



Original Article

Combined OPLS-DA and decision tree as a strategy to identify antimicrobial biomarkers of volatile oils analyzed by gas chromatography–mass spectrometry

 Felipe A. dos Santos^{a,b,1}, Ingrid P. Sousa^{a,1}, Niege A.J.C. Furtado^a, Fernando B. Da Costa^{a,b,*}
^a Laboratório de Farmacognosia, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, Brazil

^b Grupo de Pesquisa AsterBioChem, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, Brazil

ARTICLE INFO

Article history:

Received 15 April 2018

Accepted 11 August 2018

Available online 25 September 2018

Keywords:

Antimicrobial activity

Chemometrics

Decision tree

Volatile oils

Gas chromatography–mass spectrometry

Orthogonal projection to latent structures discriminant analysis

ABSTRACT

Bioguided isolation to discriminate antimicrobial compounds from volatile oils is a time- and money-consuming process. Considering the limitations of the classical methods, it would be a great improvement to use chemometric techniques to identify putative biomarkers from volatile oils. For this purpose, antimicrobial assays of volatile oils extracted from different plant species were carried out against *Streptococcus mutans*. Eight volatile oils that showed different antimicrobial effects (inactive, weakly active, moderately active and very active) were selected in this work. The volatile oils' composition was determined by GC–MS-based metabolomic analysis. Orthogonal projection to latent structures discriminant analysis and decision tree were carried out to access the metabolites that were highly correlated with a good antimicrobial activity. Initially, the GC–MS metabolomic data were pretreated by different methods such as centering, autoscaling, Pareto scaling, level scaling and power transformation. The level scaling was selected by orthogonal projection to latent structures discriminant analysis as the best pretreatment according to the validation results. Based on this data, decision tree was also carried out using the same pretreatment. Both techniques (orthogonal projection to latent structures discriminant analysis and decision tree) pointed palmitic acid as a discriminant biomarker for the antimicrobial activity of the volatile oils against *S. mutans*. Additionally, orthogonal projection to latent structures discriminant analysis and decision tree predicted as “very active” the antimicrobial activity of volatile oils, which did not belong to the training group. This predicted result is in agreement with our experimental result (MIC = 31.25 $\mu\text{g ml}^{-1}$). The present study can contribute to the development of useful strategies to help identifying antimicrobial constituents of complex oils.

© 2018 Sociedade Brasileira de Farmacognosia. Published by Elsevier Editora Ltda. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Volatile oils (VOs) have a prominent role in the search for new antimicrobial compounds. Since 2015, about 2000 scientific papers investigating the antimicrobial activity of VOs have been published (Scopus, 2018). These publications show different applications of VOs and their constituents in micro- and nanostructured systems, liposomes and films against pathogenic fungi, bacteria and resistant microorganisms. Therefore, VOs are still a significant natural source of new antimicrobials.

The most common techniques for antimicrobial assessment of VOs are the agar diffusion and dilution methods (Balouiri et al., 2016). The dilution assay is a quantitative method used to establish the minimal inhibitory concentration (MIC). MIC determination can be useful to compare the antimicrobial potential of different VOs, extracts or isolated compounds against the same microorganism strain.

Some metabolites that have been previously isolated from VOs are reported as good antimicrobials, such as menthol, carvacrol, citral, thymol, β -caryophyllene and α -cedrene (Iskan et al., 2002; Barrero et al., 2005; Dahham et al., 2015; Siroli et al., 2015). One available technique to identify VO's bioactive constituents is the bioautography assay (Balouiri et al., 2016). However, the high volatility of some constituents can interfere with the analysis,

* Corresponding author.

E-mail: febcosta@fcfrp.usp.br (F.B. Da Costa).¹ These authors contributed equally to this work.

and the conventional bioautography methods are not suitable for anaerobic and microaerophilic bacteria (Kovács et al., 2016).

Considering the limitations of the currently available techniques and the time-consuming process of a bioguided isolation approach to discriminate putative biomarkers from VOs, it would be a great improvement to use metabolomic and chemometric techniques to achieve this goal.

Metabolomics is the science of detection and identification of all metabolites (Fiehn, 2002; Krastanov, 2010). The complexity and large amount of data generated by metabolomics require chemometric techniques to elucidate and furnish interpretable information. Chemometrics can also contribute to correlate metabolomic data with biological activity of several samples simultaneously (Wiklund et al., 2008; Pan et al., 2010; Zhang et al., 2015). Metabolomics usually require hyphenated techniques, such as liquid or gas chromatography coupled to mass spectrometry (LC–MS or GC–MS). GC–MS is the most common method to determine VO's chemical constituents. Chemometrics applied to metabolomics is an innovative strategy for targeting active compounds from plant extracts and VOs (Chagas-Paula et al., 2015a).

There are many chemometric techniques, but the most commonly used are principal component analysis (PCA) for unsupervised analysis (Wold et al., 1987; Kettaneh et al., 2005) and partial least squares (PLS) for supervised (Wold et al., 2001; Kettaneh et al., 2005). The unsupervised methods show pattern recognition, trends, outliers and clustering that cover the entire sample's space.

The supervised methods are able to classify or predict a response like a biological activity or to determine the most discriminant metabolite(s), in the case of discriminant analysis (PLS-DA). Despite the recognized importance and utility of PLS-DA, some researchers have reported the use of orthogonal projection to latent structures discriminant analysis (OPLS-DA), which is a modification of the NIPALS PLS algorithm (Trygg and Wold, 2002). OPLS-DA is a powerful technique not only for interpretation and classification of sample sets (Eriksson et al., 2012), some of which from OMICS data (Boccard and Rutledge, 2013), but also to discriminate biomarkers (Wiklund et al., 2008; Pan et al., 2010; Chagas-Paula et al., 2015b). In an OPLS-DA model, the variation from matrix X that is not correlated to Y is removed. Therefore, some authors classify OPLS-DA better than PLS-DA for analysis interpretation, although both methods have the same predictive power (Trygg and Wold, 2002; Verron et al., 2004; Tapp and Kemsley, 2009).

Some authors have reported the applicability of unsupervised (Zhang et al., 2015) and supervised (Pan et al., 2010) chemometrics to discriminate samples and to discover biomarkers from GC–MS data. Other authors have demonstrated the use of S-plot and SUS-plot visualizations to identify potential biomarkers based on OPLS-DA and GC–MS data (Wiklund et al., 2008). Maree et al. (2014) used OPLS-DA to identify antimicrobial constituents from commercial VOs, and eugenol was found as a biomarker belonging to samples with good antimicrobial activity.

Beyond the methods that work with dimensionality reduction in a hyperplane such as OPLS-DA, the decision tree (DT) technique based on the J48 algorithm (Bhargava et al., 2013) can also be used for data analysis concerning classification, pattern recognition and prediction in data mining experiments (Endo et al., 2008; Zarkami, 2011). The J48 algorithm builds the model from the given data set (matrix X, with the variables) and generates a graphical representation known as DT, which shows the most important variable(s) for the classification of the model (matrix Y). In a DT, the “leaves” correspond to the classification and the “nodes” to the variables. So far, only a few publications show the applicability of DT on metabolomic studies to discover biomarkers (Chagas-Paula et al., 2015b). DT was used only once in a substructure prediction study based on GC–MS data (Canales et al., 2008), but to the best of our

knowledge, no study about the use of DT to identify biomarkers for antimicrobial activity of VO has been published so far.

Herein, we selected eight VOs from different plants with previously reported antimicrobial potential (Onawunmi et al., 1984; Canales et al., 2008; Singh et al., 2008; Bachir and Benali, 2012) and tested their effects against *Streptococcus mutans*. The chemical constitution of these VOs was obtained by GC–MS. The metabolomic data together with the biological activity of the VO (MIC values) were submitted to chemometric techniques. The OPLS-DA and DT were used to identify putative biomarkers that may be responsible for the antimicrobial activity of the VO against *S. mutans*. Additionally, based on the GC–MS data, chemometrics was used to predict the antimicrobial activity of a new VO that was not previously used in the model.

Material and methods

Plant material and VO extraction

Leaves and inflorescences of *Aldama arenaria* (Baker) E.E. Schill. & Panero, Asteraceae, were collected in a preserved Cerrado area along the Washington Luiz Highway, SP, Brazil (S 21°10.681'; W 047°51.541'; alt. 538 m), by F. B. Da Costa, F. A. Santos and I. P. Sousa. Plant collection was authorized by SISBIO (Brazilian Government's Authorization and Information in Biodiversity System, process #36391-1), and the access to genetic heritage was authorized by CNPq (National Council for Scientific and Technological Development, process #010055/2012-6).

The inflorescences were collected in February 2012, February and March 2013, and February 2014, around 9 a.m. The inflorescences were used fresh, except those collected in March 2013, which were dried outdoors for 5 days (average temperature around 24 °C). The leaves from *A. arenaria* were collected in February 2012. A voucher specimen (FBC # 103, SPFR 7652) of *A. arenaria* from the same population and period of the year is deposited at the SPFR Herbarium of the Department of Biology, FFCLRP, University of São Paulo, Ribeirão Preto, SP, Brazil.

Dried samples of the species *Cymbopogon citratus* (DC) Stapf, Poaceae, *Eucalyptus globulus* Labill., Myrtaceae, and *Zingiber officinale* Roscoe, Zingiberaceae, were purchased from the natural shop “Oficina de Ervas” in Ribeirão Preto, SP, Brazil, with the following batch numbers: 25SDM, 04SDM and 02SDM.

The volatile constituents of all plants were extracted by hydro-distillation for 4 h, using a modified Clevenger apparatus. The VOs were properly stored at –20 °C and the chemical and biological analyses were performed after the extractions.

GC–MS analysis

The VOs were analyzed on a Shimadzu QP-2010 gas chromatograph coupled to a quadrupole mass spectrometer (Shimadzu Corporation, Japan) and equipped with a DB-5MS capillary column (30 m × 0.25 mm × 0.25 μm; Agilent, USA). Helium was used as carrier gas at a flow rate of 1.3 ml/min. The oven temperature was programmed from 60 to 210 °C at 3 °C/min and the injector and detector temperatures were set at 250 and 260 °C, respectively. The injector split ratio was adjusted to 1:40. The ionizing energy was set to 70 eV.

The VO constituents were identified based on comparison of the obtained mass spectra with the mass spectral data from the libraries Wiley 7, NIST 08 and FFNSC 1.3. Additionally, the retention indices of the constituents were also compared with the literature values (Adams, 2007).

Antimicrobial activity

The antimicrobial activity of the VO was evaluated by the microdilution method. The microorganism *S. mutans* ATCC 25175 was previously activated in tryptic soy agar (BD, France) supplemented with 5% (v/v) sheep blood (EBE Farma, Brazil). The VOs were dissolved in dimethyl sulfoxide (Merck, Germany) and later diluted in tryptic soy broth (BD, France) to reach concentrations ranging from 0.97 to 2000 µg/ml. The solutions of VOs were added to a 96-well microplate with inoculum concentration of 5×10^5 colony forming units per ml. Chlorhexidine dihydrochloride (Sigma-Aldrich, USA) was used as positive control. The microplates were incubated under microaerophilic conditions at 37 °C. After 24 h incubation, microorganism viability was indicated with 0.02% (w/v) resazurin (Sigma-Aldrich, USA). The first well without visual color change of resazurin was defined as the MIC.

Chemometric analysis

The GC–MS metabolite profiling data from the VO analyzed were organized in an Excel spreadsheet (Microsoft Office Excel, 2013, Brazil). All detected peaks were aligned according to their respective retention times and mass spectrometric profiles. The peaks identified by the libraries of GC–MS were arranged in columns and the name of the samples arranged in rows. This table corresponded to the matrix X. The matrix Y was related to the antimicrobial activities of the VO, which were classified as inactive, weakly active, moderately active and very active according to their MIC values (Rios and Recio, 2005; Santos et al., 2008).

Different pretreatments such as centering, autoscaling, Pareto scaling, level scaling and power transformation were applied to the data (Van den Berg et al., 2006), in order to compare them in terms of multiple correlation coefficient (R_y^2) and predictive capacity (Q^2) by OPLS-DA. According to the obtained results, the best pretreatment method was selected with more reliability to identify putative biomarkers of VOs against *S. mutans*.

The PCA and OPLS-DA models from these five different data pretreatments were built in SIMCA (v. 13.0.3, Umetrics, Sweden). DT was carried out in Weka (v. 3.6.9, Waikato University, New Zealand), with the “minimum number of instances per leaf” equal to 1 (minimum amount of data separation per branching). The best pretreatment for the OPLS-DA model was also used in DT analysis.

Results and discussion

The antibacterial activities of the VO were compared by their respective MIC values. VOs with MIC values lower than 100 µg/ml were classified as very active, VOs with MIC values between 100 and 500 µg/ml were classified as moderately active, those with MIC values between 500 and 1000 µg/ml as weakly active and those with MIC values greater than 1000 µg/ml were considered inactive (Rios and Recio, 2005; Santos et al., 2008).

The VO from the fresh inflorescences of *A. arenaria* collected in February 2012 (*A. arenaria* 1, Table 1) was classified as “very active” based on its MIC value (15.6 µg/ml) for *S. mutans*. Due to the promising activity of this VO, the inflorescences of *A. arenaria* from the same population were collected again in 2013 (February—*A. arenaria* 2 and March—*A. arenaria* 3 and *A. arenaria* 4) and 2014 (February—*A. arenaria* 5). The leaves of *A. arenaria* were collected only once in 2012 due to its lower activity (*A. arenaria* 6, Table 1). All VOs were extracted from fresh samples in the same way, but the oil from the dried inflorescences of *A. arenaria* (*A. arenaria* 4, Table 1) was obtained for comparison of the chemical profiles.

The VO from different collections of *A. arenaria* displayed different chemical composition and, consequently, different

Table 1

Antimicrobial activity of VOs against *S. mutans* and their classification according to the MIC values.

Name of the sample for classification	MIC (µg ml ⁻¹)	Classification
<i>A. arenaria</i> 1	15.6	Very active
<i>A. arenaria</i> 2	250.0	Moderately active
<i>A. arenaria</i> 3	—	—
<i>A. arenaria</i> 4	62.5	Very active
<i>A. arenaria</i> 5	125.0	Moderately active
<i>A. arenaria</i> 6	1000	Weakly active
<i>C. citratus</i>	1000	Weakly active
<i>Z. officinale</i>	2000	Inactive
<i>Eucalyptus globulus</i>	2000	Inactive
Positive control	0.2	—

Note: *A. arenaria* 1: fresh inflorescences collected in February 2012; *A. arenaria* 2: fresh inflorescences collected in February 2013; *A. arenaria* 3: fresh inflorescences collected in March 2013 (activity predicted); *A. arenaria* 4: dried inflorescences collected in March 2013; *A. arenaria* 5: fresh inflorescences collected in February 2014; *A. arenaria* 6: fresh leaves collected in February 2012; *C. citratus*: dried leaves; *Z. officinale*: dried rhizome; *E. globulus*: dried leaves; Positive control: chlorhexidine dihydrochloride 0.12% (w/v).

Table 2

PCA and OPLS-DA constructed according to different pretreatments.

Model	PCA			OPLS-DA			
	PC	R_x^2	PredC	OrtC	R_y^2	Q^2	RMSEcv
Centering	2	0.75	3	3	0.969	0.52	0.39
Autoscaling	2	0.43	3	3	0.997	0.57	0.33
Level scaling	2	0.49	3	3	0.999	0.74	0.32
Pareto scaling	2	0.48	3	3	0.995	0.58	0.39
Power transformation	2	0.53	3	3	0.982	0.50	0.42

The results were expressed in terms of variance explained (R_x^2), calculated by two principal components (PC) for PCA and multiple correlation coefficient (R_y^2). The predictive capacity (Q^2) was calculated by three predictive components (PredC) and three orthogonal components (OrtC) for OPLS-DA.

antimicrobial activity. The samples *A. arenaria* 2 and *A. arenaria* 5 were classified as “moderately active” and the sample *A. arenaria* 6 as “weakly active”. The VO from the dried inflorescences (*A. arenaria* 4) was classified as “very active”, whereas *A. arenaria* 3 was used as external validation for prediction by OPLS and DT.

Other plant species were also used in this work. The VOs from *E. globulus* and *Z. officinale* were classified as inactive and the VO from *C. citratus* as weakly active against *S. mutans* (Table 1). Thereby, eight VOs displaying different MIC values (ranging from 15 to 2000 µg/ml) were selected for chemometric studies in order to have samples with different antimicrobial potential against the same microorganism (*S. mutans*). The analysis of the selected VO by GC–MS (Appendix B) showed a great number of different constituents, suitable for the chemometric analysis.

The VO from *A. arenaria* 1 displayed the best antimicrobial activity. Carotol (12.67%), faltarinol (6.71%) and spathulenol (5.48%) were identified as major constituents (Appendix B). The collections of *A. arenaria* inflorescences from the same population in different years provided chemically different VOs, from the qualitative and quantitative points of view. The major compounds and the proportion of monoterpenes and sesquiterpenes were different in these five VOs, resulting in different biological activities (from moderately to very active; Table 1).

Despite the chemical differences mentioned above, the PCA indicated a trend to a defined cluster for the five VOs from inflorescences and leaves of *A. arenaria* (Appendix B). The PCA carried out with the data coming from all pretreatments did not show any outlier. Moreover, all pretreatments displayed more than 40% of the variance explained (R_x^2) with only first two components (Table 2). The PCA clustering did not match with the biological activity presented by the samples (Appendix B). To obtain comparable models, OPLS-DA were standardized with three predictive (PredC) and

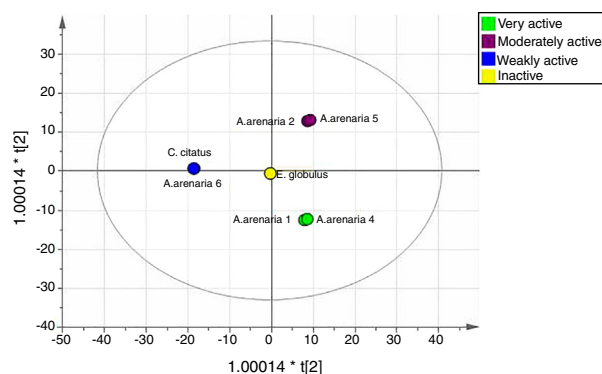


Figure 1. Score plot (PredC 2 × PredC 3) of OPLS-DA constructed with data pretreated by level scaling. The samples were classified according to the MIC values of the VOs against *S. mutans*. *A. arenaria 1*: fresh inflorescences collected in February 2012; *A. arenaria 2*: fresh inflorescences collected in February 2013; *A. arenaria 4*: dried inflorescences collected in March 2013; *A. arenaria 5*: fresh inflorescences collected in February 2014; *A. arenaria 6*: fresh leaves collected in February 2012; *C. citratus*: dried leaves; *Z. officinale*: dried rhizomes; *E. globulus*: dried leaves.

three orthogonal components (OrtC). This standardization allowed a good performance without overfitting (Table 2). The level scaling displayed the highest predictive capacity ($Q^2=0.74$) and the lowest root mean square error of cross-validation (RMSECV=0.32) (Table 2) when compared to other pretreatments.

Moreover, the results obtained for the PCA pretreatments (Table 2) are somewhat different from those reported by Van den Berg and co-workers, who also worked with GC–MS metabolomics data (Van den Berg et al., 2006). The authors stated that autoscaling and range scaling were better than other pretreatments for PCA. Nevertheless, in the present study, the centering was the best pretreatment based on R_x^2 result (Table 2). Although Van den Berg and co-workers carried out only PCA analysis, they described in detail each pretreatment, informing that level scaling was supposedly better for biomarkers identification. This information is in agreement with our OPLS-DA results.

Level scaling (Fig. 1) was selected for a more reliable discriminant analysis because it displayed better predictive capacity (Table 2) when compared to other pretreatment methods by OPLS-DA. The level scaling pretreatment separated the samples with higher antimicrobial activity from those with lower activity in

the score plot (Fig. 1). The predictive capacity is calculated by the leave-one-out seven-fold cross-validation method that ensures reliable analysis. The VO constituents correlated with good antimicrobial activity against *S. mutans* are shown in the loading plot (Fig. 2).

The regression coefficients that represent the prediction vector for the “very active” classification were selected by OPLS-DA. These coefficients are the variables that correspond to the chemical constituents detected by GC–MS and visualized in the loading plots. The variables with higher and positive magnitude (thin bars) and higher reliability (thick bars built with 95% of jack-knifed confidence intervals) were chosen (Fig. 3). Otherwise, a variable that displays a negative value has an opposite effect to any studied classification when the cross-validation is carried out. A list displaying the most important variables that contributed with the most active VO was created through the coefficient plot (Table 3).

The 15 putative biomarkers displayed in Table 3 are the variables related to the “very active” classification. These constituents are exclusive or are present at higher concentrations in the VO classified as very active. Palmitic acid was identified as the most important biomarker correlated to the good antimicrobial activity of the VO from the inflorescences of *A. arenaria 1* and 4 against *S. mutans*. This long-chain fatty acid has been reported in the literature showing antimicrobial properties against different microorganisms (Kabara et al., 1972; Ibrahim et al., 1991; Avrahami and Shai, 2004), including *S. mutans* (Huang et al., 2011). By thin-layer chromatography and bioautographic assay, Yff et al. (2002) obtained palmitic acid as the major antibacterial compound present in the ethyl acetate root extract of *Pentanisia prunelloides*. Palmitic acid also displayed inhibitory effects against other oral microorganisms such as *S. gordonii* and the Gram-negative bacteria *P. gingivalis* and *F. nucleatum* (Huang et al., 2011).

Additionally, the compounds 7,8-epoxy-1-octene, *cis*- α -bergamotene, methyl linolealdate, alloaromadendrene and veridiflorol were also correlated with the good antimicrobial activity, in addition to other nine constituents that could not be identified (n.i.) with the three GC–MS libraries used in this study. These biomarkers are at low concentrations in the respective active samples.

Palmitic acid (classified as the most important biomarker) corresponds to only 1.04 and 0.8% of the composition of *A. arenaria 1* and *A. arenaria 4* VO, respectively. In this sense, OPLS-DA can be a useful

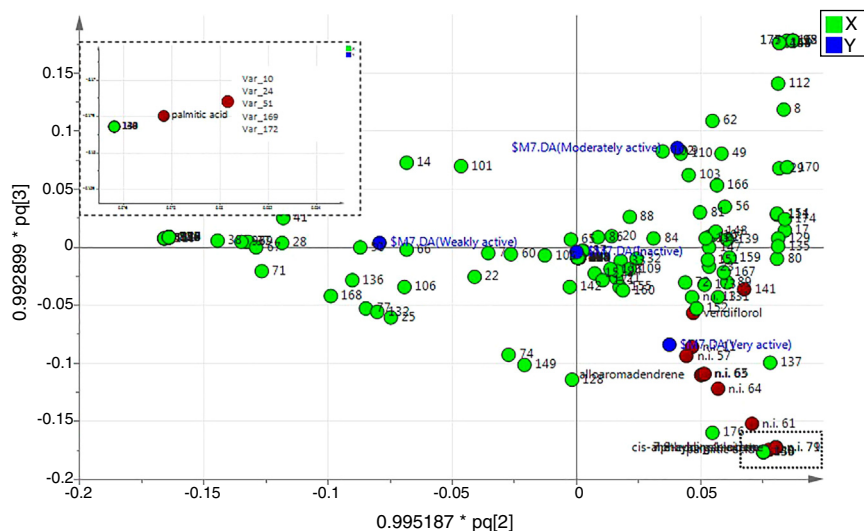


Figure 2. Loading plot (PredC 2 × PredC 3) of the OPLS-DA model constructed with data pretreated by level scaling. Green: Matrix X (VO components) classified according to OPLS-DA. Blue: Matrix Y (antimicrobial activity) classified according to OPLS-DA. Note: n.i.: metabolites not identified. The name of the numbered constituents can be seen in Appendix B.

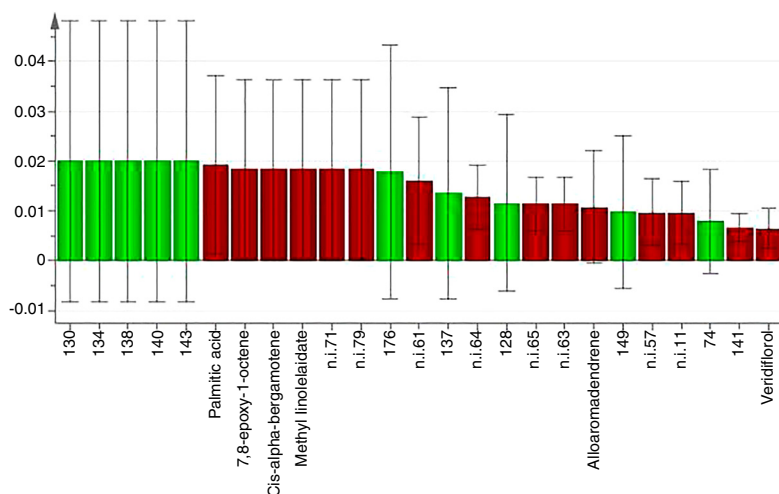


Figure 3. Putative biomarkers that contributed with the most active VO by means of a “coefficient plot” tool. The dark red bars are the statistically significant biomarkers due to their higher and positive magnitude (thin bars) and higher reliability (thick red bars).

Table 3

Summary of putative biomarkers correlated with the samples displaying good antimicrobial activity against *S. mutans* (“very active”).

Number of the variable	Name of the variable
55	Palmitic acid
10	7,8-Epoxy-1-octene
24	cis- α -Bergamotene
51	Methyl linolelaidate
169	n.i. 71
172	n.i. 79
161	n.i. 61
164	n.i. 64
165	n.i. 65
163	n.i. 63
12	Alloaromadendrene
156	n.i. 57
118	n.i. 11
141	n.i. 27
78	Veridiflorol

Note: n.i.: not identified.

technique to highlight possible active constituents that might not be associated to the biological activity without statistical analysis, once they are present at low concentrations. Therefore, the analysis by OPLS-DA can guide the isolation of constituents in complex samples such as VOs and also suggests combinations for synergistic effects.

Moreover, OPLS-DA was also used to predict the unknown antimicrobial activity of a VO (*A. arenaria* 3, Table 1) based on its chemical data pretreated by level scaling. The model predicted the

activity of the VO mainly as “very active” against *S. mutans* (Table 4). This analysis indicated that this oil sample has 41.1% of chance to be classified as “very active”, although it has 30.0% of chance to be classified as “moderately active”, 18.9% as “weakly active” and 9.9% as “inactive” (Table 4). The antimicrobial assay was carried out with this sample, and the VO displayed an MIC value equal to 31.2 $\mu\text{g/ml}$ (“very active” classification). This experimental result is in agreement with the statistical prediction, therefore experimentally validating our predictive model. Eight from the 15 biomarkers listed by OPLS-DA (Table 3) were detected in this VO: veridiflorol (0.72%), n.i. 11 (3.68%), n.i. 27 (2.13%), n.i. 57 (0.37%), n.i. 61 (0.49%), n.i. 63 (1.37%), n.i. 64 (0.56%), n.i. 65 (0.34%) (Appendix B). However, palmitic acid was not found in this active VO.

DT model (Fig. 4) was also carried out using the data pretreated with level scaling. This model was used for classification purposes only. The resulting DT classifier (Fig. 4, on the right) indicated palmitic acid (variable 55) as the most important constituent correlated to the activity, as already indicated by the OPLS-DA model (Table 3). This result reinforces the predictive validity of the statistical models used in this study.

The DT classifier also showed the variables 16 and 9 (Fig. 4, on the right) that correspond to bornyl acetate and 4-terpineol, respectively. The results for the DT classifier indicate that the VO is very active against *S. mutans* when palmitic acid is at a higher concentration than -1 (corresponding to zero, before level scaling pretreatment). Conversely, the VO is inactive when 4-terpineol is present and moderately active when bornyl acetate is part of the composition, without palmitic acid. This classification can lead to some oriented possibilities such as synergism and antagonism

Table 4

Antimicrobial activity prediction of the VO from fresh inflorescences of *A. arenaria* 3 against *S. mutans*.

Name of oil for classification*	Classification	Very active (%)	Moderately active (%)	Weakly active (%)	Inactive (%)
<i>A. arenaria</i> 1	Very active	97.54	-1.89	1.52	2.82
<i>A. arenaria</i> 2	Moderately active	-2.74	98.16	1.57	3.02
<i>A. arenaria</i> 3	—	41.10	30.03	18.94	9.94
<i>A. arenaria</i> 4	Very active	98.04	0.53	0.15	1.29
<i>A. arenaria</i> 5	Moderately active	-1.45	100.51	0.05	0.90
<i>A. arenaria</i> 6	Weakly active	-1.75	-0.22	100.49	1.47
<i>C. citratus</i>	Weakly active	-2.19	-0.95	100.97	2.17
<i>Z. officinale</i>	Inactive	-1.81	-0.47	0.64	101.65
<i>E. globulus</i>	Inactive	-1.93	-0.63	0.75	100.81

Note: The VO was classified as “very active” by OPLS-DA (41.10% of chance). *A. arenaria* 1: fresh inflorescences collected in February 2012; *A. arenaria* 2: fresh inflorescences collected in February 2013; *A. arenaria* 3: fresh inflorescences collected in March 2013; *A. arenaria* 4: dried inflorescences collected in March 2013; *A. arenaria* 5: fresh inflorescences collected in February 2014; *A. arenaria* 6: fresh leaves collected in February 2012; *C. citratus*: dried leaves; *Z. officinale*: dried rhizomes; *E. globulus*: dried leaves.

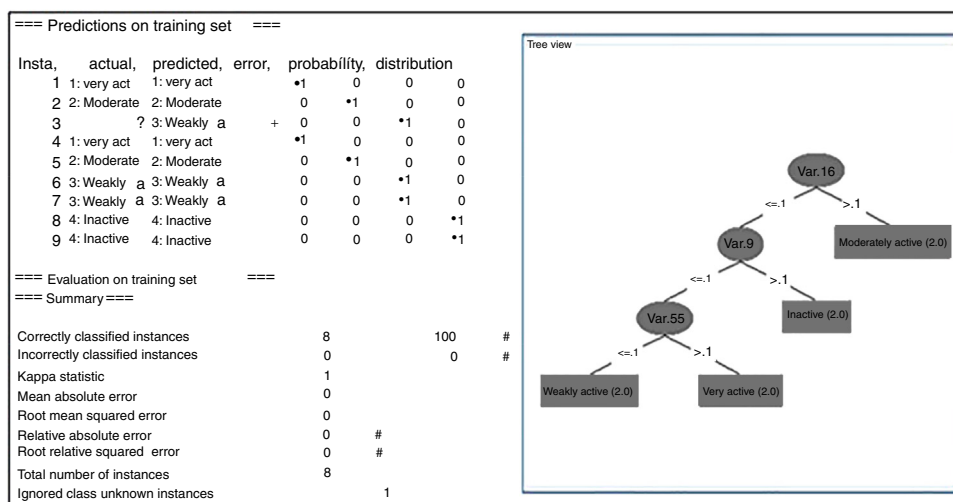


Figure 4. On the right: the DT classifier showing palmitic acid (Var. 55) as the most important variable correlated to the good antimicrobial activity against *S. mutans*. Bornyl acetate (Var. 16) was related to a moderate activity and 4-terpineol (Var. 9) was indicated as an antagonist of the antimicrobial activity. On the left: the prediction of the VO from fresh inflorescence of *A. arenaria* 3 (line 3 of the *inst#*) as “very active”, based on its chemical composition and training set.

studies involving the isolated constituents from the VO. Thus, this DT classifier showed not only a classification correlated to metabolite concentration, but also the possible interactions between different constituents.

The VO from *A. arenaria* 3 was also classified as very active against *S. mutans* by the DT (Fig. 4, on the left, *inst#* 3), thus corroborating the OPLS-DA prediction. So far, no work regarding the use of DT to identify biomarkers for antimicrobial activity of VO has been published, except for LC–MS data (Chagas-Paula et al., 2015b). Therefore, DT is an underexplored technique that is able to point out discriminant compounds in metabolomics data. Moreover, DT can be used with other chemometric techniques to improve the analytical research to identify active compounds as well as to classify unknown samples.

Conclusions

The proposed models of OPLS-DA and DT were successfully applied to identify putative biomarkers of VOs displaying good antimicrobial activity against *S. mutans*. Palmitic acid was identified as the most important biomarker by both models. Other metabolites such as 7,8-epoxy-1-octene, *cis*- α -bergamotene, methyl linolelaidate, alioaromadendrene and veridiflorol were also correlated to good antimicrobial activity by OPLS-DA, together with other nine unidentified constituents. On the other hand, bornyl acetate was associated to samples displaying moderate antimicrobial activity and 4-terpineol was associated to inactive VOs by DT. Further studies of synergism and antagonism with these metabolites would be relevant to evaluate their effects when combined. The antimicrobial activity prediction of OPLS-DA and the classification of DT were successfully applied for a VO and this result was later confirmed by our experimental antimicrobial assay. These results can contribute to the development of useful strategies to help identifying antimicrobial constituents of complex VOs.

Author's contributions

F.A.S., I.P.S. and F.B.C. contributed in collecting plant samples and their identifications. I.P.S. extracted the VOs and performed the metabolite identification and antimicrobial assays. F.A.S. performed the pretreatments of the GC–MS data and the chemometric analysis. F.A.S. and I.P.S. drafted the manuscript. F.A.S. and I.P.S. contributed equally in this work. N.A.J.C.F. and F.B.C. designed the

experiments, supervised the laboratory works and contributed to critical reading of the manuscript. All the authors have read the final manuscript and approved the submission.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

This work was supported by CNPq, CAPES and FAPESP (grants #2012/01429-8 and 2012/10249-3). Special thanks to Prof. Norberto P. Lopes and Mrs. Izabel C.C. Turatti (FCFRP-USP) for running the GC–MS experiments and Mr. Mário Ogasawara and Mrs. Maria Angélica S.C. Chellegatti (FCFRP-USP) for the laboratory assistance.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.bjp.2018.08.006](https://doi.org/10.1016/j.bjp.2018.08.006).

References

- Adams, R.P., 2007. *Identification of Essential Oil Components by Gas Chromatography/Mass Spectrometry*, 4th ed. Allured Publishing Corporation, Illinois.
- Avrahami, D., Shai, Y., 2004. A new group of antifungal and antibacterial lipopeptides derived from non-membrane active peptides conjugated to palmitic acid. *J. Biol. Chem.* 279, 12277–12285.
- Bachir, R.G., Benali, M., 2012. Antibacterial activity of the essential oils from the leaves of *Eucalyptus globules* against *Escherichia coli* and *Staphylococcus aureus*. *Asian Pac. J. Trop. Biomed.* 2, 739–742.
- Balouiri, M., Sadiki, M., Ibsouda, S.K., 2016. Methods for *in vitro* evaluating antimicrobial activity: a review. *J. Pharm. Anal.* 6, 71–79.
- Barrero, A.F., Quilez Del Moral, J.F., Lara, A., Herrador, M.M., 2005. Antimicrobial activity of sesquiterpenes from essential oil of *Juniperus thurifera*. *Planta Med.* 71, 67–71.
- Bhargava, N., Sharma, G., Bhargava, R., Mathuria, M., 2013. Decision Tree analysis on J48 algorithm for data mining. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* 6, 1114–1119.
- Boccard, J., Rutledge, D.N.A., 2013. A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock OMICS data fusion. *Anal. Chim. Acta* 769, 30–39.
- Canales, M., Hernández, T., Rodríguez-Moroy, M.A., Jiménez-Estrada, M., Flores, C.M., Hernández, L.B., Gijón, I.C., Quiroz, S., García, A.M., Avila, G., 2008. Antimicrobial activity of the extracts and essential oil of *Viguiera dentata*. *Pharm. Biol.* 46, 719–723.
- Chagas-Paula, D.A., Zhang, T., Da Costa, F.B., Edrada-Ebel, R., 2015a. A metabolomic approach to target compounds from the Asteraceae family for dual COX and LOX inhibition. *Metabolites* 5, 404–430.

- Chagas-Paula, D.A., Oliveira, T.B., Zhang, T., Edrada-Ebel, R., Da Costa, F.B., 2015b. Prediction of anti-inflammatory plants and discovery of their biomarkers by machine learning algorithms and metabolomic studies. *Planta Med.* 81, 450–458.
- Dahham, S.S., Tabana, Y.M., Iqbal, M.A., Ahamed, M.B.K., Ezzat, M.O., Majid, A.S.A., Majid, A.M.S.A., 2015. The anticancer, antioxidant and antimicrobial properties of the sesquiterpene β -caryophyllene from the essential oil of *Aquilaria crassna*. *Molecules* 20, 11808–11829.
- Endo, A., Shibata, T., Tanaka, H., 2008. Comparison of seven algorithms to predict breast cancer survival. *Int. J. Biomed. Soft Comput. Hum. Sci.* 13, 11–16.
- Eriksson, L., Rosén, J., Johansson, E., Trygg, J., 2012. Orthogonal PLS (OPLS) modeling for improved analysis and interpretation in drug design. *Mol. Inform.* 31, 414–419.
- Fiehn, O., 2002. Metabolomics—the link between genotypes and phenotypes. *Plant Mol. Biol.* 48, 155–171.
- Huang, C.B., Alimova, Y., Myers, T.M., Ebersole, J.L., 2011. Short- and medium-chain fatty acids exhibit antimicrobial activity for oral microorganisms. *Arch. Oral Biol.* 56, 650–654.
- Ibrahim, H.R., Kato, A., Kobayashi, K., 1991. Antimicrobial effects of lysozyme against gram-negative bacteria due to covalent binding of palmitic acid. *J. Agric. Food Chem.* 39, 2077–2082.
- Iskan, G., Kirimer, N., Kurkcuoglu, M., Baser, H.C., Demirci, F., 2002. Antimicrobial screening of *Mentha piperita* essential oils. *J. Agric. Food Chem.* 50, 3943–3946.
- Kabara, J.J., Swieczkowski, D.M., Conley, A.J., Truant, J., 1972. Fatty acids and derivatives as antimicrobial agents. *Antimicrob. Agents Chemother.* 2, 23–28.
- Kettaneh, N., Berglund, A., Wold, S., 2005. PCA and PLS with very large data sets. *Comput. Stat. Data Ann.* 48, 69–85.
- Kovács, J.K., Horváth, G., Kerényi, M., Kocsis, B., Emody, L., Schneider, G., 2016. A modified bioautographic method for antibacterial component screening against anaerobic and microaerophilic bacteria. *J. Microbiol. Methods* 123, 13–17.
- Krastanov, A., 2010. Metabolomics—the state of art. *Biotechnol. Biotechnol. Equip.* 1, 1537–1543.
- Maree, J., Kamatou, G., Gibbons, S., Viljoen, A., Vuuren, S.V., 2014. The application of GC–MS combined with chemometrics for the identification of antimicrobial compounds from selected commercial essential oils. *Chemometr. Intell. Lab. Syst.* 13, 172–181.
- Onawunmi, G.O., Yisak, W.A., Ogunlana, E.O., 1984. Antibacterial constituents in the essential oil of *Cymbopogon citrates* (DC.) Stapf. *J. Ethnopharmacol.* 12, 279–286.
- Pan, L., Qiu, Y., Chen, T., Lin, J., Chi, Y., Su, M., Zhao, A., Jia, W., 2010. An optimized procedure for metabolomic analysis of rat liver tissue using gas chromatography/time-of-flight mass spectrometry. *J. Pharm. Biomed. Anal.* 52, 589–596.
- Rios, J.L., Recio, M.C., 2005. Medicinal plants and antimicrobial activity. *J. Ethnopharmacol.* 100, 80–84.
- Santos, A.O.S., Ueda-Nakamura, T., Filho, B.P.D., Veiga Junior, V.F., Pinto, A.C., Nakamura, C.V., 2008. Antimicrobial activity of Brazilian copaiba oils obtained from different species of the *Copaifera* genus. *Mem. Inst. Oswaldo Cruz* 103, 277–281.
- Scopus, Elsevier, 2018. <https://www.scopus.com/> (accessed 26 February 2018).
- Singh, G., Kapoor, I.P.S., Singh, P., Heluani, C.S., Lampasona, M.P., Catalan, C.A.N., 2008. Chemistry, antioxidant and antimicrobial investigations on essential oil and oleoresins of *Zingiber officinale*. *Food Chem. Toxicol.* 46, 3295–3302.
- Siroli, L., Patrignani, F., Gardini, F., Lanciotti, R., 2015. Effects of sub-lethal concentrations of thyme and oregano essential oils, carvacrol, thymol, citral and *trans*-2-hexenal on membrane fatty acid composition and volatile molecule profile of *Listeria monocytogenes*, *Escherichia coli* and *Salmonella enteritidis*. *Food Chem.* 182, 185–192.
- Tapp, H.S., Kemsley, E.K., 2009. Note on the practical utility of OPLS. *Trends Analyt. Chem.* 11, 1322–1327.
- Trygg, J., Wold, S., 2002. Orthogonal projections to latent structures (O-PLS). *J. Chemometrics* 16, 119–128.
- Van den Berg, R.A., Hoefsloot, H.C.J., Westerhuis, J.A., Smilde, A.K., van der Werf, M.J., 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* 142, 1–15.
- Verron, T., Sabatier, R., Joffre, R., 2004. Some theoretical properties of O-PLS method. *J. Chemometrics* 18, 62–68.
- Wiklund, S., Johansson, E., Sjöström, L., Mellerowicz, E.J., Edlund, U., Shockcor, J.P., Gottfries, J., Moritz, T., Trygg, J., 2008. Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. *Anal. Chem.* 80, 115–122.
- Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. *Chemometr. Intell. Lab. Syst.* 2, 37–52.
- Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. *Chemometr. Intell. Lab. Syst.* 58, 109–130.
- Yff, B.T.S., Lindsey, K.L., Taylor, M.B., Erasmus, D.G., Jäger, A.L., 2002. The pharmacological screening of *Pentania prunelloides* and the isolation of the antibacterial compound palmitic acid. *J. Ethnopharmacol.* 79, 101–107.
- Zarkami, R., 2011. Application of classification tree-J48 to model the presence of roach (*Rutilus rutilus*) in rivers. *Caspian J. Environ. Sci.* 2, 189–198.
- Zhang, W., Zhu, S., He, S., Wang, Y., 2015. Screening of oil sources by using comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry and multivariate statistical analysis. *J. Chromatogr. A* 1380, 162–170.