org/10.1073/pnas.1303471110.

Nothias, L.F., Nothias-Esposito, M., da Silva, R., Wang, M., Protsyuk, I., Zhang, Z., Sarvepalli, A., Leyssen, P., Touboul, D., Costa, J., Paolini, J., Alexandrov, T., Litaudon, M., Dorrestein, P.C., 2018. Bioactivity-based molecular networking for the discovery of drug leads in natural product bioassay-guided fractionation. J. Nat. Prod. 81, 758–767. https://doi.org/10.1021/acs.jnatprod.7b00737.

Nothias, L.F., Petras, D., Schmid, R., Dührkop, K., Rainer, J., Sarvepalli, A., Protsyuk, I., Ernst, M., Tsugawa, H., Fleischauer, M., Aicheler, F., Aksenov, A., Alka, O., Allard, P.-M., Barsch, A., Cachet, X., Caraballo, M., Da Silva, R.R., Dang, T., Garg, N., Gauglitz, J.M., Gurevich, A., Isaac, G., Jarmusch, A.K., Kameník, Z., Kang, K.B., Kessler, N., Koester, I., Korf, A., Le Gouellec, A., Ludwig, M., Christian, M.H., McCall, L.-I., McSayles, J., Meyer, S.W., Mohimani, H., Morsy, M., Moyne, O., Neumann, S., Neuweger, H., Nguyen, N.H., Nothias-Esposito, M., Paolini, J., Phelan, V.V., Pluskal, T., Quinn, R.A., Rogers, S., Shrestha, B., Tripathi, A., van der Hooft, J.J.J., Vargas, F., Weldon, K.C., Witting, M., Yang, H., Zhang, Z., Zubeil, F., Kohlbacher, O., Böcker, S., Alexandrov, T., Bandeira, N., Wang, M., Dorrestein, P.C., 2019. Feature-based molecular networking in the GNPS analysis environment. Biorxiv 812404. https://doi.org/10.1101/812404.

Novakovic, M., Nikodinovic-Runic, J., Veselinovic, J., Ilic-Tomic, T., Vidakovic, V., Tesevic, V., Milosavljevic, S., 2017. Bioactive pentacyclic triterpene ester derivatives from *Alnus viridis* ssp. *viridis* bark. J. Nat. Prod. 80, 1255–1263. https://doi.org/10.1021/acs.jnatprod.6b00805.

Park, D., Kim, H.J., Jung, S.Y., Yook, C.S., Jin, C., Lee, Y.S., 2010. A new diarylheptanoid glycoside from the stem bark of *Alnus hirsuta* and protective effects of diarylheptanoid derivatives in human HepG2 cells. Chem. Pharm. Bull. 58, 238–241. https://doi.org/10.1248/cpb.58.238.

Pluskal, T., Castillo, S., Villar-Briones, A., Orešič, M., 2010. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinform. 11, 395. https://doi.org/10.1186/1471-2105-11-395.

R Core Development Team, 2013. R: A Language and Environment for Statistical Computing. Vienna, Austria.

Ren, X., He, T., Chang, Y., Zhao, Y., Chen, X., Bai, S., Wang, L., Shen, M., She, G., 2017. The genus *Alnus*, a comprehensive outline of its chemical constituents and biological activities. Molecules 22, E1383. https://doi.org/10.3390/molecules22081383.

Riethmüller, E., Alberti, Á., Tõth, G., Béni, S., Ortolano, F., Kéry, Á., 2013. Characterisation of diarylheptanoid- and flavonoid-type phenolics in *Corylus avellana* L. leaves and bark by HPLC/DAD-ESI/MS. Phytochem. Anal. 24, 493–503. https://doi.org/10.1002/pca.2452.

Riethmüller, E., Tóth, G., Alberti, Á., Végh, K., Burlini, I., Könczöl, Á., Balogh, G.T., Kéry, Á., 2015. First characterisation of flavonoid- and diarylheptanoid-type antioxidant phenolics in *Corylus maxima* by HPLC-DAD-ESI-MS. J. Pharm. Biomed. Anal. 107, 159–167. https://doi.org/10.1016/j.jpba.2014.12.016.

Robertson, L.P., Hall, C.R., Forster, P.I., Carroll, A.R., 2018. Alkaloid diversity in the leaves of Australian *Flindersia* (Rutaceae) species driven by adaptation to aridity. Phytochemistry 152, 71–81. https://doi.org/10.1016/j.phytochem.2018.04.011.

Rogers, S., Ong, C.W., Wandy, J., Ernst, M., Ridder, L., van der Hooft, J.J.J., 2019. Deciphering complex metabolite mixtures by unsupervised and supervised substructure discovery and semi-automated annotation from MS/MS spectra. Faraday Discuss. 218, 284–302. https://doi.org/10.1039/C8FD00235E.

Ruttkies, C., Schymanski, E.L., Wolf, S., Hollender, J., Neumann, S., 2016. MetFrag relaunched: incorporating strategies beyond *in silico* fragmentation. J. Cheminformatics 8, 3. https://doi.org/10.1186/s13321-016-0115-9.

Sati, S.C., Sati, N., Sati, O.P., 2011. Bioactive constituents and medicinal importance of genus *Alnus*. Pharmacogn. Rev. 5, 174–183. https://doi.org/10.4103/0973-7847.91115.

Sawada, Y., Nakabayashi, R., Yamada, Y., Suzuki, M., Sato, M., Sakata, A., Akiyama, K., Sakurai, T., Matsuda, F., Aoki, T., Hirai, M.Y., Saito, K., 2012. RIKEN tandem mass spectral database (ReSpect) for phytochemicals: a plant-specific MS/MS-based data resource and database. Phytochemistry 82, 38–45. https://doi.org/10.1016/j.phytochem.2012.07.007.

Sedio, B.E., Echeverri, J.C.R., Boya, C.A., Wright, S.J., 2017. Sources of variation in foliar secondary chemistry in a tropical forest tree community. Ecology 98, 616–623. https://doi.org/10.1002/ecy.1689.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504. https://doi.org/10.1101/gr.1239303.

Sumner, L.W., Amberg, A., Barrett, D., Beale, M.H., Beger, R., Daykin, C.A., Fan, T.W.M., Fiehn, O., Goodacre, R., Griffin, J.L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A.N., Lindon, J.C., Marriott, P., Nicholls, A.W., Reily, M.D., Thaden, J.J., Viant, M.R., 2007. Proposed minimum reporting standards for chemical analysis: chemical analysis working group (CAWG) metabolomics standards initiative (MSI). Metabolomics 3, 211–221. https://doi.org/10.1007/s11306-007-0082-2.

Sung, S.H., Lee, M., 2015. Anti-adipogenic activity of a new cyclic diarylheptanoid isolated from *Alnus japonica* on 3T3-L1 cells via modulation of PPARγ, C/EBPα and SREBP1c signaling. Bioorg. Med. Chem. Lett. 25, 4648–4651. https://doi.org/10.1016/j.bmcl.2015.08.032.

van der Hooft, J.J.J., Wandy, J., Barrett, M.P., Burgess, K.E.V., Rogers, S., 2016. Topic modeling for untargeted substructure exploration in metabolomics. Proc. Natl. Acad. Sci. U.S.A. 113, 13738–13743. https://doi.org/10.1073/pnas.1608041113.

Wandy, J., Zhu, Y., Van Der Hooft, J.J.J., Daly, R., Barrett, M.P., Rogers, S., 2018. Ms2lda.org: web-based topic modelling for substructure discovery in mass spectrometry. Bioinformatics 34, 317–318. https://doi.org/10.1093/bioinformatics/btx582.

Wang, M., Carver, J.J., Phelan, V.V., Sanchez, L.M., Garg, N., Peng, Y., Nguyen, D.D., Watrous, J., Kapono, C.A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A.V., Meehan, M.J., Liu, W.T., Crüsemann, M., Boudreau, P.D., Esquenazi, E., Sandoval-Calderón, M., Kersten, R.D., Pace, L.A., Quinn, R.A., Duncan, K.R., Hsu, C.C., Floros, D.J., Gavilan, R.G., Kleigrewe, K., Northen, T., Dutton, R.J., Parrot, D., Carlson, E.E., Aigle, B., Michelsen, C.F., Jelsbak, L., Sohlenkamp, C., Pevzner, P., Edlund, A., McLean, J., Piel, J., Murphy, B.T., Gerwick, L., Liaw, C.C., Yang, Y.L., Humpf, H.U., Maansson, M., Keyzers, R.A., Sims, A.C., Johnson, A.R., Sidebottom, A.M., Sedio, B.E., Klitgaard, A., Larson, C.B., Boya, C.A.P., Torres-Mendoza, D., Gonzalez, D.J., Silva, D.B., Marques, L.M., Demarque, D.P., Pociute, E., O'Neill, E.C., Briand, E., Helfrich, E.J.N., Granatosky, E.A., Glukhov, E., Ryffel, F., Houson, H., Mohimani, H., Kharbush, J.J., Zeng, Y., Vorholt, J.A., Kurita, K.L., Charusanti, P., McPhail, K.L., Nielsen, K.F., Vuong, L., Elfeki, M., Traxler, M.F., Engene, N., Koyama, N., Vining, O.B., Baric, R., Silva, R.R., Mascuch, S.J., Tomasi, S., Jenkins, S., Macherla, V., Hoffman, T., Agarwal, V., Williams, P.G., Dai, J., Neupane, R., Gurr, J., Rodríguez, A.M.C., Lamsa, A., Zhang, C., Dorrestein, K., Duggan, B.M., Almaliti, J., Allard, P.M., Phapale, P., Nothias, L.F., Alexandrov, T., Litaudon, M., Wolfender, J.L., Kyle, J.E., Metz, T.O., Peryea, T., Nguyen, D.T., VanLeer, D., Shinn, P., Jadhav, A., Müller, R., Waters, K.M., Shi, W., Liu, X., Zhang, L., Knight, R., Jensen, P.R., Palsson, B., Pogliano, K., Linington, R.G., Gutiérrez, M., Lopes, N.P., Gerwick, W.H., Moore, B.S., Dorrestein, P.C., Bandeira, N., 2016. Sharing and community curation of mass spectrometry data with global natural products social molecular networking. Nat. Biotechnol. 34, 828–837. https://doi.org/10.1038/nbt.3597.

Wansi, J.D., Lallemand, M.C., Chiozem, D.D., Toze, F.A.A., Mbaze, L.M., Naharkhan, S., Iqbal, M.C., Tillequin, F., Wandji, J., Fomum, Z.T., 2007. α-Glucosidase inhibitory constituents from stem bark of *Terminalia superba* (Combretaceae). Phytochemistry 68, 2096–2100. https://doi.org/10.1016/j.phytochem.2007.02.020.

Watrous, J., Roach, P., Alexandrov, T., Heath, B.S., Yang, J.Y., Kersten, R.D., Van Der Voort, M., Pogliano, K., Gross, H., Raaijmakers, J.M., Moore, B.S., Laskin, J., Bandeira, N., Dorrestein, P.C., 2012. Mass spectral molecular networking of living microbial colonies. Proc. Natl. Acad. Sci. U.S.A. 109, 1743–1752. https://doi.org/10.1073/pnas.1203689109.

Wink, M., 2010. Introduction: biochemistry, physiology and ecological functions of secondary metabolites. In: Wink, M. (Ed.), Annual Plant Reviews Volume 40: Biochemistry of Plant Secondary Metabolism, second ed. Blackwell Publishing Ltd., pp. 1–19. https://doi.org/10.1002/9781444320503.ch1.

Wolfender, J.L., Litaudon, M., Touboul, D., Queiroz, E.F., 2019. Innovative omics-based approaches for prioritisation and targeted isolation of natural products-new strategies for drug discovery. Nat. Prod. Rep. 36, 855–868. https://doi.org/10.1039/c9np00004f.

Worley, B., Powers, R., 2012. Multivariate analysis in metabolomics. Curr. Metabolomics 1, 92–107. https://doi.org/10.2174/2213235x130108.

Yu, Y.B., Miyashiro, H., Nakamura, N., Hattori, M., Jong, C.P., 2007. Effects of triterpenoids and flavonoids isolated from *Alnus firma* on HIV-1 viral enzymes. Arch. Pharm. Res. 30, 820–826. https://doi.org/10.1007/BF02978831.

Zhou, S., Allard, P.-M., Wolfrum, C., Ke, C., Tang, C., Ye, Y., Wolfender, J.-L., 2019. Identification of chemotypes in bitter melon by metabolomics: a plant with potential benefit for management of diabetes in traditional Chinese medicine. Metabolomics 15, 104. https://doi.org/10.1007/s11306-019-1565-7.