Food Research 7 (3) : 247 - 259 (June 2023)
Journal homepage: https://www.myfoodresearch.com

FOOD RESEARCH

FULL PAPER

Role of chemometric classification for future prediction: application on different geographical origins of Jordanian Guava

Abu Mualla, S. and *Al Bakain, R.

Department of Chemistry, School of Science, The University of Jordan, 11942, Amman, Jordan

Article history:
Received: 13 June 2022
Received in revised form: 21 July 2022
Accepted: 6 September 2022
Available Online: 30 June 2023

Keywords:
Guava analysis,
Gas chromatography,
Principal component analysis,
Cluster analysis,
Region's prediction

DOI:
https://doi.org/10.26656/fr.2017.7(3).317

Abstract

In this study, the guava-origin fruits were collected from different cultivated regions in Jordan, then scanned using gas chromatography–mass spectrometry (GC-MS) to reveal the chemical constituents. The chemical contents were then used with the help of multivariate analysis to classify the regions. Guava fruit was collected from; Northern Shouneh-1, Northern Shouneh-2, Madaba, Saham Al-Kfarat, and Southern Shouneh. Hydrodistillation was implemented to extract the essential oils from guava fruits. Comprehensive chemical profiling of the extracted essential oils was achieved using (GC -MS). A total of thirty-eight chemical compounds have been detected and identified with variances from one region to another. Limonene, longifolene, β-copaene, and t-muurolo were found in high concentrations among the other detected compounds. The GC-MS data were subjected to Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) to reveal the similarities/differences between guava fruit regions. The Northern Shouneh-1 and Madaba regions' fruits showed high similarity to each other due to the distinct contents of limonene and longifolene. On the other hand, cadinol was the main compound in Saham Kfarat and Southern Shouneh regions. Finally, Northern Shouneh-2 guava samples showed different content than other regions due to the distinguished levels of t-muurolol. Guava classification based on the GC-MS profile will meet the practical needs of its applications in food production and will contribute to the standardization of commercially available cultivars in Jordan.

1. Introduction

Guava (*Psidium guajava* L.) is considered an important fruit that is widely distributed in tropical and subtropical regions (Batista Silva *et al*., 2017; Yadav *et al*., 2022). Guava can be implanted in Latin America, the Caribbean, Asia, Africa, Australia, and the United States (Moon *et al*., 2018). It adapts to all climatic conditions but prefers dry climates (Naseer *et al*., 2018). Guava can be used in the production of certain marketed foods such as juice, jelly nectar, stuffed candies, gelatins, pastes, tinned and confectionery. In addition, the extracted essential oil contents from guava can be implemented in the treatment of some diseases such as anti-diarrheal, gastroenteritis and stomach problems (Gutiérrez *et al*., 2008; Deguchi and Miyazaki, 2010; Ishartani *et al*., 2018).

Essential oils are extensively found in fresh fruits including guava (Bakkali *et al*., 2008; Gutiérrez *et al*., 2008; Adorjan and Buchbauer 2010). Guava leaves, fruit, and seeds contain a significant amount of essential oil, phenols, tannins, lectins, and vitamins. Many types of flavonoids are present in the guava leaves, especially quercetin. In addition, it has antioxidant properties that are attributed to the polyphenols found in the leaves (Joseph and Priya 2011; Irshad *et al*., 2020). The concentration of the chemical components in guava fruit differs from one region to another depending on the type of species or cultivar as well as cultivation conditions such as soil type, weather, and irrigation system (Porat *et al*., 2011; Irshad *et al*., 2020).

Fruit quality is associated with management and climatic conditions during the production phase (Arroyo *et al*., 2020). Guava has significantly impacted the plant accession, growing environment, and cultural practices (Moon *et al.,* 2018). It is the hardiest among tropical fruit trees and excels most of the other fruit crops in productivity and adaptability, whereas the adult trees of guava can survive temperatures down to -4°C, but the younger plants are vulnerable to freezing in the condition of low temperature (Joseph and Priya 2011; Irshad *et al*.,

*Corresponding author.
Email: *r.bakain@ju.edu.jo*

eISSN: 2550-2166 / © 2023 The Authors.
Published by Rynnye Lyan Resources

2020). It is susceptible to chilling injury when stored below 10°C. Thus, the spoilage and losses of this horticultural produce require attention during transportation and marketing (Nair *et al*., 2018).

Extraction of the essential oils from natural solid matrices of plants, such as guava fruits can be achieved through several techniques including steam distillation, hydrodistillation (Gavahian and Farahnaky 2018), soxhlet extraction and cold pressing (Kokolakis and Golfinopoulos 2013). Among these methods, hydrodistillation is the most widely used extraction process to extract essential oils from plants (Al-Hyasat *et al.,* 2021). Distillation is frequently achieved by prolonged heating and stirring in water or solvent using a Clevenger apparatus (Kokolakis and Golfinopoulos 2013). Many studies have used this technique to obtain volatile oils from different plant sources (Silva *et al*., 2005; Fasola *et al.,* 2011; Al-Hyasat *et al.,* 2021).

Gas Chromatography coupled with Mass Spectrometry (GC-MS) is the most commonly used technique for analyzing essential oils since it could; separate complex mixtures into their individual components, identify and quantify them easily (Fasola *et al.,* 2011; Al Bakain *et al*., 2020), and provide comprehensive chromatographic profiling of the whole chemical contents (Al Bakain *et al*., 2021a).

The multivariate method includes Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) are useful tools in analytical chemistry for classifications. These tools were performed in analytical studies for many purposes; to identify the most relevant compounds in distinguishing plant varieties, to find the variation in chemical profiles as a result of growing plants in different batches and with variations in growth time, to confirm whether the cultivars in the cluster analysis would also be grouped together, to reveal the compounds that were responsible for grouping cultivars between clusters and to predict the geographical origin of the sample using linear discriminant analysis (Al Bakain *et al*., 2021b).

There is currently no available systematic clustering for the guava-origin fruits in Jordan, which is necessary to explore the similarities/differences if any among plant -origin samples. Therefore, the present study was carried out to fingerprint the chemical profile of the essential oils in guava fruit planted in different regions in Jordan. For the first time, guava plant samples obtained from the five cultivars in Jordan were collected, extracted and analyzed using GC-MS. The fruits were fully scanned to detect all the chemical compounds. The validated method was evaluated for selectivity and precision. GC-MS data have been subjected to PCA and HCA to reveal the significance of the chemical contents for guava-based origin classification, to confirm whether the cultivars in the cluster analysis would also be grouped together and to reveal the compounds that were responsible for grouping cultivars between clusters. To the best of our knowledge, this study is the first to classify the Jordanian guava samples and to use the PCA and HCA to predict the region for future Jordanian samples analyzed blind.

2. Materials and methods

2.1 Guava samples

Guava fruits were harvested in the autumn season at the end of September to the beginning of October 2019, at the onset of ripening from the cultivars in Jordan. The five Jordanian cultivators where guava was planted were Northern Shouneh-1, Northern Shouneh-2, Madaba, Saham Al-Kfarat and Southern Shouneh. From each cultivar, fresh guava fruits were collected and immediately transferred in clean bags to the laboratory for sample processing and chemical analysis.

2.2 Chemicals and reagents

All reagents used in this study were of HPLC high purity analytical grade: n-hexane (99.5%, TEDIA, USA), HPLC water (Sigma Aldrich, USA) and sodium sulfate anhydrous (Na_2SO_4) (100%, Biochem Chemopharma, France).

2.3 Essential oils extraction

The extraction of guava essential oils was carried out by hydrodistillation using a Clevenger-type apparatus according to the following procedure. After collecting the guava samples from the cultivars, 2 kg of washed guava fruit was inserted into stainless steel blender container. The content has been mixed with 2.0 L of HPLC water until homogenous, and then, immediately, the mixture was transferred into 5.0 L Clevenger apparatus (Chitransh borosilicate, India) and then finally, 1.0 L of HPLC water was added. Hydrodistillation extraction was implemented for 4 hrs at a temperature between 80-90°C. Guava essential oils were collected in 10.0 mL flask, and then about 10-20 mg of anhydrous sodium sulphate was added to eliminate any remaining water in the extracted oil. A concentration of 0.0015% (v/v) extracted essential oil/n-hexane was prepared for the GC-MS injection. Finally, the extract was filtered using the Acrodisc syringe filter (PAU-Gelman Lab, 0.45 mm, 25 mm diameter) membrane micro-filter and collected in a screw-capped amber vial of 1.5 mL, stored in a freezer (at least -10°C) prior to analysis. The extractions and the GC-MS injections were carried out in triplicate. The injection volume was 1.0 mL. The content of each essential oil was reported as an average of the

https://doi.org/10.26656/fr.2017.7(3).317 © 2023 The Authors. Published by Rynnye Lyan Resources

triplicate injections. Quantification of guava chemical content for each compound was achieved by internal normalization, by comparing the ratio of the peak area of the component in the sample to the summation of all peak areas in the guava sample.

2.4 Gas chromatography-mass spectrometer detector

The injection of 1.0 µL sample into the GC-MS instrument was carried out in triplicate. The GC-MS chemical profiles of the guava extracts were all generated using a VARIAN Model CP-3800 Gas Chromatographic system coupled with Saturn 2200 GC/MS/MS, CP-8400 autosampler. The GC capillary column was SLB®-5ms fused silica, 30 m length × 0.25 mm internal diameter, 0.25-µm film thickness (Sigma-Aldrich, USA). The front injector temperature was set at 250°C. The temperature was programmed in a linear generic gradient mode from 60°C to 250°C at a rate of 3° C/min, then held for 2 mins at 250°C. The total overall runtime was 65 min/sample. The carrier gas was helium of 99.999% purity at a constant flow rate of 1.0 mL/min. Solutes detection was achieved using a mass spectrometer with a scan range from 35-500 m/z, and samples were injected in a split ratio mode of 1:10 with an ionization source of 180°C and ionization voltage of 70 eV.

2.5 Identification and quantification of the extracted essential oils

Selectivity was determined by injecting solvent blank to confirm that there were no artifact signal peaks at the targeted retention time. The quantitative analysis for each essential oil component expressed in a content% was calculated according to the following estimation:

$$\text{Chemical content \%} = \frac{\text{Peak area of the component}}{\text{Summation of all peak areas}} \times 100\% \quad (1)$$

A hydrocarbon mixture of n-alkanes standard (C_8-C_{20}) was analyzed separately on GC-MS under the same chromatographic conditions of the sample analysis. The compounds were identified (i.e. name, molecular weight and structure) by comparing the retention indices relative to (C_8-C_{20}) alkanes standard and with the MS computer library using the database of National Institute Standard and Technology (NIST 05 MS Library and AMDIS 2.6).

2.6 Multivariate analysis of guava samples

The main objectives for applying multivariate analysis in analytical chemistry are grouping, visualization, clustering, discriminating, and classifying objects (samples, compounds, or materials) with modeling relationships among different analytical data. In grouping or clustering, samples would be grouped according to their chemical composition, elemental pattern, or technological properties (Al-Hyasat *et al*., 2021). While in classification, samples would be classified into known class membership based on their elemental pattern, chemical composition, or spectra (Al-Hyasat *et al*., 2021). Generally, the main methodology adopted for the purpose of grouping or clustering is known as the unsupervised method. For unsupervised methods, the grouping of analytical data is carried out by projecting the high dimensional data into a lower dimensional space and since there is no supervisor to relate the membership of samples to classes, then unsupervised methods are performed in a supervised manner (Al-Hyasat *et al*., 2021). Both PCA and HCA are typical examples of unsupervised methods. HCA and PCA are performed to confirm if the guava obtained from different regions would be grouped together based on the GC-MS measured active ingredients including volatile oils with a total number of 38 solutes (38 detected peaks). PCA can reveal the compounds that are responsible for grouping guava samples. Guava classification will be the foundation for food production and informative guidance for individual growers. Indeed, these results would help to show the similarities and differences between guava-origin samples in Jordan.

2.6.1 Arrangement of chromatographic data

In general, chromatographic data can be arranged as a data matrix X of n samples or location rows and *l* variables or compounds. For the current case, matrix X has the size of 38×5 or 5 samples (obtained from 5 regions) × 38 detected compounds. Matrix X was subjected to HCA and PCA as discussed below.

2.6.2 Quantitative guava classification by Principal Component Analysis and Hierarchical Clustering Analysis

These tools are powerful for the interpretation of large data tables. PCA is a data compression method based on the correlation among variables. It aims to group those correlated variables, while replacing the original ones with a new set, called the principal components (PCs), onto which the data are projected (Al -Bakain *et al.,* 2020; Al-Hyasat *et al*., 2021). These PCs are completely uncorrelated and are built as a simple linear combination of the original variables. It is important to point out here that the first PCs contain most of the variability in the data set, albeit in a much lower dimensional space. The first principal component, PC1, is defined in the direction of maximum variance of the whole data set. PC2 is the direction that describes the maximum variance in the orthogonal subspace to PC1. The subsequent components are taken orthogonal to those previously chosen, and describe the maximum of

https://doi.org/10.26656/fr.2017.7(3).317

© 2023 The Authors. Published by Rynnye Lyan Resources

the remaining variance (Al-Hyasat *et al*., 2021). Once the redundancy is removed, only the first few principal components are required to describe most of the information contained in the original data set. In this study, two data matrices were built. The data matrix X (38×5) (i.e. 38 compounds from 5 regions) is decomposed into two matrices, T (score matrix) and L (loading matrix) using a suitable PCA algorithm. The first step in PCA is the computation of loadings. Mathematically, the loadings are the Eigenvectors of the matrix (XXT). There are several methods to estimate the eigenvectors, such as singular value decomposition (SVD) and NIPALS (non-linear iterative partial least-squares) in the order of explained proportion of the variations in X, until a certain pre-established number of components. The loadings are grouped into a matrix L. The collected loadings are orthonormal, meaning that they are both orthogonal and normalized (Al Bakain *et al*., 2011; Al-Bakain *et al*., 2012). The relationship between the: original matrix X, loading matrix L and score matrix T is described as:

$$X = TLT \quad (2)$$

Mathematically, matrix X is decomposed in the product of two matrices, T and L, on the condition that L is formed by orthonormal columns. T is obtained as: T = XTL. In this work, the size of X is 38×38 while size T is 38×*h* and L is *h*×38, where *h* is the number of factors needed to decompose matrix X. The optimum number of factors (*h*) is necessary to create an optimum number of loadings and scores and produce informative discrimination among samples/guava.

Regarding HCA, the main strategy of unsupervised methods is based on cluster analysis where the guava samples are aggregated stepwise according to the similarity of their features or variables (i.e., contents) (Al -Hyasat *et al*., 2021). As a result, hierarchically ordered clusters are created. In HCA, the collected data is shown in a certain way to emphasize their natural clusters and patterns in a two-dimensional space. The results are often presented in the form of a dendrogram which allows quick visualization of clusters and correlations among tested samples. The similarity between samples can be evaluated following suitable distance measures commonly applied in pattern recognition. Euclidean distance d between samples is estimated as:

$$d_{i,k} = \left[\sum_{k=}^{K} (x_{i,k} - x_{j,k})^2\right]^{1/2} \quad (3)$$

Where K and i/j are the numbers of variables measured for samples and indices for samples, respectively. Estimations would be made using the main principal components of the original data after decomposition by PCA. Initially, $d_{i,k}$ is estimated between all samples (i.e., every sample is to be compared with the remaining samples) to create the distance matrix (Al-Hyasat *et al*., 2021). The similarity or aggregation between samples is then estimated using the weighted average linkage method. In order to classify guava samples-origin and to specify the compounds responsible for clustering, the GC-MS scan data were subjected to HCA and PCA analysis. PCA and HCA are commonly employed to reduce the complexity of multivariate data sets without losing important information, observe the variance in data sets, and visualize data clustering (Al Bakain *et al*., 2020). In this study, 38 chemical compounds are the original variables (38 dimensions) in PCA. By calculating the covariance matrix between these 38 dimensions, PCA can generate 38 PCs that are orthogonal to each other and can explain 100% of the total variance of the orthogonal data. Each PC is correlated with the original 38 variables. All detected chemical contents were rather necessary for guava fruit clustering. Accordingly, the number of variables used in clustering was 38 (detected compounds) × 5 (regions).

2.7 Data treatment

Data acquisition and processing were conducted using MS Workstation software version 6.6 (SP1). The statistical analysis including PCA and HCA was performed using Chemoface 1.61 software which works under Matlab® (Mathworks, 8.6, USA) and *XL*STAT software (Excel, Microsoft®). The peak area ratio was calculated for each solute (peak ratio = peak area of solute /summation of peak areas of the sample) and further used to quantify the content of the detected solute.

3. Results and discussion

3.1 GC analysis and variability of guava contents in different samples

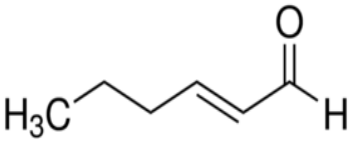
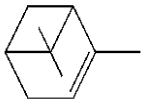
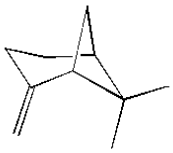
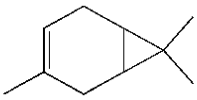
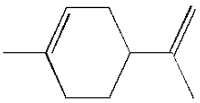
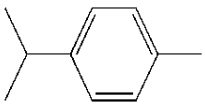
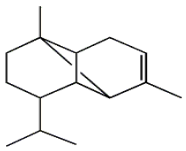
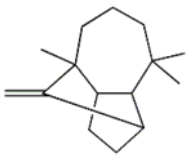
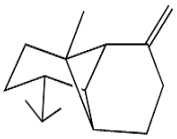
Guava fruit samples were collected from the five different plant cultivars in Jordan; Northern Shouneh-1 (NS1), Northern Shouneh-2 (NS2), Madaba (MA), Saham Al-Kfarat (SK), and Southern Shouneh (SS). The guava fruits essential oils contents have been screened using GC-MS. Thirty-eight compounds were detected in this study and identified in the guava samples over the 5 locations as shown in Table 1.

The same procedure was applied for all guava samples, therefore the comparison between all fruit samples is valid. The 38 detected guava compounds had variable levels and appeared within 40 mins. Figure 1, presented a model example of GC-MS chromatogram of Northern Shouneh-2-origin guava. GC-MS analysis in Figure 1 showed that limonene, longifolene, ß-copaene,

https://doi.org/10.26656/fr.2017.7(3).317

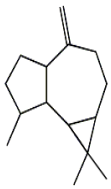
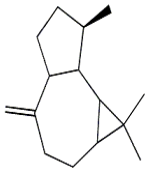
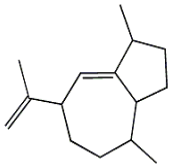
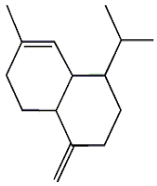
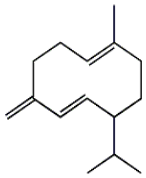
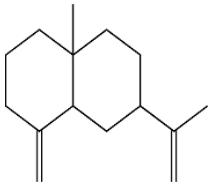
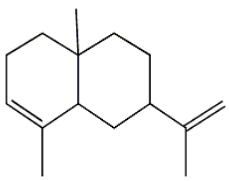
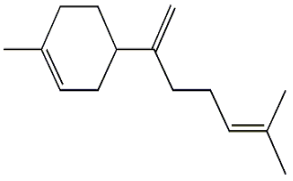
© 2023 The Authors. Published by Rynnye Lyan Resources

Table 1. The 38 solutes identified in guava fruits samples obtained from the five cultivars in Jordan.

No.	Common name	Structure	Molecular formula	Molecular mass (g/mol)	Retention time (min)
1.	(E)-2-Hexenal		$C_6H_{10}O$	98.14	4.5
2.	α-Pinene		$C_{10}H_{16}$	136.24	6.3
3.	β-Pinene		$C_{10}H_{16}$	136.24	7.9
4.	β-Carene		$C_{10}H_{16}$	136.24	8.5
5.	Limonene		$C_{10}H_{16}$	136.24	9.5
6.	p-Cymene		$C_{10}H_{14}$	134.22	9.6
7.	α-Copaene		$C_{15}H_{24}$	204.36	23.9
8.	Longifolene		$C_{15}H_{24}$	204.36	25.9
9.	β-Copaene		$C_{15}H_{24}$	204.36	26.6

https://doi.org/10.26656/fr.2017.7(3).317
© 2023 The Authors. Published by Rynnye Lyan Resources

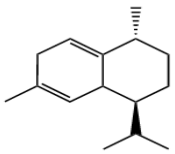
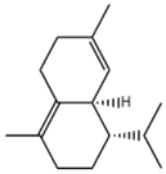
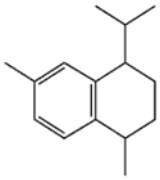
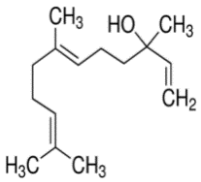
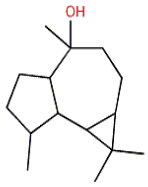
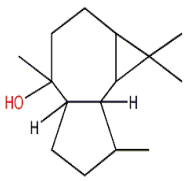
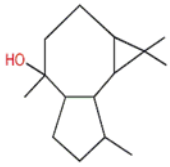
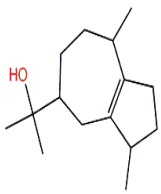
Table 1 (Cont.). The 38 solutes identified in guava fruits samples obtained from the five cultivars in Jordan.

No.	Common name	Structure	Molecular formula	Molecular mass (g/mol)	Retention time (min)
10.	Aromadendrene		$C_{15}H_{24}$	204.36	27.2
11.	Alloaromadendrene		$C_{15}H_{24}$	204.36	27.4
12.	γ-Gurjunene		$C_{15}H_{24}$	204.36	27.9
13.	γ-Muurolene		$C_{15}H_{24}$	204.36	28.6
14.	Germacrene D		$C_{15}H_{24}$	204.36	28.7
15.	β-Selinene		$C_{15}H_{24}$	204.36	28.9
16.	α-Selinene		$C_{15}H_{24}$	204.36	29.1
17.	β-Bisabolene		$C_{15}H_{24}$	204.36	29.4

https://doi.org/10.26656/fr.2017.7(3).317

© 2023 The Authors. Published by Rynnye Lyan Resources

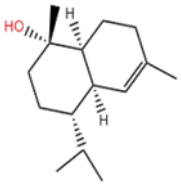
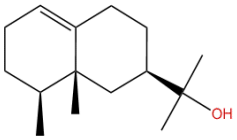
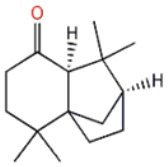
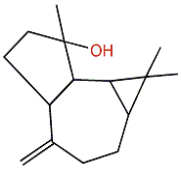
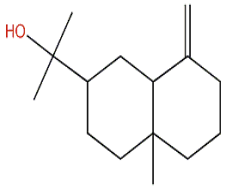
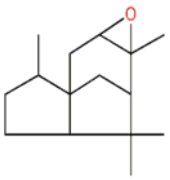
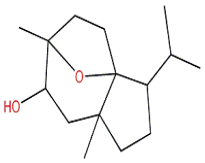
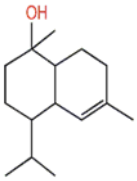
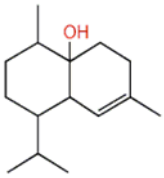
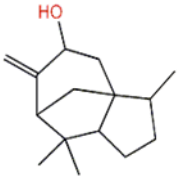
Table 1 (Cont.). The 38 solutes identified in guava fruits samples obtained from the five cultivars in Jordan.

No.	Common name	Structure	Molecular formula	Molecular mass (g/mol)	Retention time (min)
18.	Cadina-1,4-diene		$C_{15}H_{24}$	204.36	29.8
19.	δ-Cadinene		$C_{15}H_{24}$	204.36	30.0
20.	α-Calamenene		$C_{15}H_{22}$	202.33	30.4
21.	Trans-nerolidol		$C_{15}H_{26}O$	222.37	31.6
22.	Ledol		$C_{15}H_{26}O$	222.37	31.7
23.	Globulol		$C_{15}H_{26}O$	222.37	31.9
24.	Viridiflorol		$C_{15}H_{26}O$	222.37	32.5
25.	Guaiol		$C_{15}H_{26}O$	222.37	32.6

https://doi.org/10.26656/fr.2017.7(3).317

© 2023 The Authors. Published by Rynnye Lyan Resources

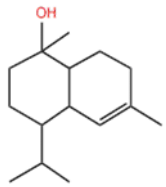
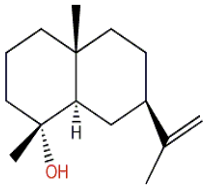
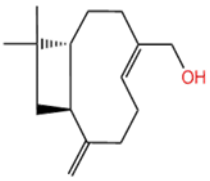
Table 1 (Cont.). The 38 solutes identified in guava fruits samples obtained from the five cultivars in Jordan.

No.	Common name	Structure	Molecular formula	Molecular mass (g/mol)	Retention time (min)
26.	t-Muurolol		$C_{15}H_{26}O$	222.37	33.0
27.	Eremoligenol		$C_{15}H_{26}O$	222.37	33.3
28.	Cis-Isolongifolanone		$C_{15}H_{24}O$	220.35	33.4
29.	Spathulenol		$C_{15}H_{24}O$	220.35	33.6
30.	β-Eudesmol		$C_{15}H_{26}O$	222.37	33.8
31.	Cedrene epoxide		$C_{15}H_{24}O$	220.35	34.1
32.	Daucol		$C_{15}H_{26}O_2$	238.37	34.5
33.	Torreyol		$C_{15}H_{26}O$	222.37	34.7
34.	Cubenol		$C_{15}H_{26}O$	222.37	34.7
35.	Cedrenol		$C_{15}H_{24}O$	220.35	34.8

https://doi.org/10.26656/fr.2017.7(3).317

© 2023 The Authors. Published by Rynnye Lyan Resources

Table 1 (Cont.). The 38 solutes identified in guava fruits samples obtained from the five cultivars in Jordan.

No.	Common name	Structure	Molecular formula	Molecular mass (g/mol)	Retention time (min)
36.	Cadinol		$C_{15}H_{26}O$	222.37	34.9
37.	Kongol		$C_{15}H_{26}O$	222.37	35.3
38.	14-Hydroxy-caryophyllene		$C_{15}H_{24}O$	220.35	35.4

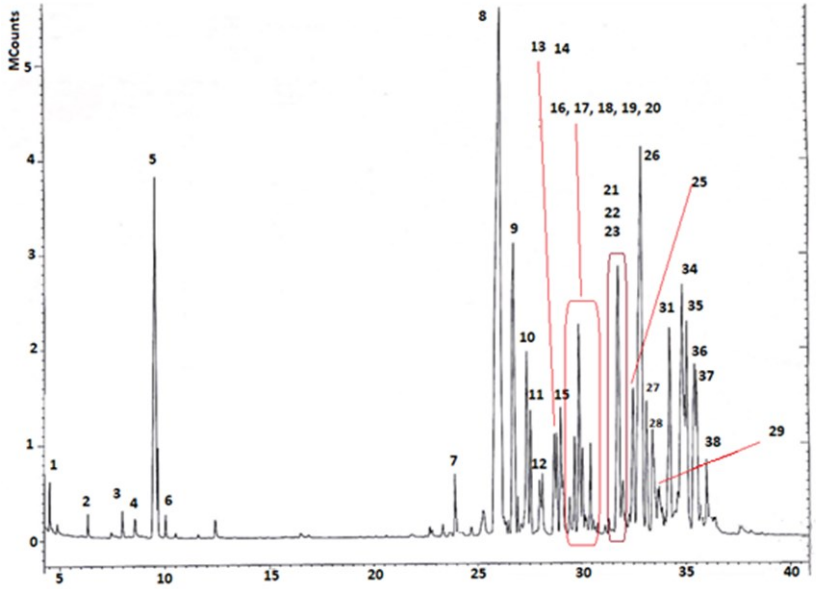


Figure 1. GC-MS of Northern Shouneh-2-origin guava.

cadina 1, 4-diene, globulol, guaiol, t-muurolol, cedrene epoxide, cedrenol and 14-hydroxy-caryophyllene have high intensity of a content% > 3%.

The precision was evaluated as intra-day precision and expressed by the relative standard deviation (RSD%) of the peak area variations of limonene as tested analyte in repeated samples. The precision was measured by the analysis of five replicates in a single day. The precision (%RSD) was 4.30%. The chemical composition % of guava (peak area%) is shown in Table 2.

3.2 Comparison between guava chemical contents of Jordanian samples

A total of thirty-eight compounds were identified in guava essential oils. The detected compounds were related to four different classes: monoterpene hydrocarbons (5 compounds), aldehyde (1 compound), sesquiterpenes hydrocarbons (14 compounds) and oxygenated sesquiterpenes (18 compounds). The main five solutes of the guava essential oil that contained the highest contents levels at all regions, as shown in Table 2, were: limonene (6.7 - 26.2%), longifolene (18.8 - 31.1%), guaiol (3.2 - 8.0%), ß-copaene (2.8 - 6.6%) and cadinol (2.0 - 11.7%). Our results were in agreement with other studies on guava fruits contents from Brazil, India, Egypt, and France (Paniandy *et al*., 2000; Biegelmeyer *et al*., 2011; El-Ahmady *et al*., 2013; Borah *et al*., 2021). El-Ahmady found that limonene was the major content in the Egyptian guava in comparison to the other compounds as reported in our study (El-Ahmady *et al*., 2013). Paniandy *et al*. (2000) have reported that β-copaene was a distinct compound in French guava as highlighted in our results. Cadinol was mentioned as the

https://doi.org/10.26656/fr.2017.7(3).317
© 2023 The Authors. Published by Rynnye Lyan Resources

Table 2. Identification of compounds extracted from the essential oil of Jordanian guava fruit.

Name of Compound	Retention time (Rt)/min	Retention index (RI)		Percentages of Compounds %				
		Calculated	Reported	NS1%	NS2%	MA%	SK%	SS%
(E)-2-Hexenal	4.5	843	852	ND	0.6	1.2	ND	0.8
α-Pinene	6.3	928	929	0.5	0.2	4.1	ND	ND
β-Pinene	7.9	984	979	0.6	0.3	0.7	ND	ND
β-Carene	8.5	1003	1004	0.8	1.4	0.5	0.7	0.4
Limonene	9.5	1030	1033	19.3	6.7	26.2	9.5	8.0
p-Cymene	9.6	1032	1032	3.1	2.4	3.9	2.6	4.4
α-Copaene	23.9	1374	1377	2.4	0.6	0.8	1.0	0.7
Longifolene	25.9	1418	1410	31.1	22.4	18.8	20.9	19.0
β-Copaene	26.6	1438	1432	5.6	6.6	3.6	2.8	3.3
Aromadendrene	27.2	1455	1445	3.8	2.6	2.3	2.7	2.5
Alloaromadendrene	27.4	1459	1460	0.8	1.1	0.7	0.6	0.4
γ-Gurjunene	27.9	1472	1460	ND	0.2	0.7	1.2	ND
γ-Muurolene	28.6	1488	1481	4.6	0.4	3.4	4.7	2.4
Germacrene D	28.7	1491	1485	0.7	0.3	0.4	ND	0.4
β-Selinene	28.9	1496	1495	4.0	0.9	2.7	4.3	2.3
α-Selinene	29.1	1499	1505	1.6	1.1	0.6	1.8	3.0
β-Bisabolene	29.4	1507	1512	1.9	1.2	0.5	1.6	2.4
Cadina-1,4-diene	29.8	1518	1518	1.5	3.2	1.2	1.0	1.7
δ-Cadinene	30.0	1522	1521	ND	1.0	0.5	ND	ND
α-Calamenene	30.4	1534	1525	ND	0.9	0.5	ND	0.3
Trans-nerolidol	31.6	1564	1564	2.3	0.6	1.2	5.3	3.0
Ledol	31.7	1566	1565	0.9	0.6	2.7	ND	5.6
Globulol	31.9	1572	1575	ND	5.4	0.2	0.8	0.6
Viridiflorol	32.5	1588	1590	1.8	ND	4.1	5.9	5.1
Guaiol	32.6	1591	1595	4.1	3.2	5.3	5.6	8.0
t-Muurolol	33.0	1599	1608	0.6	12.3	0.9	1.1	1.9
Eremoligenol	33.3	1608	1606	0.4	1.9	0.5	0.8	0.6
Cis-Isolongifolanone	33.4	1612	1606	0.4	1.4	0.5	1.0	1.5
Spathulenol	33.6	1616	1619	ND	1.2	0.2	ND	0.4
β-Eudesmol	33.8	1622	1624	0.4	ND	0.3	1.4	1.1
Cedrene epoxide	34.1	1631	1631	1.1	3.7	1.3	2.3	1.1
Daucol	34.5	1642	1640	ND	ND	0.6	1.0	1.9
Torreyol	34.7	1645	1645	ND	ND	1.7	2.1	2.0
Cubenol	34.7	1647	1646	1.3	0.6	0.6	3.4	2.0
Cedrenol	34.8	1650	1651	ND	4.8	ND	1.0	1.1
Cadinol	34.9	1652	1652	3.6	2.0	2.6	11.7	11.3
Kongol	35.3	1662	1655	0.6	2.5	0.5	1.3	1.0
14-Hydroxy-caryophyllene	35.4	1666	1661	ND	6.2	3.9	ND	ND

ND, Not detected

main component in the guava fruit collected from India (Borah *et al*., 2021). Regarding the Brazilian guava, limonene, cadinol, and ß-copaene were observed at high levels and considered as main guava constituents (Biegelmeyer *et al*., 2011).

t-Muurolol has a distinct level of in the Northern Shouneh-2 region of a value of 12.3%, where cadinol has big differences in contents% between samples, where Southern Shouneh and Saham Al-Kfarat have high levels of cadinol in comparison to the other samples. Guaiol and longifolene in our study were found specifically in high contents in the Jordanian guava essential oils, but not in the previous studies. Hence, this result could be highlighted as a fingerprint for the Jordanian guava fruit. Regarding the variation of the fruits themselves, within plants, it may vary significantly from one growing region to another and from plant to plant within the same growing area due to the mineral contents of the groundwater, soil, growing medium and air pollution. The variations in contents in the studied guava would be necessary to obtain good clustering objectives as will be appeared in the next sections.

https://doi.org/10.26656/fr.2017.7(3).317

© 2023 The Authors. Published by Rynnye Lyan Resources

The distribution of the contents according to their retention times which mainly affected by two main factors: volatility and polarity. It was noticed that any peak before 3 mins is mainly for the solvent; n-hexane that is considered very volatile and non-polar. The compounds detected between 4-10 mins are generally nonpolar or have the highest volatility or both characteristics. Compounds detected between 10-20 mins are moderately nonpolar or moderately volatile. Finally, late-detected compounds have the lowest volatility with the highest polarity (Al-Hyasat *et al*., 2021).

The quantitative and qualitative outcomes showed that many differences in the percentage levels of each content are referred to many factors: 1) geographic variations; climate, seasonal changes, humidity, and nurture of regions, 2) environmental variations; soil composition, and pH of the soil, 3) type of organic fertilizer and insecticide, 4) farm plants around, 5) the method of watering the plants and the number of times (Bakkali *et al.* 2008). Moreover, genetic factors may affect the secondary metabolism, climate, soil composition, plant organ, age, and vegetative cycle stage, harvesting time of the biological material in different development stages, interaction with microorganisms, insects, and different post-harvest techniques (Bakkali *et al.* 2008).

3.3 Guava characterization by Principal Component Analysis and Hierarchical Clustering Analysis

The GC-MS scan data were subjected to HCA and PCA analysis. PCA and HCA are commonly employed to reduce the complexity of multivariate data sets without losing important information, observe the variance in data sets, and visualize data clustering. The number of variables used in clustering was 38 (detected compounds) × 5 (regions). The resulting HCA clustering of states is provided in Figure 2.

As indicated in Figure 2, three main clusters collect the 5 regions of Jordan according to guava essential oil contents. Cluster A collects Madaba and Northern Shouneh-1 regions and Cluster B collects only Northern Shouneh-2, and cluster C has Saham Al-Kfarat and Southern Shouneh. HCA analysis revealed similar and/or comparable contents of guava fruit samples. Accordingly, guava samples obtained from Northern Shouneh-2 are significantly different from the rest of the samples in other regions (Figure 2). The outcomes are related to the distinct constituents of t-muurolol than other regions as shown in Figure 3.

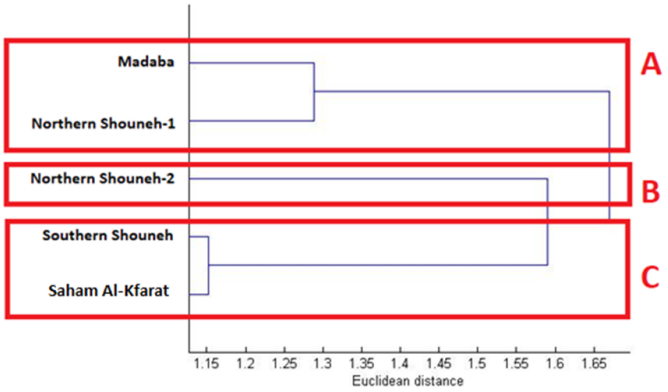


Figure 2. Dendrogram of the 5 guava-origin samples contents.

Regarding the PCA outcomes (Figure 3), PC1 describe 42.25% of the total variance of the data set and has positive loading for only Northern Shouneh-2 samples (Figure 3-A). On the other hand, PC2 and PC3 have 29.75% and 18.88%, respectively of the variance of the data set with a major contribution for the loading of the other 4 regions. These 3 PCs collect 90.88% of the total data variance.

Figure 3-B reveals the most significant oil contents for grouping. Limonene, longifolene, cadinol and t-muurolol were not correlated with other compounds and were more important for guava origin grouping. This result is highlighted in Table 2 for the highest content marker compounds in the studied guava fruit samples.

As shown in the biplot outcome (Figure 3-C), limonene and longifolene were responsible to separate

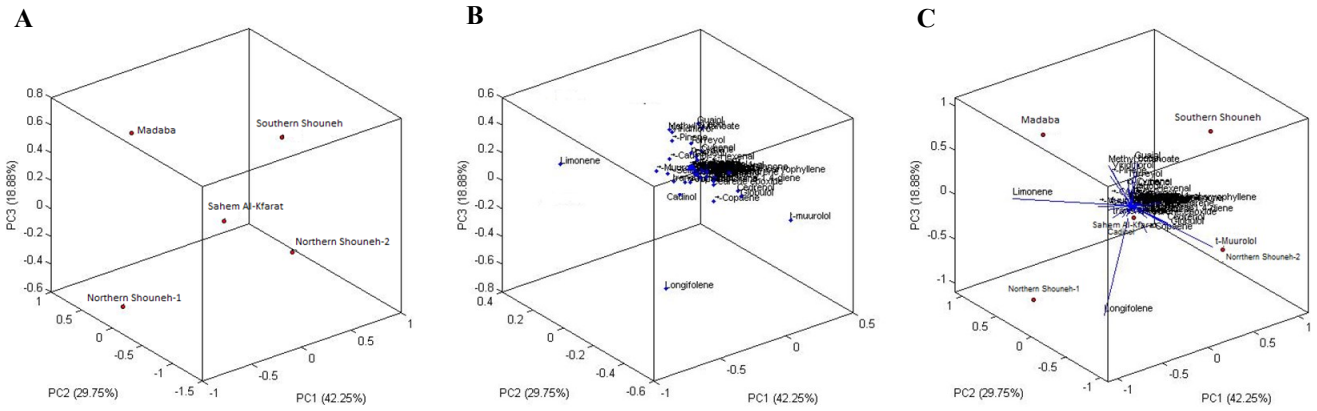


Figure 3. PCA outputs: A) score, B) loading and C) bi-plots obtained from guava fruit oils components from the five Jordanian regions.

https://doi.org/10.26656/fr.2017.7(3).317
© 2023 The Authors. Published by Rynnye Lyan Resources

Madaba and Northern Shouneh-1 guava-origin samples far from the other samples which were proved in Table 2 since these compounds had the highest levels in these two regions in comparison the other regions. t-muurolol was of a large impact to separate Northern Shouneh-2 than other regions due to its large distinct level in this region with 12.30% content, whereas in the other regions', t-muurolol content varied between 0.6%- 1.9%. Regarding Southern Shouneh and Saham A-Kfarat, the distinct constituent of cadinol was the reason behind the separation of these regions together. The rest of the contents do not show any impact or effect on classifications since they showed comparable contents in the guava essential oil in the five regions. In fact, PCA gives an intuitive explanation of where to find the longer radial separation of the compound from the centre (meaning that the most important constituents are in the region). In this work, limonene, longifolene, cadinol and t-muurolol were the most significant contents for regions classification.

4. Conclusion

The five guava-origin fruits were scanned using GC-MS to reveal the chemical constituents. The chemical contents were then used with the help of multivariate analysis to classify the regions. Three resulted clusters were obtained: the first cluster consisted of Madaba and Northern Shouneh-1 which was related to the limonene and longifolene, the second cluster had only the Northern Shouneh-2 due to the distinct content of t-muurolol, and finally, the third cluster of Southern Shouneh and Sahem Al-Kfarat that have different content of cadinol than the other regions. The results were fairly efficient since classifications were based intensively on all constituents considering all compounds in guava.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

Prof. Ramia Al Bakain gratefully acknowledges the financial support provided to this project by (1) L'Agence Universitaire de la Francophonie-Moyen Orient (AUF - Moyen Orient), (2) Elsevier Foundation and 3) the University of Jordan. Indeed, great appreciation goes to Mr. Ismail Abaza for his experimental and technical support.

References

Adorjan, B. and Buchbauer, G. (2010). Biological properties of essential oils: an updated review. *Flavour and Fragrance Journal*, 25(6), 407-426. https://doi.org/10.1002/ffj.2024.

Al Bakain, R., Rivals, I., Sassiat, P., Thiébaut, D., Hennion, M-C., Euvrard, G. and Vial, J. (2011). Comparison of different statistical approaches to evaluate the orthogonality of chromatographic separations: application to reverse phase systems. *Journal of Chromatography A*, 1218(20), 2963–2975. https://doi.org/10.1016/j.chroma.2011.03.031.

Al Bakain, R., Rivals, I., Sassiat, P., Thiébaut, D., Hennion, M-C., Euvrard, G. and Vial, J. (2012). Impact of the probe solutes set on orthogonality evaluation in reverse phase chromatographic systems. *Journal of Chromatography. A,* 1232, 231–241. https://doi.org/10.1016/j.chroma.2011.12.056.

Al Bakain, R., Al-Degs, Y., Cizdziel, J. and Elsohly, M. (2021a). Comprehensive chromatographic profiling of cannabis from 23 USA States marketed for medical purposes. *Acta Chromatographica*, 33(1), 78-90. https://doi.org/10.1556/1326.2020.00767.

Al Bakain, R., Al-Degs, Y., Cizdziel, J. and Elsohly, M. (2020). Comprehensive classification of USA cannabis samples based on chemical profiles of major cannabinoids and terpenoids. *Journal of Liquid Chromatography and Related Technologies*, 43(5-6), 172-184. http:// doi: 10.1080/10826076.2019.1701015.

Al Bakain, R., Al-Degs, Y., Cizdziel, J. and Elsohly, M. (2021b). Linear discriminant analysis based on gas chromatographic measurements for geographical prediction of USA medical domestic cannabis. *Acta Chromatographica*, 33(2), 179-187. https:// doi.org/10.1556/1326.2020.00782.

Al-Hyasat, M., Al Bakain, R., Vial, J. and Thiebaut, D. (2021). Geographical Prediction of Parsley Origins Based on Chromatographic Fingerprinting and Quantitative Analysis: Application to Mediterranean Cultivars. *American Journal of Chemistry*, 11(4), 66-80. http://doi:10.5923/j.chemistry.20211104.02.

Arroyo, B.J., Bezerra, A.C., Oliveira, L.L., Arroyo, S.J., Melo, E.A. and Santos, A. (2020). Antimicrobial active edible coating of alginate and chitosan add ZnO nanoparticles applied in guavas (*Psidium guajava L.*). *Food Chemistry*, 309, 125566. https:// doi.org/10.1016/j.foodchem.2019.125566.

Bakkali, F., Averbeck, S., Averbeck, D. and Idaomar, M. (2008). Biological effects of essential oils-a review. *Food and Chemical Toxicology*, 46(2), 446–475. https://doi.org/10.1016/j.fct.2007.09.106.

Batista Silva, W., Cosme Silva, G.M., Bortolini Santana, D., Rodrigues Salvador, A., Barbosa Medeiros, D., Belghith, I., Martins da Silva, N., Menezes Cordeiro, M.H. and Polete Misobutsi, G. (2017). Chitosan delays ripening and ROS production in guava (*Psidium guajava L.*) fruit. *Food Chemistry*, 242,

 © 2023 The Authors. Published by Rynnye Lyan Resources

232-238. http://doi: 10.1016/ j.foodchem.2017.09.052.

Biegelmeyer, R., Andrade, J.M.M., Aboy, A.L., Apel, M.A., Dresch, R.R., Marin, R., Raseira, M.D.C.B. and Henriques, A.T. (2011). Comparative analysis of the chemical composition and antioxidant activity of red (*Psidium cattleianum*) and yellow (*Psidium cattleianum var. lucidum*) strawberry guava fruit. *Journal of Food Science,* 76(7), C991–C996. http:// doi: 10.1111/j.1750-3841.2011.02319.x.

Borah, A., Pandey, S.K., Haldar, S. and Lal, M. (2021). Essential oil compositions, pharmacological importance and agro technological practices of Patchouli (*Pogostemon cablin Benth*.): A review. *Journal of Essential Oil Bearing Plants*, 24 (6), 1212-1226. http:// doi: 10.1080/0972060X.2021.1995511.

Deguchi, Y. and Miyazaki, K. (2010). Anti-hyperglycemic and anti-hyperlipidemic effects of guava leaf extract. *Nutrition and* Metabolism, 7, 9. https://doi.org/10.1186/1743-7075-7-9.

El-Ahmady, S.H., Ashour, M.L. and Wink, M. (2013). Chemical composition and anti-inflammatory activity of the essential oils of Psidium guajava fruits and leaves. *Journal of Essential Oil Research*, 25(6), 475-481. http:// doi:10.1080/10412905.2013.796498.

Fasola, T.R., Oloyede, G.K. and Aponjolosun, B.S. (2011). Chemical composition, toxicity and antioxidant activities of essential oils of stem bark of Nigerian species of Guava (*Psidium guajava* Linn.). *EXCLI Journal*, 10, 34–43.

Gavahian, M. and Farahnaky, A. (2018). Ohmic- assisted hydrodistillation: a review. *Trends in Food Science and Technology*, 72, 153-161. https:// doi.org/10.1016/j.tifs.2017.12.014.

Gutiérrez, R.M.P., Mitchell, S. and Solis, R.V. (2008). *Psidium guajava*: A Review of Its Traditional Uses, Phytochemistry and Pharmacology. *Journal of Ethnopharmacology*, 117(1), 1-27. https:// doi.org/10.1016/j.jep.2008.01.025

Irshad, Z., Hanif, M.A., Ayub, M.A., Jilani, M.I. and Tavallali, V. (2020). Chapter 26 - Guava in Medicinal Plants of South Asia. In Hanif, M.A., Nawaz, H., Khan, M.M. and Byrne, H.J. (Eds.) Medicinal Plants of South Asia: Novel Sources for Drug Discovery, p. 341–354. Netherlands, Elsevier. https://doi.org/10.1016/B978-0-08-102659-5.00026-4

Ishartani, D., Rahman, F., Hartanto, R., Utami, R. and Khasanah, L. (2018). Physical, chemical and sensory characteristics of red guava (*Psidium guajava*) velva at different fruit ripening time. *IOP Conference Series: Earth and Environmental Science*, 102, 012075. https://doi.org/10.1088/1755-1315/102/1/012075

Joseph, B. and Priya, M. (2011). Review on nutritional, medicinal and pharmacological properties of Guava (*Psidium guajava Linn*). *International Journal of Pharmacy and Biological Sciences*, 2(1), 53-69.

Kokolakis, A.K. and Golfinopoulos, S.K. (2013). Microwave-assisted techniques (MATs); a quick way to extract a fragrance: a review. *Natural Product Communication*, 8(10), 1493-1504. https:// doi.org/10.1177/1934578X1300801040.

Moon, P., Fu, Y., Bai, J., Plotto, A., Crane, J. and Chambers, A. (2018). Assessment of fruit aroma for twenty-seven guava (*Psidium guajava*) accessions through three fruit developmental stages. *Scientia Horticulturae*, 238 (12), 375–383. https:// doi.org/10.1016/j.scienta.2018.04.067.2018.

Nair, M.S., Saxena, A. and Kaur, C. (2018). Characterization and Antifungal Activity of Pomegranate Peel Extract and its Use in Polysaccharide-Based Edible Coatings to Extend the Shelf-Life of Capsicum (*Capsicum annuum L.*). *Food and Bioprocess Technology*, 11, 1317–1327. https://doi.org/10.1007/s11947-018-2101-x.

Naseer, S., Hussain, S., Naeem, N., Pervaiz, M. and Rahman, M. (2018). The phytochemistry and medicinal value of *Psidium guajava* (guava). *Clinical Phytoscience*, 4, 32. https://doi.org/10.1186/s40816-018-0093-8.

Paniandy, J.C., Chane-Ming, J. and Pieribattesti, J.C (2000). Chemical composition of the essential oil and headspace solid-phase microextraction of the guava fruit (*Psidium guajava* L.). *Journal of Essential Oil Research*, 12(2), 153-158. https:// doi.org/10.1080/10412905.2000.9699486.

Porat, R., Tietel, Z., Zippori, I. and Dag, A. (2011). Aroma volatile compositions of high- and low-aromatic guava varieties. *Journal of the Science Food and Agricultural*, 91(15), 2794-2798. https:// doi.org/10.1002/jsfa.4523.

Silva, L., Nelson, D., Drummond, M., Dufossé, L. and Glória, M. (2005). Comparison of hydrodistillation methods for the deodorization of turmeric. *Food Research International*, 38(8-9), 1087-1096. https:// doi.org/10.1016/j.foodres.2005.02.025.

Yadav, A., Kumar, N., Upadhyay, A., Fawole, O.A., Mahawar, MK., Jalgaonkar, K., Chandran, D., Rajalingam, S., Zengin, G., Kumar, M. and Mekhemar, M. (2022). Recent Advances in Novel Packaging Technologies for Shelf-Life Extension of Guava Fruits for Retaining Health Benefits for Longer Duration. *Plants (Basel),* 11(4), 547. https:// doi.org/10.3390/plants11040547.

https://doi.org/10.26656/fr.2017.7(3).317 © 2023 The Authors. Published by Rynnye Lyan Resources