G Model
CULHER-3353; No. of Pages 12
ARTICLE IN PRESS
Journal of Cultural Heritage xxx (2017) xxx–xxx
Available online at ScienceDirect www.sciencedirect.com
Elsevier Masson France EM|consulte www.em-consulte.com/en

Original article

An analytical strategy based on Fourier transform infrared spectroscopy, principal component analysis and linear discriminant analysis to suggest the botanical origin of resins from *Bursera*. Application to archaeological Aztec Samples

Carmen Piña-Torres [a], Paola Lucero-Gómez [a,*], Stephany Nieto [a], Alfonsina Vázquez [a], Lauro Bucio [b], Irma Belio [c], Rito Vega [c], Carole Mathe [d], Catherine Vieillescazes [d]

[a] *Laboratory of Chemistry Applied to Art and Archaeology & LANCIC, Chemistry Institute, National Autonomous University of Mexico (UNAM), Mexico City, Mexico*
[b] *Physics Institute, National Autonomous University of Mexico (UNAM), Mexico City, Mexico*
[c] *Cuerpo de Materiales, Sinaola University, Sinaloa, Mexico*
[d] *IMBE UMR, Avignon University/CNRS/IRD/AMU, Restoration Engineering of Natural and Cultural Heritage, Faculty of Sciences, 84911 Avignon, France*

ARTICLE INFO

Article history:
Received 18 May 2017
Accepted 12 February 2018
Available online xxx

Keywords:
FTIR
Copal
Aztec
Botanical origin

ABSTRACT

Bursera species are the source of oleoresins that have been used by pre-Columbian American cultures as adhesives, raw materials for molding figurines, ritual offerings, among other uses. Spread along different museum collections all over the world, pre-Columbian artefacts contain these resins. The preservation and understanding of the technology of fabrication of these pieces constitute a major concern for conservators, historians and archaeologists. Few studies have so far dealt with the chemical composition and the botanical origin of Mexican copal, owing maybe to the difficulty on the procuration of resins from known botanical origin. In this work, fresh resins from six Mexican *Bursera* species, namely *B. bipinnata*, *B. excelsa*, *B. grandifolia*, *B. laxiflora*, *B. penicillata* and *B. stenophylla*, were analyzed by Fourier-transformed infrared spectroscopy (FTIR). Main spectral band positions were selected for chemometric analysis using principal component analysis (PCA), based on the loading plot of chemometric analysis. Sample distribution patterns were investigated with PCA. Score plots revealed a sample agglomeration with good differentiation in 5 out of the 6 species. This method was validated by linear discriminant analysis (LDA) with a 95.2% of global positive recognition for certified origin species. To compare the efficiency of this approach, high performance liquid chromatography coupled to diode array detection (HPLC-DAD) and FTIR results were coupled to PCA and LDA, for the same set of samples. "FTIR showed 94.4% of samples correctly assigned on the confusion matrix and 91% on the cross validation one. HPLC-LDA showed 100% of correct assignment in the confusion matrix and 95% on the cross validation one. These results are encouraging, as FTIR is much faster and less expensive than chromatographic techniques and it could more readily be available in conservation laboratories. Finally, an application to the identification of the botanical origin of four archaeological Aztec copal samples was performed and the model suggested an origin on *B. bipinnata*/*B. stenophylla* for these archaeological samples.

1. Introduction

Fourier-transform infrared spectroscopy is a sensitive and microdestructive diagnostic tool. It has been widely applied to the analysis of different organic and inorganic materials in the field of Cultural Heritage.

FTIR applications range from the monitoring of the ageing of ketonic [1] and mastic resins [2] to the study of the nature of ceramic organic residues [3,4]. Different authors have used this method for investigating adhesives from a Cypriot pottery [5] to the analysis of binders [6,7], the investigation of illuminated manuscripts [8,9] to historical varnishes [10] or from standards to natural resins used as paint varnishes [11].

Natural products are complex mixtures of organic molecules that can be identified by various analytical methods. The modern word «copal» is derived from the Nahuatl word «copalli» meaning resin. From the Spain of 16th century Spanish "copal" became a term used in the markets all over the world, and soon referred

* Corresponding author. Laboratory 2-5 of Chemistry Applied to Art and Archaeology, Circuito Exterior, Ciudad Universitaria, C.P. 04510 Delegación Coyoacán, Cd. de México, Mexico. Tel.: +52 55 53530415.
E-mail addresses: plucero@unam.mx, paolaluceg@gmail.com (P. Lucero-Gómez).

https://doi.org/10.1016/j.culher.2018.02.006
1296-2074/

Please cite this article in press as: C. Piña-Torres, et al., An analytical strategy based on Fourier transform infrared spectroscopy, principal component analysis and linear discriminant analysis to suggest the botanical origin of resins from *Bursera*. Application to archaeological Aztec Samples, Journal of Cultural Heritage (2017), https://doi.org/10.1016/j.culher.2018.02.006

to almost any resin and even gums [12], thus generating great confusion among its customers. In this work, the term copal only describes the specific resin produced by *Bursera* trees in Mexico context.

Natural resins are exuded of trees and insects and because of their wide availability they have been used all over the world since ancient times [13]. In pre-Columbian Mexican times their use as adhesive materials [14] in archaeological Aztec findings [15,16] such as featherworks [17], dental fillings and even dental incrustations with semi-precious stones [18,19], was recorded.

Additionally, these resins have often been used in the artistic field as binders for paintings by such renowned artists as Diego Rivera, David A. Siqueiros and Dr. Atl [20]. At the archaeological site of Templo Mayor in Mexico City, copal has been found in different offerings shaped as figurines that could represent Aztec gods or offerings for spiritual rituals in cone-shaped conglomerates [15].

To our knowledge few studies have been performed on Mexican resins [16,21] and even less researches were performed on resins from living specimens [15]. This may be due to the great number of *Bursera* trees producing resins. Along the Pacific coast of Mexico, the botanic genus *Bursera* counts more than 80 species. *Bursera* genus is a relative of the Old World *Boswellia* and *Commiphora* genus [22]. The resins of *Bursera* are composed of mono- and sesquiterpenes [23] mixed with triterpenes [24,25].

The genus exhibits great variations in the number of these terpenes and their relative concentrations [13]. Among the triterpenes that have been found in *Bursera* resins are: α-amyrin (98.5%), 3-epi-lupeol (98.0%) and β-amyrin (82.0%) from a commercial Mexican copal [24]. Stacey and Catwright characterized a sample attributed to *Bursera excelsa* by its abundant lupeol, β-amyrin and α-amyrenone and also a pair of amyrin isomers, possibly 3-*epi*-α-amyrin and 3-*epi*-β-amyrin [16]. *B. simaruba* includes a variety of triterpenes such as lupeol, *epi*-lupeol, *epi*-glutinol, α- and β-amyrin, and a lupane type diol [13].

Resins being organic materials can be identified by various analytical methods, the most widely employed technique is gas chromatography coupled with mass spectrometry (GC-MS) [26–30]. Pyrolysis (Py) can be coupled to GC-MS for the study of polymerized components [31–33] and IR spectroscopy [34–41].

Chromatographic techniques present many disadvantages compared to FTIR: their detection limit is usually higher than that of FTIR, they cannot document the organization of heterogeneous samples [42] and implicate pretreatments like purification, extraction and derivatization that may be time consuming and modify original composition. Such necessary procedures could also cause contamination.

FTIR is a widely accepted analytical research tool in the field of Cultural Heritage that has been used for over 50 years [43]. However, IR alone does not allow the discrimination of materials with close chemical compositions.

Resins are constituted by secondary metabolites. Their production on the plant is under strong gene regulation and usually they occur only in particular groups of taxonomically related plants [3].

As it is known, FTIR band positions allow characterizing the chemical functions of the various molecules that occur in a mixture. When the FTIR spectra of a natural resin are obtained, the band position for each functional group is slightly shifted from a species to another, owing to the interactions between functional groups of the molecules that are present in the mixture. This difference among resin composition from a species to another is the crucial factor that allows the application of FTIR and multivariate analysis on the discrimination of resin samples from different botanical origin. Multivariate statistical methods aid the extraction and interpretation of information from the spectra.

PCA is a wide spread mathematical technique used for pattern recognition [44,45]. It allows the grouping of samples into categories by maximizing the variance between categories and minimizing the variance within groups by projecting data in a reduced hyperspace, defined by the principal components [46]. These components are linear combinations of the original variables, with the first principal component having the largest variance, the second principal component having the second largest variance and so on.

The aim of many methods of multivariate analysis is data reduction. Quite frequently, there is some correlation between the variables, so some of the information is redundant. PCA allows reducing the amount of data when there is correlation present [47].

LDA is a supervised pattern recognition method. This type of analysis starts with a number of objects whose group membership is known, this group of objects is known as learning or training objects. The aim of supervised pattern recognition methods is to use these objects to find a rule for allocating a new object of unknown group to the correct group.

In order to do so, linear discriminant analysis (LDA) finds a linear discriminant function (LDF), Y, which is a linear combination of the original measured variables:

$$Y = a_1X_1 + a_2X_2 + \ldots + a_nX_n$$

The original n measurements for each object are combined into a single value of Y, so the data is reduced from n dimensions to one dimension. The coefficients of the terms are chosen in such a way that Y reflects the difference between groups as much as possible: objects in the same group will have similar values of Y and objects in different groups will have very different values of Y. Then an unknown object will be classified according to its Y-value.

The success of LDA at allocating an object correctly can be tested in several ways. The simplest is to use the classification rule to classify each object in the group and to record whether the resulting classification is correct. The table summarizing the results of this procedure is called confusion matrix in the present work. This method tends to be overoptimistic since the object being classified was part of the set, which was used to form the rule (Table 1).

Another method, which uses the data more economically, is cross-validation, sometimes called the 'leave-one-out method'. As its name suggests, this finds the LDF with one object omitted and checks whether this LDF then allocates the omitted object correctly. The procedure is then repeated for each object in turn and again a success rate can be found. [44].

In every case the selection of bands to be considered for PCA or LDA analysis is crucial [48]. If selected bands are not resolved, this may introduce noise to the mathematical model that would prevent accurate grouping of the samples into the hyperspace [46,48,49]. To prevent such a situation, the bands must be selected for the discrimination of the loading plot produced by the PCA. The loading plot is a graphic that allows to distinguish which variables contribute to the separation and to what extent [46,49–51], the higher the absolute value of the vector representing a variable, the higher its impact on separation.

For multivariate analysis, a matrix constituted by the resin samples (rows) and the position of the bands of their FTIR spectra (columns) was analyzed by principal component analysis (PCA) to display the structure of the data.

Table 1
Samples of contemporary resins of certified botanical origin.

Identification	Number of samples	Physical state
B. bipinnata	8	Solid
B. excelsa	4	Solid
B. grandifolia	6	Solid
B. laxiflora	10	Semi-liquid
B. penicillata	6	Solid
B. stenophylla	8	Solid