ORIGINAL ARTICLE

Unbiased Metabolomics of Volatile Secondary Metabolites in Essential Oils Originated from Myrtaceae Species

Ganapaty Manickavasagam[1,2] · Priscilla Wong Chie San[3] · Sara Ghorbani Gorji[4] · Bunleu Sungthong[5] · Yong Yoke Keong[6] · Melissa Fitzgerald[4] · Firli Rahmah Primula Dewi[7] · Vuanghao Lim[1]

Abstract

In this study, an untargeted metabolomic approach was applied to analyze volatile secondary metabolites in five different essential oils (i.e., anise myrtle oil, honey myrtle oil, lemon myrtle oil, lemon-scented tea tree oil, and peppermint gum oil) from the Myrtaceae family, originating from the Australian bush plants. A total of 503 volatile secondary metabolites were tentatively identified across the five oils using static headspace two-dimensional gas chromatography time-of-flight mass spectrometry (GC×GC-TOFMS). Both principal component analysis and partial least squares discriminant analysis generated using the identified metabolites revealed a clear discrimination into five clusters of oils. Nevertheless, only 24 of 503 identified metabolites significantly differentiated the five essential oils and these metabolites were identified as biomarkers based on their variable importance in the projection score. In terms of cluster analysis, the essential oils were clustered into five different groups using hierarchical cluster analysis. However, K-mean clustering showed that lemon myrtle oil and lemon-scented tea tree oil were clustered together due to their very close chemical composition compared to the other studied essential oils. Additionally, metabolite set enrichment analysis revealed that ketones represented the primary class in all essential oils, followed by aryl-aldehydes. In summary, despite belonging to the same Myrtaceae family, these oils can be well discriminated based on their respective species.

Keywords Essential oils · Metabolites · Metabolomics · Myrtaceae family · Static headspace GC×GC-TOFMS

✉ Vuanghao Lim
  vlim@usm.my

1 Advanced Medical and Dental Institute, Universiti Sains Malaysia, Bertam, 13200 Kepala Batas, Pulau Pinang, Malaysia

2 School of Chemical Sciences, Universiti Sains Malaysia, 11800 Gelugor, Pulau Pinang, Malaysia

3 School of Distance Education, Universiti Sains Malaysia, 11800 Gelugor, Pulau Pinang, Malaysia

4 School of Agriculture and Food Sciences, The University of Queensland, Brisbane, QLD 4072, Australia

5 Integrative Pharmaceuticals and Innovative of Pharmaceutical Technology Research Unit, Faculty of Pharmacy, Mahasarakham University, Kantharawichai 44150, Maha Sarakham, Thailand

6 Department of Human Anatomy, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

7 Department of Biology, Faculty of Science and Technology, Universitas Airlangga, Surabaya 60115, Indonesia

1 Introduction

Dicotyledonous woody plants are generally classified into 52 families and one of its largest families is known as the Myrtaceae family [1]. Family Myrtaceae is also widely identified as guava, myrtle, clove, or eucalyptus family. The Myrtaceae family is positioned within the order Myrtales, comprising more than 175 genera and 6000 species [2, 3]. Not to mention, the Myrtaceae family has undergone evolutionary changes, transitioning from its earlier, less advanced forms found in moist rainforest climates to more specialized forms adapted to the extremely dry and highly seasonal conditions of the western, semi-arid region. Therefore, the Myrtaceae family is structurally and floristically dominant and distributed across all continents except Antarctica, with its primary occurrence concentrated in the tropical and temperate regions of the Southern Hemisphere, such as eastern Australia, Malesia (the Australasian region), and the Neotropics [4].

🖄 Springer

The prominent genera in the mentioned regions mainly comprised *Eucalyptus*, *Syncarpia*, *Callistemon*, *Psidium*, *Angophora*, *Eugenia*, *Syzygium*, *Melaleuca*, *Myrtus*, *Myrcia*, and *Corymbia*. *Melaleuca* (236), *Eugenia* (3000), *Psidium* (150), *Myrcia* (765), *Syzygium* (1000), *Eucalyptus* (520), and *Myrtus* (150) represent a few of the most substantial Myrtaceae genera with the most documented species [5–7]. However, the numbers do not end there. New species are constantly being discovered in remote tropical forests, and many of them are yet to be officially identified at the species level [4]. It is not unanticipated that there are a variety of ways that different species have been used by humans, given the vastness and diversity of the Myrtaceae family. As a result, several species within the Myrtaceae family have been intentionally grown, resulting in significant economic value both in their country of origin and in plantations worldwide [4]. Regardless of the species and genus, Myrtaceae family species are utilized in various industries for distinct purposes, such as plantation (charcoal, gasoline, pulp, and firewood), agriculture (windbreaks, pesticides, honey, and livestock breeding), horticulture (ornamental species, cut flowers, and foliage), medicine (anti-diabetic, anti-microbial, anti-viral, anti-fungal, traditional medicines, and vector control), and food (sauce, paste, flavourings, condiments, herbs, tea, and fruits) [4, 8–10].

One of the common products of Myrtaceae family plants is essential oils. These essential oils can be extracted and isolated from various organs of plants, such as aerials, fruits, leaves, flowers, and bark [11]. At room temperature, essential oils are typically liquid and colourless. Except for vetiver, sassafras, and cinnamon, other essential oils possess a distinct aroma, a low density (< 1 kg/m^3), and demonstrate significant optical activity. These essential oils in herbs are responsible for the various fragrances that plants generate. They are frequently applied in grooming products, scented sanctuaries, and aromatherapeutic practices. The latter is meant to be a therapeutic treatment that includes reflexology, inhalations, or steam baths with these oils [12]. In terms of medicinal and therapeutic properties, essential oils have been reported to exhibit anti-bacterial, anti-cancer, anti-protozoal, anti-viral, anti-fungal, anti-diabetic, anti-inflammatory and anti-septic [12, 13]. A wide range of compounds has been reported in essential oils from different species of the Myrtaceae family, such as phenolic compounds, flavonoids, fatty acids, saponins, steroids, tannins, catechins, and volatile organic compounds [14–16]. However, volatile secondary metabolites are the most gained constituent of the species of the Myrtaceae family among researchers. To date, more than 100 volatile organic compounds have been reported and these compounds are generally classified as secondary metabolites [17, 18]. The common classes of volatile secondary metabolites reported in the Myrtaceae family primarily comprised oxygenated and hydrocarbon sesquiterpenes with monoterpenes [19]. The mevalonate pathway is responsible for the production of these distinct terpenes. Besides volatile secondary metabolites, these essential oils also consist of polyphenols generated through the shikimate pathway [12].

A review of the existing literature indicated that no reports are available on untargeted metabolic profiling of volatile secondary metabolites in essential oils from different species within the Myrtaceae family from the Australian bush plants. The reported studies are mostly limited to the investigation of the chemical composition of the essential oils from Myrtaceae family rather than integrating the chemical composition and metabolomics [20, 21]. Thus, this study aimed to employ metabolomic approaches to qualitatively identify metabolites for the determination of major classes of compounds and discrimination with clustering patterns of essential oils originating from five different species of the Myrtaceae family. The selected essential oils for the untargeted metabolic profiling of volatile secondary metabolites are anise myrtle oil, honey myrtle oil, lemon myrtle oil, lemon-scented tea tree oil, and peppermint gum oil. Subsequently, volatile secondary metabolites with greater potential (known as biomarkers) to differentiate between the five essential oils, originating from different species within the same family, can be ascertained. Following this, the biomarkers from this research work can be useful in studying the respective roles of essential oils from different species in the field of medicinal chemistry, such as spasmolytic, analgesic, anticonvulsant, expectorant, anti-inflammatory, and anti-microbial activities [22, 23].

2 Materials and Methods

2.1 Collection of Essential Oils and Sample Preparation

The selected five different essential oils formerly originating from Australia were purchased from Essentially Australia® (https://essentiallyaustralia.com.au/). These essential oils belong to the Myrtaceae family, specifically anise myrtle [*Syzygium anisatum* (*S. anisatum*)] oil, honey myrtle [*Melaleuca teretifolia* (*M. teretifolia*)] oil, lemon myrtle [*Backhousia citriodora* (*B. citroodora*)] oil, lemon-scented tea tree [*Leptospermum petersonii* (*L. petersonii*)] oil, and peppermint gum [*Mentha piperita* (*M. piperita*)] oil. The essential oils were stored in amber bottles at a storage temperature of 4 °C and protected from direct light exposure. These oils were used directly as provided without any preconcentration and modification steps. The sample (0.5 g) was weighed and placed in 20-mL gas chromatography

headspace vials with silicon caps. The vials were then stored at a temperature of 80 °C to prepare them for subsequent analyses. After 10 min of spinning at 80 °C, 1.5 mL of headspace was collected with a 2.5 mL headspace syringe.

2.2 Analysis of Volatile Metabolites Using GC×GC-TOFMS

The volatile metabolite compositions in all five essential oils were extensively analyzed using static headspace extraction coupled with separation by GC×GC-TOFMS (LECO Pegasus 4D, Castle Hill, Australia). The GC×GC-TOFMS comprised autosampler and two-dimensional gas chromatography. The autosampler settings are as follows: incubator and agitator turned on, incubator temperature of 80 °C, fill speed of 50 μL/s, injection speed of 100 μL/s, injection volume of 1.5 mL, and syringe temperature of 80 °C. The key conditions of the gas chromatography are as follows: splitless injection mode, injector temperature of 250 °C, helium (99.9999%) as the carrier gas, and flow rate of 1 mL/min. The chromatographical separation condition of two-dimensional gas chromatography is divided into primary and secondary conditions. In the primary condition, Agilent DB-624UI was used as the primary column (mid-polar; composition of 6% cyanonpropyl phenyl, 94% polydimethyl siloxane; 30 m x 250 μm x 1.4 μm). In the secondary condition, Restek Stabilwax was used as the secondary column (polar; composition of crossbond polyethylene glycol, 0.9 m x 250 μm x 0.50 μm) and the secondary column temperature of 15 °C offset primary. Other conditions of GC×GC as follows: modulator temperature of 25 °C offset primary, modulation time of 2.5 s, hot pulse time of 0.4 s, and cool time of 0.85 s. The mass spectrometry conditions are as follows: transfer line of 240 °C, voltage of 1700 V, electron energy of 70 V, scan rate of 200 spectra/s, a mass range of 35–350 m/z, and ion source of 240 °C. During the analysis, a void vial was utilized as a blank to determine background contamination and all five essential oils were mixed to provide a quality assurance standard.

2.3 Data Processing and Statistical Analysis

Each essential oil was static headspace extracted and analyzed in three independent replicates. The GC×GC-TOFMS data was processed using LECO ChromaTOF 4.50 software, which included pre-processing tasks such as baseline correction. For the identification of volatile metabolites, a library matching (NIST 11 v2.0) approach was comprehensively applied and putative identification was determined based on a similarity threshold of at least 80% with the NIST library to establish a reasonably high level of confidence in the identification process. Subsequently, the

MetaboAnalyst version 5.0 JavaServer was used to further analyze the processed data from GC×GC-TOFMS for multivariate statistical analysis (MSVA) and univariate statistical analysis [24]. Prior to both analyses, the dataset was pre-processed with log transformation, normalized by the median, and not subjected to scaling. In the initial stage of MSVA, the principal component analysis (PCA) and partial least square-discrimination analysis (PLS-DA) were applied to the processed data matrix of essential oils. PCA was used to access the clustering trend of the comprehensive depiction of the essential oils; meanwhile, PLS-DA was used to analyze further the correlation of metabolites in the essential oils. Afterwards, the variable selection, namely variable importance of projection (VIP), was used to select the top 25 possible chemical markers (VIP greater than 1.5) [25]. Besides, hierarchical clustering analysis (HCA) was used to visualize the metabolite profiles and reveal the relationship between metabolites and samples [26]. In addition, K-mean cluster analysis is another classification technique which widely used to minimize the distance within clusters and maximize the distance between different clusters [27].

3 Results and Discussion

3.1 Bilinear Models

3.1.1 Principal Component Analysis

In the initial stage, the metabolite profiles (representative total ion chromatogram of each essential oil presented in Figure S1) were compared in PCA score plots by submitting the combined anise myrtle oil (AM), honey myrtle oil (HM), lemon myrtle oil (LM), lemon-scented tea tree oil (LST) and peppermint gum oil (PG) in triplicate samples. Figure 1a is the PCA 2D-score plot combination of PC1 and PC2 of the metabolites in different oils. As illustrated in Fig. 1a, all LST and LM samples were predominantly distributed on the positive region (right side) of PC1 of the PCA score plot. Conversely, HM, AM and PG samples were present predominately on the negative region (left side). In conjunction with that, Madala et al. [28] proved that PCA score plots showing different clusters of samples suffice to understand the apparent clustering of samples due to their biochemical constituents. The PCA analysis dataset of the five essential oils showed distinct differences in clustering patterns. The five significant clusters scattered far apart except for LM and LST samples clustering side by side. However, the closer distribution of LM and LST in the PCA plot is probably due to the relatively small differences in the metabolic profile among LM and LST, or these oils are the most chemically similar. There was a great degree of separation between HM
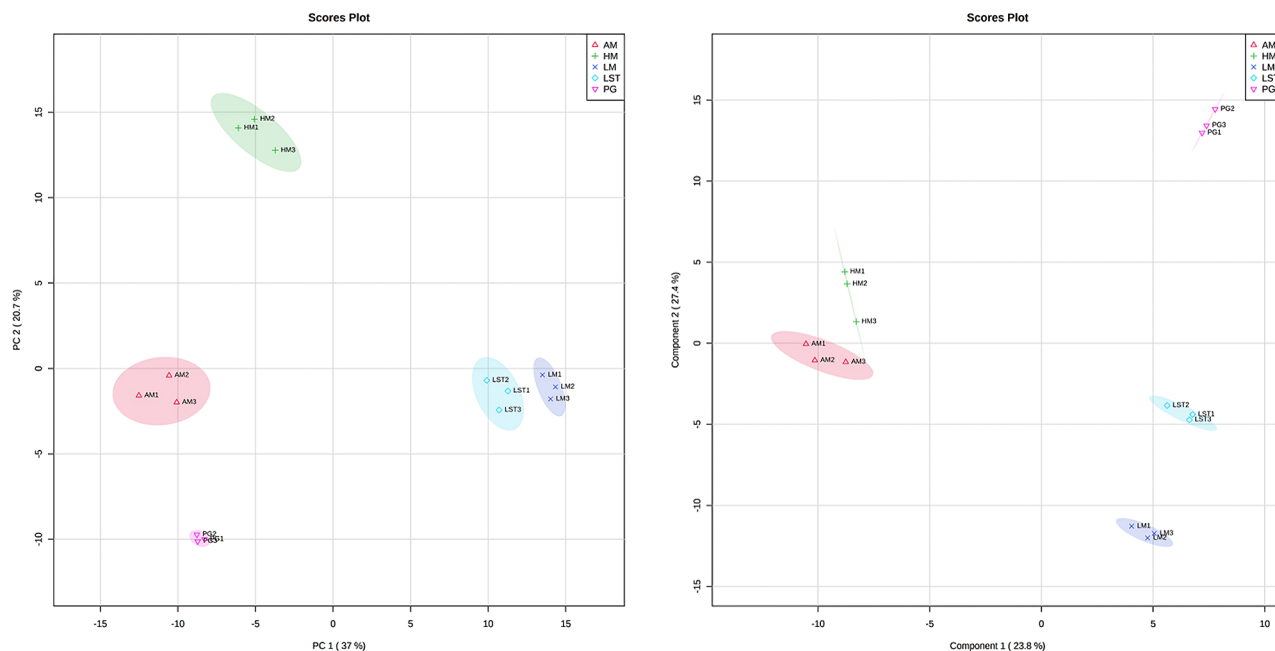
Fig. 1 (a) 2D-score plot of PCA model based on five essential oils, (b) 2D-score plot of PLS-DA model based on five essential oils

and the other four essential oil samples along the PC2 axis, indicating that the metabolic profiles of HM were prominently different from those of the other four oils. In a recent study, Elhawary et al. [29] identified metabolomics profiles of five essential oils belonging to the same Rosa varieties, and the PCA analysis gave rise to five different clusters, which were similar to the present study. In this study, the five selected essential oils belonging to the same Myrtaceae family showed five different clusters in the PCA score plot. Out of these five samples, PG gave the smallest area of a range of metabolites. The small range area indicates a minor variation among the replicates. PCA biplot was used to see the relationship between loading scores and exhibited using correlation scaling (Figure S2). It demonstrates divergence between all metabolites found in the five oil samples. Based on the biplot, isothiocyanato-methane was the metabolite which significantly discriminated HM from other oils. In addition, 3-methyl- pentanoic acid methyl ester discriminated PG from others. HM. On the other hand, PG, AM, LM and LST each formed their own cluster and did not overlap, meaning that each cluster did not share the same metabolites and/or the same metabolites with significantly different concentrations. Similarly, in a previous study of untargeted metabolomics profiling of aroma compounds in Australian indigenous fruits, the results showed the discrimination of the samples in three clusters in the PCA score plot, indicating the aroma components in the fruits studied were different [25]. However, Wan Salleh et al. [30] reported that five essential oils from leaves of different Beilschmiedia species were able to cluster into three groups instead of five.

Group 1 consists of essential oils from *B. kunstleri* and *B. maingayi*, group 2 consists of *B. penangiana* and *B. glabra*, and group 3 consists solely of *B. madang*. This is attributed to different volatile secondary metabolites responsible for the characterization of clusters. Also, Carmargo et al. [31] reported that essential oils from leaves and branches of nine species of the Celastraceae family were able to cluster into 2 distinct groups, which can be attributed to the presence of oil storage structures in the parts of the species used in the extraction process. Therefore, there are two possibilities associated with the excellent discrimination pattern of essential oils from five different species, specifically different compositions of volatile secondary metabolites or the purchased essential oils of different species extracted from different parts of the plant species.

3.1.2 Partial Least Squares-Discriminant Analysis

To further verify the difference in metabolic profiles among these essential oils, PLS-DA was carried out to identify and classify the biomarker, thus helping in improving data interpretations [32]. PLS-DA 2D-score plot shows component 1 and component 2 with the highest percentage of correlation, 23.8% and 27.4%, respectively (Fig. 1b). There are five distinct clusters of the five samples shown in the plot. LM and LST samples were clustering near each other, the same as HM and AM samples, leaving only the PG samples discriminated far from those four samples. Interpretation can be made more clearly by referring to the 2D-score plot in Fig. 1b and Table S1, as the sample groups were classified

more precisely. The scattering and non-overlapping clusters showed different biomarker compounds existed in each essential oil sample even though the samples were all under the same family. In addition, a total of the top 25 potential biomarker metabolites were sorted, as shown in Figure S3 and the details are tabulated in Table 1. VIP plot summarised the significance of the variables to clarify the X data, correlated them to the Y data and introduced variables with the strongest influence in the model [33]. In this model, the VIP values were set at greater than 2.0 and

Table 1 Detailed information on the top 25 metabolites from five different essential oils identified by VIP based on the PLS-DA model

No.	Peak no.	Metabolite	Group (VIP Score)	CAS no.	Molecular formula	Exact mass	Level
1	455	3-Methyl-1-pentanol	Primary alcohol (4.004)	589-35-5	$C_6H_{14}O$	102.17	L2
2	458	1-Octen-3-ol	Alkenyl alcohol (3.408)	3391-86-4	$C_8H_{16}O$	128.21	L2
3	61	3-Methyl-pentanoic acid methyl ester	Ester (3.401)	2177-78-8	$C_7H_{14}O_2$	130.19	L2
4	260	1-(2-Furanyl)-2-3-methyl-butanone	Ketone (2.983)	20907-04-4	$C_9H_{12}O_2$	152.19	L2
5	221	Butanal	Aldehyde (2.941)	123-72-8	C_4H_8O	72.11	L2
6	419	3-Methyl-2-buten-1-ol acetate	Carboxylic acid ester (2.531)	1191-16-8	$C_7H_{12}O_2$	128.17	L2
7	369	3,4-Dimethyl-(Z)-2-pentene	Alkene 2.509	4914-91-4	C_7H_{14}	98.19	L2
8	178	3,5-Dimethyl-cyclohexanol	Cyclic alcohol (2.491)	5441-52-1	$C_8H_{16}O$	128.21	L2
9	140	4-Methyl-decane	Alkane (2.457)	2847-72-5	$C_{11}H_{24}$	156.31	L2
10	153	Cyclopentanol	Cyclic alcohol (2.436)	96-41-3	$C_5H_{10}O$	86.13	L2
11	206	2-Methyl-butanoic acid methyl ester	Fatty acid ester 2.401	868-57-5	$C_6H_{12}O$	116.16	L2
12	33	Cis-2,6-dimethyl-2,6-octadiene	Diene (2.389)	2492-22-0	$C_{10}H_{18}$	138.25	L2
13	380	2-Methyl-2-heptene	Alkene (2.356)	627-97-4	C_8H_{16}	112.21	L2
14	222	3,7,7-Trimethyl- (1α,3α,6α)-bicyclo[4.1.0]heptane	Terpenoid (2.258)	554-59-6	$C_{10}H_{18}$	138.25	L2
15	449	2-Methyl-1-penten-3-one	Ketone (2.247)	25044-01-3	$C_6H_{10}O$	98.14	L2
16	85	Isothiocyanato-methane	Organo sulphur (2.163)	556-61-6	C_2H_3NS	73.12	L2
17	74	Nonanal	Fatty aldehyde (2.154)	124-19-6	$C_9H_{18}O$	142.24	L2
18	72	Octanal	Aldehyde (2.149)	124-13-0	$C_8H_{16}O$	128.21	L2
19	232	2-Methyl-5-(1-methylethyl)-(1α, 2α, 5α)-bicyclo[3.1.0]hex-3-en-2-ol	Bicyclic monoterpenoid (2.144)	97631-68-0	$C_{10}H_{16}O$	152.24	L2
20	301	4-Nonene	Alkene (2.119)	2198-23-4	C_9H_{18}	126.24	L2
21	282	1,5-Dimethyl-(1 S)-6,8-dioxabicyclo[3.2.1]octane	Terpenoid (2.04)	28401-39-0	$C_8H_{14}O_2$	142.19	L2
22	378	(Z)-2-Cctene	Alkene (2.046)	7642-04-8	C_8H_{16}	112.21	L2
23	62	3-Methyl-pentanal	Aldehyde (2.034)	15877-57-3	$C_6H_{12}O$	100.16	L2
24	367	2-Propen-1-ol	Allyl alcohol (2.028)	107-18-6	C_3H_6O	58.08	L2
25	197	4,5-Epoxy-trans-carane	Terpene (1.999)	6909-20-2	$C_{10}H_{16}O$	152.23	L2

VIP: variable of importance; CAS: chemical abstracts service; Level of identification based on the guidelines by Salek et al. [35], where L2 signifies putatively identified compounds through NIST14 MS spectrum library matching

loadings correlation coefficient [p(corr)] values above 0.5 to be significant for sample separation in the PLS-DA model. As a result, only 24 compounds were significant biomarkers from the top 25 shortlisted metabolites. With reference to Tables 1, 3-methyl-1-pentanol has the highest VIP score, which is 4.0041 in component 1, meaning that it is the most significant biomarker in all samples. This compound was found most abundantly in the PG samples but the least in the HM samples. Followed by 1-octen-3-ol and 3-methyl-pentanoic acid methyl ester are the second and third biomarkers with VIP scores of more than 3 (3.4081 and 3.4008, respectively). Similarly, the rest of the 21 biomarkers and their distribution (highest and lowest) in respective oils can be observed in Table 1. There are several factors that can influence the type and composition of volatile secondary metabolites in essential oils of different species and eventually lead to possibilities to become putative biomarkers to differentiate essential oils, such as harvesting time of the plant where the essential oils are originating from, part of the plant where the essential oils are being extracted, agricultural methods and practices, soil type, altitude, climate, growth conditions of plants, and developmental stage of plants [34].

3.2 Cluster Analyses: Hierarchical and Partitional Clusterings

The heatmap of HCA depicted a data matrix as a grid, with each row and column representing a variable and sample, respectively. The cells contain the values of the pairwise associations between them (variable and sample) and values in each cell are then colour-coded using a gradient scale, with warmer colours (e.g., red) representing higher values and cooler colours representing lower values (e.g., blue). The heatmap was constructed based on the top 25 metabolites according to *t*-test (Figure S4). It is observed that there were five compounds that were highly positive-correlated in the PG samples but negatively correlated in HM, LM, and LST samples, namely 3-methyl-3-buten-1-ol acetate, 4,4-dimethyl-2-cyclopenten-1-one, 2-hexanone,4-methyl, 4-methyl-1-pentanol, and 4-methyl-pentanoic acid methyl ester. Both isopropyl acetate and 1-methylethyl formic acid ester were positively correlated in PG and AM samples but negatively correlated in HM, LM, and LST samples. 2-(2-propenyl)-furan and 3-methyl-phenol were positively correlated in PG, HM and AM samples but negatively correlated in LM and LST samples. Meanwhile, 1,1,2-trimethyl-cyclopentane, 3-methyl-2-pentanone, 2,3,5-trimethyl-furan, and 3-methyl-phenol were positively correlated in PG and AM samples but negatively correlated in HM, LM and LST samples. Out of the top 25 metabolites (based on *T*-test), only (z)-2-octene and isothiocyanato-methane were

negatively correlated in PG samples. In addition, only ethylbenzene and α-myrcene showed a positive correlation in LM samples. Similarly, in the previous study by Southwell et al. [36] on essential oil from leaves of LM, the result showed that the oil was found to contain myrcene (0.1–0.7%).

On the other hand, the clustering of different essential oils was visualized using hierarchical clustering. Based on the dendrogram in Fig. 2a, each triplet analysis of PG, AM, HM, LM, and LST samples forming own significant cluster (five groups). The PCA model supports the formation of five clusters. The PG sample was correlated to the AM sample and both samples correlated with HM. PG, AM, and HM forming a cluster also correlate with clusters formed by LM and LST samples. Similarly, Yong Lee et al. [37] reported three distinct clusters for three different essential oils (i.e., spearmint, apple mint, and chocolate mint).

Since the performance of hierarchical clustering can be reduced based size of the data matrix, thus partitional clustering using K-mean is preferred as it has outperformed clustering in terms of entropy and purity using a cosine similarity measure [38]. In conjunction with that, K-mean clustering was applied for partitional clustering. Four significant clusters of the five samples in this study were formed using K-mean clustering and each cluster was discriminated far from each other (Fig. 2b). The centroid is the centre of the cluster for each distinct cluster. PG sample formed the smallest clusters and the distance from the centroid is the lowest. This also indicates that the samples in the PG cluster possessed the least variability compared to the other four sample clusters. The triplet analysis of each oil formed individual clusters for AM, HM, and PG samples (Table S2). However, triplicate analyses of LM and LST were grouped into a single big cluster. Combining samples from two different essential oils (LM and LST) in a single cluster might be due to the similar metabolites in these two essential oils. This is supported by the closer distribution of the LM and LST in the PCA plot. Not to mention, these two oils also correlated with each other in hierarchical clustering. Therefore, the K-mean performs better in the clustering of samples, as reported by Abdalla [38].

3.3 Metabolite Set Enrichment Analysis

Metabolite set enrichment analysis is an approach designed to assist metabolomics scholars in identifying and interpreting patterns of metabolite concentration changes in a biologically meaningful way [39]. A total of 503 compounds were identified by utilizing GC×GC-TOFMS. Ketones were identified as the major class (with the greatest enrichment ratio) among the five essential oils, followed by aryl-aldehydes and phenylacetaldehyde. Butanone and acetone were the main components of the ketone class, followed by
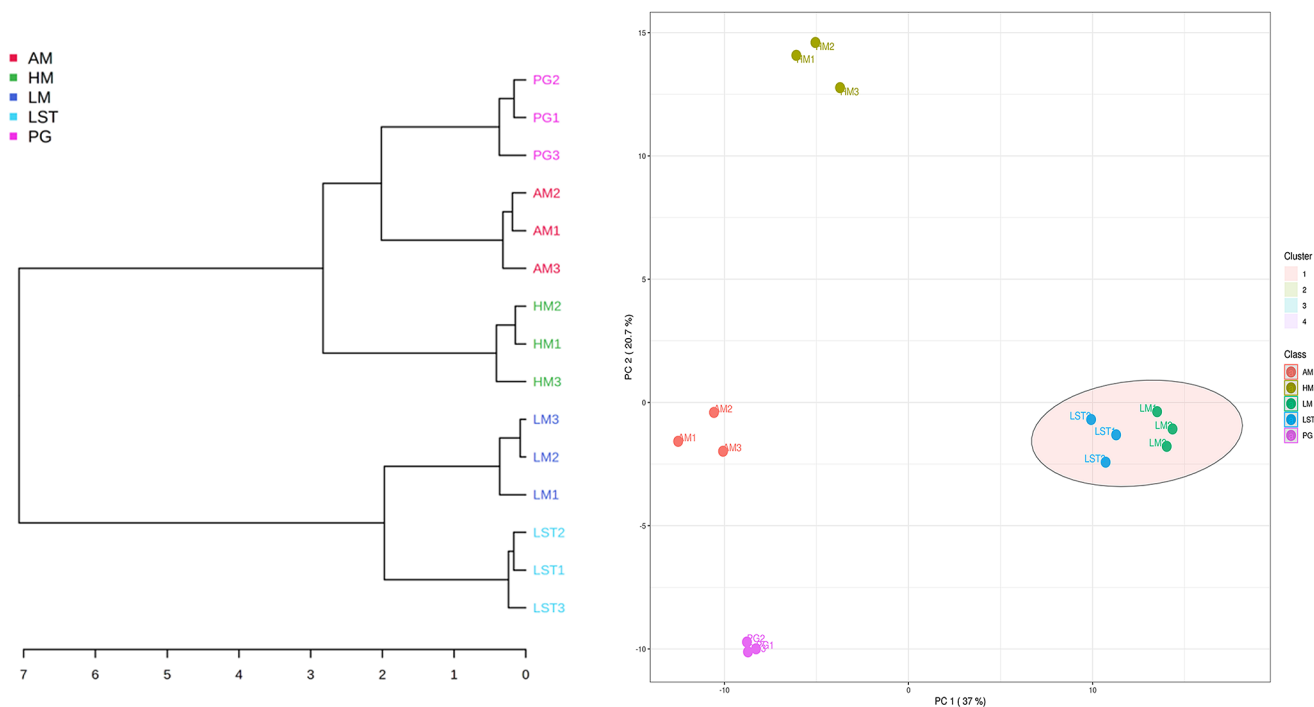
Fig. 2 (**a**) Hierarchical clustering dendrogram of five essential oils, (**b**) K-mean clustering of five essential oils

aryl-aldehyde represented by 2-furancarboxaldehyde (Figure S5). Similarly, a study by Clarke [40] also presented that essential oils originating from aniseed, sage, hyssop, fennel, rosemary, caraway, pennyroyal, peppermint, spearmint, and dill also contain ketones as the primary metabolite. Also, Lu et al. [41] and Mugao et al. [42] reported ketones as one of the major classes of volatile secondary metabolites in essential oils originating from peppermint (*Mentha piperita* L.) and ginger, garlic, tick berry, and Mexican marigold, respectively. The higher abundance of ketone in essential oils could be associated with the reason that plant species naturally produce higher amounts of ketone-containing compounds as part of their essential oil composition. These compounds may be synthesized through specific biochemical pathways within the plant. On a positive note, certain ketones found in essential oils have been associated with various therapeutic effects, including analgesic, anti-inflammatory, and antimicrobial properties. For example, menthone and camphor are known for their cooling and pain-relieving effects [43, 44]. In terms of *p*-value enrichment analysis, wax monoester possessed a *p*-value of 2.48×10^{-16} showing the greatest actual hits and the class represented by hexyl acetate, propyl propionate, 3-methylbutyl formate, 3-methylbutyl propanoate, 2-methypropyl propanoate; hexyl propionate, methyl 3-methylbutanoate, isopentyl isopentanoate, methyl pentanoate; pentyl 3-methylbutanoate, 2-methylpropyl 2-ethyl-2-methylbutanoate, and ethyl hexanoate.

4 Conclusion

The primary objectives of the study such as (i) identification of volatile secondary metabolites via an untargeted approach and determination of major classes of these metabolites, (ii) discrimination and clustering patterns associated with the identified metabolites, and (iii) identification of biomarkers among the identified metabolites, were accomplished. The findings of this study suggest that a wide range of volatile secondary metabolites are present in essential oils. Although these essential oils originated from different species of a similar family, both supervised (PLS-DA) and unsupervised (PCA) approaches produced a similar outcome on the discrimination pattern where the five essential oils were well discriminated according to their species. Also, it can be deduced that although more than 500 metabolites were identified, not all of them are significant enough to be used as biomarkers. In this context, only 24 metabolites were deemed biomarkers, with VIP values exceeding 2.0. Besides, it is important to note that although both hierarchical and partitional clusterings are well-known for the clustering of samples, different clustering models do not necessarily have similar outputs. Thus, partitional clustering (K-mean) is recommended for future studies as it produces a more accurate result. The combination of metabolite set enrichment analysis with other analyses is essential as it provides more explicit information on the significant class of metabolites in all the samples, like the ketone class in

this study. For future studies that investigate volatile secondary metabolites incorporated with metabolomics, it is recommended to involve validation steps for the supervised models, such as cross-validation and permutation tests. This is because supervised models tend to overfit the data and produce incorrect information on the discrimination pattern. However, the validation steps are not a must in the current study as both supervised and unsupervised models had the same pattern.

Declarations

Ethics Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent Not applicable.

Conflict of Interest Authors have no conflict of interest to declare.

References

1. Grattapaglia D, Vaillancourt RE, Shepherd M, Thumma BR, Foley W, Külheim C, Potts BM, Myburg AA (2012) Progress in Myrtaceae genetics and genomics: Eucalyptus as the pivotal genus. Tree Genet Genomes 8:463–508. https://doi.org/10.1007/s11295-012-0491-x

2. Soliman FM, Fathy MM, Salama MM, Saber FR (2015) Botanical study of Eucalyptus Cinerea F. Muell. Ex Benth. Cultivated in Egypt. World J Pharm Pharm Sci 4:310–328

3. Duarte LDS, Pereira MTM, Pascoal VDÁB, Pascoal ACRF (2020) Campomanesia genus-A literature review of nonvolatile secondary metabolites, phytochemistry, popular use, biological activities, and toxicology. Eclet Quim 45:12–22. https://doi.org/10.26850/1678-4618eqj.v45.2.2020.p12-22

4. Hardstaff LK, Sommerville KD, Funnekotter B, Bunn E, Offord CA, Mancera RL (2022) Myrtaceae in Australia: Use of cryobiotechnologies for the conservation of a significant plant family under threat. Plants 11:1–15. https://doi.org/10.3390/plants11081017

5. Rahman AHMM, Zaman R (2015) Taxonomy and traditional medicinal plant species of Myrtaceae (myrtle) family at Rajshahi district, Bangladesh. Int J Adv Res 3:1057–1066

6. Beltrame BM, Klein-Junior LC, Schwanz M, Henriques AT (2021) Psidium L. Genus: a review on its chemical characterization, preclinical and clinical studies. Phyther Res 35:4795–4803. https://doi.org/10.1002/ptr.7112

7. Science E (2020) Effects of different drying methods on volatile composition of Melaleuca alternifolia essential oil effects of different drying methods on volatile

8. composition of Melaleuca alternifolia essential oil. https://doi.org/10.1088/1755-1315/559/1/012010

8. Mitra SK, Irenaeus TKSS, Gurung MR, Pathak PK (2012) Taxonomy and importance of Myrtaceae. Acta Hortic 959:23–34. https://doi.org/10.17660/ActaHortic.2012.959.2

9. Santos AA, Hancox TJJ, Picanço MC, Delaporte K, Hogendoorn K (2021) Potential distribution of Leptospermum species (Myrtaceae) in Australia for bioactive honey production purposes. New Zeal J Crop Hortic Sci 51:81–92. https://doi.org/10.1080/01140671.2021.1951306

10. Cascaes MM, Guilhon GMSP, de Aguiar Andrade EH, das Graças Bichara Zoghbi M, da Silva Santos L (2015) Constituents and pharmacological activities of Myrcia (Myrtaceae): A review of an aromatic and medicinal group of plants. Int J Mol Sci 16:23881–23904. https://doi.org/10.3390/ijms161023881

11. Butnariu M, Sarac I (2018) Essential oils from plants. J Biotechnol Biomed Sci 1:35–43. https://doi.org/10.14302/issn.2576

12. Dhifi W, Bellili S, Jazi S, Bahloul N, Mnif W (2016) Essential oils' chemical characterization and investigation of some biological activities: a critical review. Medicines 3:25. https://doi.org/10.3390/medicines3040025

13. Goyal MR, Ayeleso AO (2018) Bioactive compounds of medicinal plants, 1st edn. Apple Academic. https://doi.org/10.1201/b10300-9

14. Faleiro JH, Gonçalves RC, dos Santos MNG, da Silva DP, Naves PLF, Malafaia G (2016) The Chemical featuring, toxicity, and Antimicrobial Activity of Psidium cattleianum (Myrtaceae) leaves. New J Sci 2016:1–8. https://doi.org/10.1155/2016/7538613

15. Antonelo FA, Rodrigues MS, Cruz LC, Pagnoncelli MG, Cunha MAA da, Bonatto SJR, Busso C, Wagner Júnior A, Montanher PF, da Cunha MAA, Bonatto SJR, Busso C, Júnior AW, Montanher PF (2023) Bioactive compounds derived from Brazilian Myrtaceae species: Chemical composition and antioxidant, antimicrobial and cytotoxic activities. Biocatal Agric Biotechnol 48:102629. https://doi.org/10.1016/j.bcab.2023.102629

16. Nassar MI, Aboutabl ESA, Ahmed RF, El-Khrisy EDA, Ibrahim KM, Sleem AA (2010) Secondary metabolites and bioactivities of Myrtus communis. Pharmacognosy Res 2:325–329. https://doi.org/10.4103/0974-8490.75449

17. Vivaldo G, Masi E, Taiti C, Caldarelli G, Mancuso S (2017) The network of plants volatile organic compounds. Sci Rep 7:1–18. https://doi.org/10.1038/s41598-017-10975-x

18. Sørensen M, Rinnan R, Woodrow I, Møller BL, Neilson EHJJ (2020) The entangled dynamics of eucalypt leaf and flower volatile emissions. Environ Exp Bot 176:104032. https://doi.org/10.1016/j.envexpbot.2020.104032

19. García YM, Ramos ALCC, Paula ACCFF, De, Nascimento MH, Do, Augusti R, de Araújo RLB, de Lemos EEP, Melo JOF (2021) Chemical physical characterization and profile of fruit volatile compounds from different accesses of myrciaria floribunda (H. West ex wild.) o. Berg through polyacrylate fiber. Molecules 26:1–13. https://doi.org/10.3390/molecules26175281

20. Reis AS, Silva L, de Martins S, de Paula CF (2021) Analysis of the volatile oils from three species of the gender Syzygium. Res Soc Dev 10:e13510716375. https://doi.org/10.33448/rsd-v10i7.16375

21. da Costa JS, Andrade WMS, de Figueiredo RO, Santos PVL, Freitas JJ da S, Setzer WN, da Silva JKR, Maia JGS, Figueiredo PLB (2022) Chemical composition and variability of the volatile components of Myrciaria species growing in the Amazon region. Molecules 27:2234. https://doi.org/10.3390/molecules27072234

22. da Fonsêca DV, da Silva Maia Bezerra, Filho C, Lima TC, de Almeida RN, de Sousa DP (2019) Anticonvulsant essential oils and their relationship with oxidative stress in epilepsy. Biomolecules 9:835. https://doi.org/10.3390/biom9120835

23. Bai L, Wang W, Hua J, Guo Z, Luo S (2020) Defensive functions of volatile organic compounds and essential oils from northern white-cedar in China. BMC Plant Biol 20:500. https://doi.org/10.1186/s12870-020-02716-6

24. Chong J, Wishart DS, Xia J (2019) Using MetaboAnalyst 4.0 for comprehensive and integrative metabolomics data analysis. Curr Protoc Bioinforma 68:1–128. https://doi.org/10.1002/cpbi.86

25. Lim V, Gorji SG, Daygon VD, Fitzgerald M (2020) Untargeted and targeted metabolomic profiling of Australian indigenous fruits. Metabolites 10. https://doi.org/10.3390/metabo10030114

26. Manickavasagam G, Saaid M, Lim V, Saad MIZM, Azmi NAS, Osman R (2023) Quality assessment and chemometrics application on physicochemical characteristics, antioxidant properties, and 5-HMF content of Malaysian stingless bee honey from different topographical origins. J Food Sci 88:1466–1481. https://doi.org/10.1111/1750-3841.16535

27. Manickavasagam G, Saaid M, Lim V (2024) Exploring stingless bee honey from selected regions of Peninsular Malaysia through gas chromatography–mass spectrometry–based untargeted metabolomics. J Food Sci. https://doi.org/10.1111/1750-3841.16903

28. Madala NE, Piater LA, Steenkamp PA, Dubery IA (2014) Multivariate statistical models of metabolomic data reveals different metabolite distribution patterns in isonitrosoacetophenone-elicited Nicotiana tabacum and Sorghum bicolor cells. Springerplus 3:1–10. https://doi.org/10.1186/2193-1801-3-254

29. Elhawary EA, Mostafa NM, Labib RM, Singab AN (2021) Metabolomic profiles of essential oils from selected rosa varieties and their antimicrobial activities. Plants 10. https://doi.org/10.3390/plants10081721

30. Salleh WMNHW, Shaharudin SM (2021) Multivariate statistical analysis of the essential oils of five beilschmiediaspecies from Peninsular Malaysia. Bol Latinoam Y Del Caribe Plantas Med Y Aromat 20:61–70. https://doi.org/10.37360/blacpma.21.20.1.5

31. Camargo KC, Duarte LP, Vidal DM, Pereira HV, Pereira RCG, de Aguilar MG, de Sousa GF, Filho SAV, Mercadante-Simões MO, Messias MCTB, de Oliveira DM (2020) Chemodiversity of essential oils from nine species of celastraceae. Chem Biodivers 17. https://doi.org/10.1002/cbdv.202000107

32. Manickavasagam G, Saaid M, Lim V, Tey ZJ, Osman R (2023) Topographical discrimination of Heterotrigona itama honey by their major-to-trace elements analysis using ICP-OES employed with chemometrics. Chem Pap. https://doi.org/10.1007/s11696-023-02905-1

33. Hur S-J, Lee H-W, Shin A-H, Park SJ (2014) Metabolic perturbation of an Hsp90 C-domain inhibitor in a lung cancer cell line, A549 studied by NMR-based chemometric analysis. J Korean Magn Reson Soc 18:10–14. https://doi.org/10.6564/jkmrs.2014.18.1.010

34. Moghaddam M, Mehdizadeh L (2017) Chemistry of Essential Oils and Factors Influencing Their Constituents, in: Soft Chem. Food Ferment., Elsevier,: pp. 379–419. https://doi.org/10.1016/B978-0-12-811412-4.00013-8

35. Salek RM, Steinbeck C, Viant MR, Goodacre R, Dunn WB (2013) The role of reporting standards for metabolite annotation and identification in metabolomic studies. Gigascience 2:1. https://doi.org/10.1186/2047-217X-2-13

36. Southwell IA, Russell M, Smith RL, Archer DW (2000) Backhousia Citriodora F. Muell. (Myrtaceae), a superior source of citral. J Essent Oil Res 12:735–741. https://doi.org/10.1080/10412905.2000.9712204

37. Yong Lee S, Sik Yang D, Woo Park K, Yi Kim D, Hong Kim J (2011) A multivariate statistical approach to comparison of essential oil composition from three Mentha species. J Hort Sci Technol 29:382–387. https://www.researchgate.net/publication/236734165

38. Abdalla HI (2021) A brief comparison of K-means and agglomerative hierarchical clustering algorithms on small datasets. Proceeding 2021 Int Conf Wirel Commun Netw Appl 623–632. https://doi.org/10.1007/978-981-19-2456-9_64

39. Xia J, Psychogios N, Young N, Wishart DS (2009) MetaboAnalyst: a web server for metabolomic data analysis and interpretation. Nucleic Acids Res 37:652–660. https://doi.org/10.1093/nar/gkp356

40. Clarke S (2009) Essential chemistry for aromatherapy, 2nd edn. Elsevier,

41. Lu L, Cao H, Li H, Zhang H, Li S, Wang J (2022) Diversity and profiles of volatile compounds in twenty-five peppermint genotypes grown in China. Int J Food Prop 25:1472–1484. https://doi.org/10.1080/10942912.2022.2082465

42. Mugao LG, Gichimu BM, Muturi PW, Mukono ST (2020) Characterization of the volatile components of essential oils of selected plants in Kenya. Biochem Res Int 2020:1–8. https://doi.org/10.1155/2020/8861798

43. Kamatou GPP, Vermaak I, Viljoen AM, Lawrence BM (2013) Menthol: a simple monoterpene with remarkable biological properties. Phytochemistry 96:15–25. https://doi.org/10.1016/j.phytochem.2013.08.005

44. Fazmiya MJA, Sultana A, Rahman K, Heyat MB, Bin, Sumbul, Akhtar F, Khan S, Appiah SCY (2022) Current insights on bioactive molecules, antioxidant, anti-inflammatory, and other pharmacological activities of Cinnamomum camphora Linn. Oxid Med Cell Longev 2022:1–23. https://doi.org/10.1155/2022/9354555