Analyst

RSC Publishing

PAPER

View Article Online
View Journal | View Issue

Identification of *Dactylopius* cochineal species with high-performance liquid chromatography and multivariate data analysis†

Cite this: *Analyst*, 2013, **138**, 6081

Ana Serrano,[a] Micaela Sousa,[b] Jessica Hallett,[a] Monique S. J. Simmonds,[c] Mark Nesbitt[c] and João A. Lopes*[d]

Identification of American cochineal species (*Dactylopius* genus) can provide important information for the study of historical works of art, entomology, cosmetics, pharmaceuticals and foods. In this study, validated species of *Dactylopius*, including the domesticated cochineal *D. coccus*, were analysed by high-performance liquid chromatography with a diode array detector (HPLC-DAD) and submitted to multivariate data analysis, in order to discriminate the species and hence construct a reference library for a wide range of applications. Principal components analysis (PCA) and partial least squares discriminant analysis (PLSDA) models successfully provided accurate species classifications. This library was then applied to the identification of 72 historical insect specimens of unidentified species, mostly dating from the 19th century, and belonging to the Economic Botany Collection, Royal Botanic Gardens, Kew, England. With this approach it was possible to identify anomalies in how insects were labelled historically, as several of them were revealed not to be cochineal. Nevertheless, more than 85% of the collection was determined to be species of *Dactylopius* and the majority of the specimens were identified as *D. coccus*. These results have shown that HPLC-DAD, in combination with suitable chemometric methods, is a powerful approach for discriminating related cochineal species.

Received 7th January 2013
Accepted 30th July 2013

DOI: 10.1039/c3an00052d

www.rsc.org/analyst

Introduction

Study of *Dactylopius* cochineal species

For nearly two decades, the development of green chemistry and sustainable development has led to increased concerns about using raw materials, eliminating waste and avoiding the use of toxic and/or hazardous reagents and solvents, in the manufacture and application of chemical products.[1] Consequently, a renewed interest in natural dyes has recently emerged, among several research areas, for diverse economic purposes.[2]

Cochineal, one of the most valuable natural red dyes used in historical works of art, is currently an important source of red for dyes, lake pigments, food additives, cosmetics and pharmaceutical colorants, due to its low toxicity and its intense hues and colourfastness.[24] This red dye, mainly composed of carminic acid (a red hydroxyanthraquinone compound), can be obtained from the American cochineal insects belonging to *Dactylopius* genus (Hemiptera: Dactylopiidae), which comprises ten species, namely *D. austrinus* De Lotto, *D. confertus* De Lotto, *D. salmianus* De Lotto, *D. zimmermanni* De Lotto, *D. opuntiae* Cockerell, *D. bassi* Dov & Marotta, *D. confusus* Cockerell, *D. tomentosus* Lamarck, *D. ceylonicus* Green and the domesticated *D. coccus* Costa.[3,6] The latter is considered the most popular, as it holds higher contents of carminic acid[4,5] and, therefore, it has been widely exploited as a food additive.[5,7,8] It was always considered a desirable dye for colouring textiles and paintings, until the 19th century.[9]

The study of cochineal dyes is usually made with high-performance liquid chromatography with a diode array detector (HPLC-DAD), although visual interpretation or quantification of chromatographic peaks does not allow the accurate discrimination between species of *Dactylopius*.[9,10] Recent studies have emphasized the importance of applying multivariate statistical methods to the chromatographic data. For instance, Serrano *et al.* have shown that principal components analysis (PCA) can provide precise information of the date and the provenance of works of art, and additionally, on the trade routes in which American cochineal dyestuffs circulated.[9] On the other hand, multivariate analysis has been used for discriminating the geographical origin of cochineal insects: hierarchical cluster analysis (HCA) and PCA were used by Méndez *et al.*[10] to

[a]*CHAM (Centre for Overseas History), Faculdade de Ciências Sociais e Humanas, Universidade Nova de Lisboa & Universidade dos Açores, Avenida de Berna, 1069-061 Lisboa, Portugal*

[b]*Madeira Interactive Technologies Institute, Polo Científico e Tecnológico da Universidade da Madeira, Caminho da Penteada, 9020-105 Funchal, Portugal*

[c]*Royal Botanic Gardens, Kew, Richmond TW9 3AB, Surrey, England, UK*

[d]*REQUIMTE, Departamento de Ciências Químicas, Faculdade de Farmácia, Universidade do Porto, 4099-030 Porto, Portugal. E-mail: joaolopes@ff.up.pt; Fax: +351 226093483; Tel: +351 220428664*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c3an00052d

discriminate the origin of *D. coccus*, based on the percentage of pigments (estimated by HPLC-DAD) and colour quality variables. Moreover, Chavéz-Moreno *et al.*[3] used the same methods (HCA and PCA) on the chromatographic profiles of five species of *Dactylopius* to characterize their host plant and geographical origin (Argentina or Mexico).

HCA and PCA yield dendrograms and scores/loadings maps, respectively, which allows a visual interpretation of the chromatograms. However, these are not classification methods (especially PCA), as they do not assign group memberships. Although much less reported, supervised methods, such as independent modelling of class analogies (SIMCA), linear discriminant analysis (LDA), partial least squares regression (PLSR) or Kohonen neural networks, enable estimates of accuracy and uncertainty of predictions.[11] For example, Vandenabeele and Moens,[12] besides using HCA and PCA, evaluated the LDA method to differentiate natural from synthetic dyestuff profiles, acquired by Raman spectroscopy.

This study focuses on the combination of HPLC-DAD with supervised chemometric methods to distinguish among species of *Dactylopius*. Chromatographic data were analysed by PCA in order to unveil the major patterns, and for the first time, a PLSR method (PLS discriminant analysis) was used to discriminate among the five species of *Dactylopius*. The method was calibrated using a set of reference specimens (known identity and geographical origin) to create a cochineal species reference database. Subsequently, this approach was applied to 72 historical insect specimens of unidentified species, belonging to the Economic Botany Collection (EBC), Royal Botanic Gardens, Kew, England. In order to correctly identify the cochineal species in this historical collection, their chromatographic data were compared with the cochineal reference database, and hence, predictions were quantitatively supported and probabilities were assigned to each estimate.

Theory

Given the complexity of the cochineal chromatograms, multivariate data analysis was required to analyse the results. In this paper principal component analysis (PCA) and partial least squares discriminant analysis (PLSDA) were adopted. The PCA model was adopted to analyse chromatograms of closely related cochineal specimens, in order to identify the main chromatographic features.[13] PLSDA modelling was used to develop chemometric calibration models for the discrimination of cochineal species. This multivariate model is based on the well-known partial least squares regression (PLSR) model,[14,15] and it is specifically used for classifying samples. In this study, if the PLSDA assignment probability for an unknown sample is higher than 0.95, the assignment is considered valid.[15] Otherwise, the result is considered to be non-conclusive.

Chemometric models validation

When dealing with latent variables methods, such as PCA and PLSDA, it is absolutely essential to analyse the statistical validity of the new sample projections, and therefore, the new sample predictions. By projecting a new sample onto a model, it is necessary to verify if the sample lies within the space covered by the model and, in this case, the projection is said to be valid. For PCA and PLSDA methods, Hoteling's T^2 and squared prediction error (SPE) statistics are typically used to assess the validity of the projections. Both are roughly sums-of-squares of the scores and residuals. The former assesses the similarity between the scores of a new sample and the scores of the model calibration samples (deviation within the model space). On the other hand, the latter estimates the residual (error), when the sample is projected onto the model (orthogonal to the model space). The dimension of the model space is defined by the number of latent variables (in PLSDA) or principal components (in PCA) selected. Usually, confidence limits are stipulated for these two statistics, based on the calibration samples. Consequently, these statistics must be under the corresponding confidence limit, normally considering a significance level of 0.05.[14]

Experimental

Insect samples

A total of 75 reference insect specimens of five *Dactylopius* cochineal species were acquired from several entomologists: 33 cochineal specimens of *D. coccus*, *D. ceylonicus*, *D. confusus*, *D. opuntiae* and *D. tomentosus* species, dating from the 17th to early 20th centuries, and from Sri Lanka, Canary Islands (Spain), Madeira (Portugal), USA, Mexico and Argentina, were provided by Douglas Miller; 33 specimens of *D. coccus* and *D. opuntiae*, from Mexico, Peru, Chile, and the Canary Islands (Spain), were given by Liberato Portillo and Mónica González; and 9 specimens of *D. coccus* were purchased from Dott. Alessandro Bizarri (Florence, Italy), Zecchi (Florence, Italy), and Kremer-Pigmente (Aichstetten, Germany). Also, 72 historical insect specimens of unidentified species, belonging to 24 historical sets (3 replicates from each set were analysed), from India, Indonesia, Madeira (Portugal), Honduras, Mexico, Ecuador and Peru, mainly collected during the 19th century, were obtained from the Economic Botany Collection (EBC), Royal Botanic Gardens, Kew (England).

The reduced number of available *D. tomentosus* insects did not allow further chemometric analysis for this species, as only three specimens were available for comparison. Therefore, a dataset comprising 125 chromatograms, resulting from the analysis of replicates from dye extracts belonging to 72 reference insect specimens of *D. coccus*, *D. opuntiae*, *D. confusus* and *D. ceylonicus* species, was selected to construct the reference database using chemometric methods. Subsequently, the same chemometric methods were used to determine the species of 63 insect specimens identified as cochineal among the 72 historical insect specimens of unidentified species (Table 1).

HPLC-DAD analysis

The chromatographic method was adapted from the method described in Serrano *et al.* 2011 (ref. 9) which was based on the validated method described by Casteele *et al.* 1983.[16] Samples preparation and HPLC-DAD analysis were performed using deionised water (Millipore Simplicity® Simpak 2, $R = 18.2$ MΩ cm, USA), methanol (99.9%) from Panreac (Barcelona, Spain),

Table 1 Total number of insect specimens analyzed, as well as the total number of chromatogram replicates obtained with HPLC-DAD analyses and used for chemometric analysis

	Cochineal species	Number of specimens	Number of chromatograms
Reference database	*D. coccus*	45	57
	D. opuntiae	12	23
	D. confusus	9	24
	D. ceylonicus	6	21
	Total	72	125
Historical insects identified as cochineal		63	63
Total		135	187

and perchloric acid from Riedel-de-Haën (Seelze, Germany). The insect specimens were extracted with water, in accordance with historical dyeing recipes.[17] Finely powdered insects (circa 0.2–0.3 mg) from different cochineal species were extracted in 400 μl water, for 10 minutes, in a 60 °C water-bath, with constant mechanical agitation. The dye extracts were then filtered and diluted in water, when necessary (1 : 5, v/v). Prior to HPLC-DAD analysis, they were mixed with methanol and aqueous perchloric acid (H_2O–MeOH–0.3% (v/v) aqueous perchloric acid (50 : 20 : 30, v/v/v)), obtaining a final pH around 4–5.[9] HPLC-DAD analyses were performed on the dye extracts from the insect specimens, using a Thermofinnigan Surveyor HPLC-DAD system with a Thermofinnigan Surveyor PDA 5 diode-array detector (Thermofinnigan, USA), an autosampler and a gradient pump. The sample separation was performed with a reversed-phase column, Zorbax Eclipse Plus C18 (Agilent Tecnologies, CA, USA) with 5 μm particle size column (150 × 2.1 mm), with a flow rate of 0.5 ml min^{-1} at 35 °C constant temperature. The samples were injected *via* a Rheodyne injector with a 25 μl loop. A gradient elution of two solvents, A: pure methanol and B: 0.3% (v/v) aqueous perchloric acid (v/v), was applied: 0–2 min 7A : 93B isocratic, 8 min 15A : 85B linear, 25 min 75A : 25B linear, 27 min 80A : 20B linear, 29 min 95A : 5B linear, 33–40 min 7A : 93B isocratic.[9,18,19]

Data pre-processing

Chromatograms were acquired from 0 to 30 minutes and sampled every second, yielding a 125 × 1800 matrix for the reference insect specimens and a 63 × 1800 matrix for the unidentified specimens. All chromatograms were baseline corrected to remove baseline drifts. Peak alignment was equally performed using the positions of four common peaks: carminic acid, dcIV, dcVII and kermesic and flavokermesic acid (ka + fk) compounds. Ensuring that peaks are conveniently aligned is essential for the application of PCA and PLSDA directly on chromatographic data. Chromatograms were then subjected to standard normal variate processing to remove differences originated by samples concentration differences (eliminating the effect of varying peaks intensity). PCA and PLSDA were used directly on a restricted time region of the processed chromatograms. Before applying PCA or PLSDA, chromatograms were mean-centred. Chromatograms processing was carried out using Matlab version 7.4 release 2007a (Mathworks, Natick, MA).

Multivariate data analysis

The chromatographic data were analysed with PCA to reveal chromatographic patterns and verify which unique chromatographic patterns (or principal components) were directly related to species differentiation. Hence, PCA models were evaluated using the chromatogram retention time between 15 and 25 min, at an absorbance of 275 nm, as the relevant cochineal chromophore peaks were found to be within this region. However, the optimized PCA model for the species discrimination was built using a sub-region of the chromatograms (19.8 to 24.2 min). PCA model projections were evaluated with Hotelling's T^2 and SPE statistics (confidence limits were assessed by using the reference database samples).[13]

As PCA is neither a supervised nor a classification method, a second chemometric approach, PLSDA, was necessary to specifically evaluate the chromatographic based discrimination method between cochineal species. PLSDA requires a more complex validation method for statistical assessment of the results (often cross-validation) and for this reason, many authors validate PLSDA models by considering different calibration and test datasets, which are usually randomly divided.[20] However, this approach does not ensure results with identical probability values of correct assignments, especially when dealing with small datasets. A more reliable validation method previously described by Preisner *et al.* was proposed in this study.[21] Each PLSDA model was calibrated with 70% of randomly selected chromatograms and tested with the remaining 30%. Note that this process is entirely performed using the 72 specimens of the reference cochineal dataset (specimens of the historical dataset are not used to optimise the model). To each random division, a correct balance between the different classes in calibration and test sets was ensured. Additionally, the chromatograms for each insect specimen were permanently in both the calibration and the test sets to avoid over-fitting. The optimal number of latent variables (LV) was estimated by cross-validation (CV), using the calibration set. As model results can be dependent on a particular division between calibration and the test, the classification results were obtained using the average classification performance of 200 PLSDA models, each considering different calibration/test subsets. This procedure was repeated 200 times to ensure statistical validity of the results, at a significance level of 0.01. Results are expressed as confusion matrices, which were obtained through the standardization of the correct classification rate for each predicted class, only considering the test sets. The confusion matrices entries, expressed as percentages, compare the identity of each chromatogram, or sample, with the corresponding prediction. Their aim not only relies on the analysis of the number of correctly predicted species, but also diagnoses those which may be incorrectly predicted, by identifying the most (dis)similar characteristics in terms of chromatographic information. After the optimization of PCA and PLSDA models using the reference cochineal samples, the chromatograms of the historical insect specimens of unidentified species were projected onto these models. Data processing was carried out using Matlab version 7.4 release 2007a

(Mathworks, Natick, MA). PCA and PLSDA models were calculated using the PLS toolbox version 4.2.1 (Eigenvector Research, Manson, WA).

Results and discussion

Dactylopius reference database

One of the major aims of this study was to create a comprehensive cochineal reference database using HPLC-DAD and multivariate data analysis for species of *Dactylopius*. The major North American cochineal species (*D. coccus*, *D. opuntiae*, *D. confusus*, *D. ceylonicus* and *D. tomentosus*) (ref. 3) were analysed with HPLC-DAD, and it was possible to detect carminic acid, as well as other minor compounds, namely dcII, dcIV, dcVII and kermesic and flavokermesic acids (ka + fk) chromophores, through the examination of their elution order, retention time and respective UV spectra, in agreement with the literature.[22] The dcV and dcX compounds, reported by Chavéz-Moreno *et al.*,[3] were not found. Although each species of *Dactylopius* could produce chromatograms with a significant variability, their elution profiles were still very similar, as shown in Fig. 1. For this reason, it was extremely difficult to differentiate among the species by visual examination of the chromatograms. In addition, the quantification of minor markers, according to Wouters and Verhecken,[22] should not be considered an advisable method for discriminating cochineal species, as previously shown by Serrano *et al.*[9]

Chemometric analysis of cochineal reference specimens

The chromatograms belonging to the cochineal species reference database were submitted to PCA. The chromatographic region between 19.8 and 24.2 min was used as it was found to be the most conservative region. The peak corresponding to the carminic acid (CA) centered at 18.98 min was not considered because it does not provide any discrimination between the species. Moreover, it is highly affected by the concentration of the dye extract and, for this reason, it becomes difficult to normalize by using processing correction methods, such as standard normal variate or derivatives. Also, the chromatographic region between 18.98 and 19.8 min was not considered in the analysis, as the variability observed was not consistent with any discriminant feature that could be related to a cochineal species.

The 125 cochineal processed chromatograms were modelled using PCA. The statistically significant number of components was 5.[11] A detailed analysis was performed on all five components in order to observe their relationship with species differentiation. The first (28.6% of captured variance) and the third (10.3% of captured variance) principal components were found to be the most discriminant regarding the species attribute (Fig. 2A). It is worthwhile to notice that these two components only capture 38.9% of the total data variance, showing that other major sources of variability are still present in the chromatograms. It was found to be very difficult to model these data with only a few components since many variations were observed between samples of different species and even within species. For this reason, it was found to be extremely important to apply a supervised method, such as PLSDA, for further analysis of the chromatographic data.

The domesticated *D. coccus* species could be distinguished from the wild species of *Dactylopius* (*D. opuntiae*, *D. confusus* and *D. ceylonicus*), as samples exhibit strong negative scores on component 1, in contrast to the other species (Fig. 2A). These were also differentiated into three groups, each corresponding to a species, and showing some degree of separation. It is clear that *D. coccus* samples exhibit a higher congregation, indicating a lower variability of the corresponding chromatograms. Nevertheless, it is noticeable that some scores are lying on clusters corresponding to other species. The validation of this chart is only ensured by the analysis of the Hotelling's T^2 and SPE statistics (Fig. 2B).

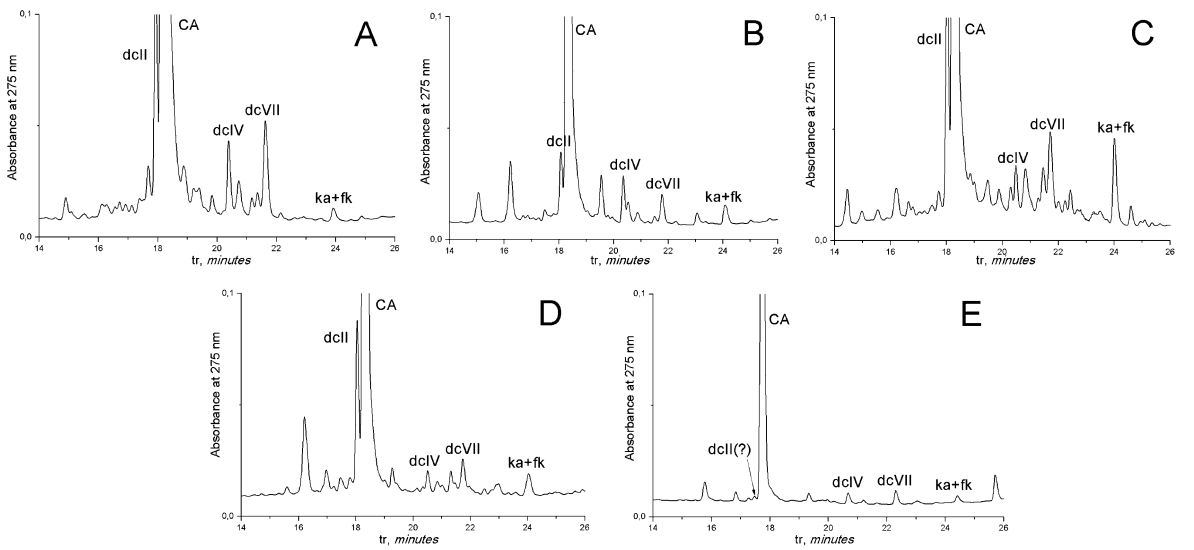

Fig. 1 HPLC-DAD chromatograms, acquired at 275 nm, of dye extracts from species of *Dactylopius*: (A) *D. coccus*, (B) *D. opuntiae*, (C) *D. confusus*, (D) *D. ceylonicus* and (E) *D. tomentosus*.

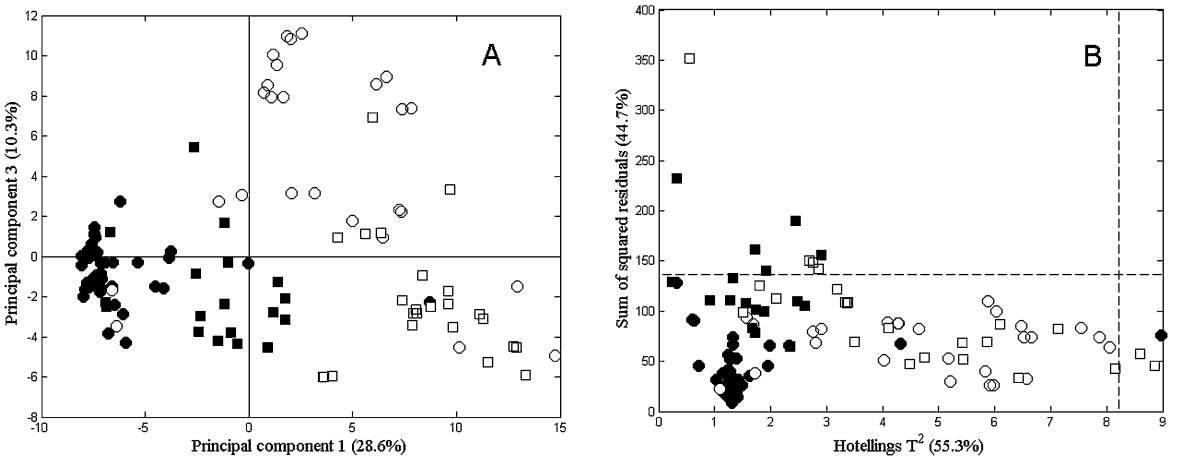

Fig. 2 (A) PCA scores representing 38.9% of total data variance of HPLC-DAD chromatograms (19.8 to 24.15 min) acquired at 275 nm for cochineal reference insects and (B) Hotelling's T^2 and SPE statistics for the same specimens with indication of 95% confidence limits. Caption: *D. coccus* (black circles), *D. opuntiae* (white circles), *D. confusus* (white squares) and *D. ceylonicus* (black squares).

The representation of these statistics reveals that there are two samples (one *D. ceylonicus* and one *D. confusus*) which are significantly above the 95% confidence limit for the SPE statistic, although they are within the correct groups in Fig. 2A. On the other hand, the *D. coccus* samples with high values of the Hotelling's T^2 statistic (4.31 and 8.98) correspond to those pointed out in Fig. 2A, lying on the *D. confusus* and *D. ceylonicus* clusters. The high variability observed for this particular sample makes this result non-conclusive.

Subsequently, a PLSDA model was built using the same chromatograms and chromatographic regions of the reference database, used in the PCA model (Table 2). Although PLSDA is a relatively robust supervised model it is not totally insensitive to variations in the independent block (the chromatograms) that are not related to the dependent variable (the species).[13] For that reason it was decided to keep the spectral region used before in PCA. When trying to model these data using the entire chromatographic region containing peaks (15–25 min), the model performance was not only substantially poorer but the number of components was higher than that needed using the restricted time region (data not shown).

This PLSDA model was optimized in terms of the number of latent variables using cross-validation and revealing an optimal number of 5. The PLSDA model confirmed the results of the PCA that the accurate discrimination among the four species of *Dactylopius* is difficult. The major difficulties arose with the differentiation between *D. opuntiae* and *D. confusus*, and between *D. coccus* and *D. ceylonicus*. PLSDA yielded circa 78% of correct assignments. For further analyses of this result, samples which had low reproducibility were detected by looking at the PLSDA scores. A hierarchical cluster analysis was applied to the first five PLSDA scores, and the samples, for which the replicates were not fully linked, were considered outliers. In total, eight samples had a large variation between replicates. In general, the chromatographic results of the insect specimens belonging to the same insect species have shown similar reproducibility, although some exceptions were detected. This was indeed the case of these eight samples (or chromatograms).

As a natural organic dyestuff, cochineal comprises other biological molecules and residues besides the red coloring agent and, for this reason, the weight of the dyestuff measured in each sample preparation might not correspond accurately to the final concentration of the dye extract. As a result, some chromatograms might display a different chromatographic behaviour, commonly expressed in smaller or larger peak areas, from that shown by other chromatograms obtained with the analysis of other insect species. For this evidence these eight samples were excluded and a new PLSDA model was constructed. It provided an increase of the number of correct assignments to 94.3% (Table 2). This model demonstrated that a small number of samples of *D. opuntiae* and *D. confusus* were not correctly predicted, *e.g.*, about 5% of the samples only considering these two species.

The results achieved with this model showed that the reproducibility of the HPLC-DAD analysis is extremely important for obtaining more accurate multivariate models. Before attempting to use this model for any prediction, the reproducibility of the chromatograms should be validated. This can be ensured by projecting the chromatograms onto the PCA and PLSDA models to verify the consistency of the Hotelling T^2 and SPE statistics.

To further analyse the chromatographic regions, responsible for the discrimination of the four insect species, the squared regression coefficients for the PLSDA model were superimposed

Table 2 Confusion matrix for the PLSDA model of the cochineal species, considering the chromatographic region between 19.80 and 24.15 min, and five latent variables (values are in percentages)

Predicted species (%)	*D. coccus*	*D. opuntiae*	*D. confusus*	*D. ceylonicus*	Total
D. coccus	47.0	1.8	0.1	0.8	49.6
D. opuntiae	0.7	16.5	1.3	0.0	18.4
D. confusus	0.0	1.1	20.1	0.0	21.2
D. ceylonicus	0.0	0.0	0.0	10.7	10.7
Total	48	19	22	12	100.0

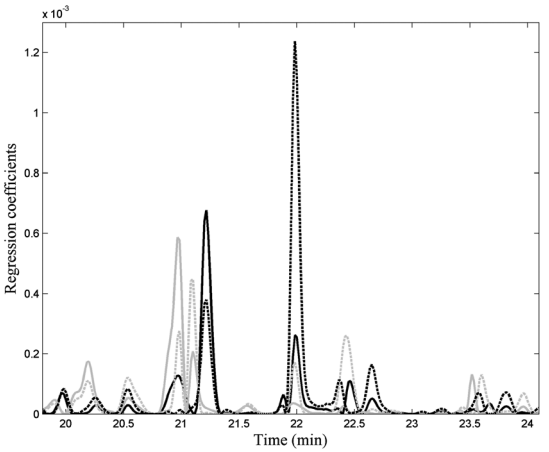

Fig. 3 Regression coefficients for the PLSDA model discriminating the four species: *D. coccus* (——), *D. opuntiae* (——), *D. confusus* (— —) and *D. ceylonicus* (— —).

for comparison purposes (Fig. 3). A PLSDA model generates one regression coefficient for each modelled class. Hence, there are four classes (species) modelled, and four regression coefficients. From a first observation, very prominent similarities seem to exist between coefficients for *D. coccus* and *D. ceylonicus*, confirming the resemblance between these two species, as shown before. This is particularly evident with the peaks at 21.2 and 22 min; indeed, the peak at 21.2 min is highly characteristic of these two species. Nevertheless, there are minor differences in the chromatographic region of these species, such as those displayed around 22.5 min. Also, a similar result was observed for *D. opuntiae* and *D. confusus*, as revealed by the presence of peaks at 20.97 and 21.09 min. The latter is very characteristic of these two species, although another peak at 22.43 min seems to be more characteristic of *D. confusus*.

Due to the similarity exhibited by the species (*D. coccus* and *D. ceylonicus*, and *D. opuntiae* and *D. confusus*), two other PLSDA models were built, each only using the information from the similar species. The PLSDA model built to discriminate between *D. opuntiae* and *D. confusus* revealed a 100% correct discrimination. The same result was obtained for a model discriminating between *D. coccus* and *D. ceylonicus*. For a two class discrimination model, there are only two regression coefficients and these are symmetric. Therefore, each model can be represented by only one coefficient (Fig. 4), allowing the visualization of the locations and improving discrimination between the two groups of species. For the *D. coccus*/*D. ceylonicus* model, the discriminating regions are centered at 21.21, 21.98 and 22.65 min. For the *D. opuntiae*/*D. confusus* model, the discriminating regions are centered at 20.98, 21.10 and 22.45 min. Indeed, these models perform very accurate distinction of the reference database samples.

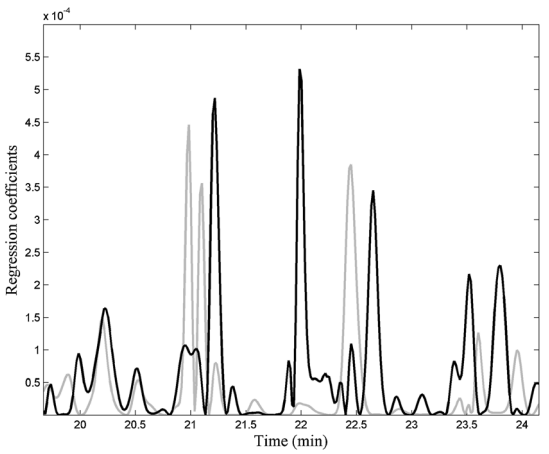

Fig. 4 Regression coefficients for two PLSDA models discriminating between two species: *D. coccus* and *D. ceylonicus* (——) and between *D. opuntiae* and *D. confusus* (——).

Historical cochineal insects from Kew's Economic Botany Collection

From the group of 72 historical insect specimens analysed, only 63 were identified as cochineal, owing to the detection of carminic acid and the minor cochineal compounds, namely dcII, dcIV, dcVII and fk + ka. These results were in agreement with the provenances and dates provided by the labels from the EBC specimens. Indeed, these correspond to a period of global interest and experimentation in the culture of the domesticated cochineal *D. coccus*, in such regions as India or Australia.[4,23–27] As for the remaining nine specimens, yellow unidentified compounds were detected in six of them (EBC 55338 and 55340 sets), while three others produced colorless extracts (EBC 73057 set), hence excluding these as being species of *Dactylopius* (Table 3). Therefore, circa 13% of the historical EBC insects labelled as cochineal were considered here as insects of an unknown source.

Chemometric assisted predictions

The chromatographic data corresponding to the historical cochineal insect specimens of unidentified species were then analysed using the PCA and PLSDA models previously obtained. These data were processed using a similar strategy as followed for the calibration data. The results were projected onto the PCA model yielding score projections as represented in Fig. 5A, and then these projections were analysed together with the corresponding statistics (Fig. 5B).

Samples 23 (EBC 54394), 31 (EBC 54404) and 42 (EBC 58236.2) do not fit the model and therefore, results for these samples cannot be considered valid. Samples 7, 8 and 9 (EBC 54389) appeared within the *D. opuntiae* group (Fig. 5A) and displayed higher values relative to Hotelling's T^2 (Fig. 5B) when compared with the remaining samples.

This was also reported for samples 40 and 41 (EBC 58236.2), which were found in the region between the *D. coccus* and the *D. opuntiae* groups (Fig. 5A). The other projected scores have shown that the majority of the historical cochineal samples were located within the region spanned by *D. coccus*, thus indicating a high similarity with this species. The purpose of projecting the unknown specimens in the PCA model was not to classify them (as stated before, PCA is not a classification

Table 3 Historical cochineal specimens which were labelled as "cochineal" in the Economic Botany Collection (EBC), Royal Botanic Gardens, Kew, with their respective species attributions, as determined by HPLC-DAD, PCA and PLSDA analyses

EBC sample sets classification[a]	Chemometric analyses samples nr.	Collection date	Source, donor	Samples description	Species attribution/ observations
54387	1–3	1856	India, Madras, from James A. Mann	Dry, brown, medium dimensions, 10.4 ± 3.0 mg	*D. coccus* Specimens 1–3 in *D. coccus* region
54388	4–6	1918	Peru, Callas, from London Drug Market	Dry, black, medium dimensions, 14.2 ± 2.8 mg	*D. coccus* Specimens 4–6 in *D. coccus* region
54389	7–9	1899	Ecuador, Chimborazo Province, Guana, from Edward Whymper	Dark dry cake mixture of cochineal and other ingredients, 1.0 ± 0.2 mg	*D. opuntiae* (?) Specimens 7–9 near *D. opuntiae* region
54390	10–12	Before 1879	Indonesia, Java, Buitenzorg, Tyikoppo, donated by India Museum	Dry, dark brown, medium dimensions, 14 ± 2 mg	*D. coccus* Specimens 10–12 in *D. coccus* region
54391	13–15	Before 1879	Indonesia, Java, Buitenzorg, Bandok, donated by India Museum	Dry, dark brown, medium dimensions, 16.2 ± 1.9 mg	*D. coccus* Specimens 13–15 in *D. coccus* region
54392	16–18	Late 19th century	A. S. Hill & Son, London[b]	Dry, dark brown, medium dimensions, 21.7 ± 5.2 mg	*D. coccus* Specimens 16–18 in *D. coccus* region
54393	19–21	Before 1879	India, Andhra Pradesh, Scinde, Hyderabad, donated by India Museum	Dry, shiny red, medium dimensions, 2.8 ± 1.1 mg	*D. coccus* Specimens 19–21 in *D. coccus* region
54394	22–24	1867	India, Calcutta, from International exhibition, Paris	Dry, dark brown, medium dimensions, 11 ± 2.8 mg	*D. coccus* Specimen 23 not valid Specimens 22 and 24 in *D. coccus* region
54402	25–27	Late 19th century	Honduras and Vera Cruz	Dry, light brown, little dimensions, 1.6 ± 0.4 mg	*D. coccus* Specimens 25–27 in *D. coccus* region
54403	28–30	Before 1879	India, Punjab, donated by India Museum	Dry, dark brown, medium dimensions, 11 ± 4.7 mg	*D. coccus* Specimens 28–30 in *D. coccus* region
54404	31–33	Probably 1851	Mexico, Oaxaca, from J. Sadler, probably International exhibition, London	Dry, dark brown, medium dimensions, 10.8 ± 1.7 mg	*D. coccus* Specimen 31 not valid Specimens 32 and 33 in *D. coccus* region
54410	34–36	1977	Madeira, from Jane Stubbs	Dry, greyish brown, medium dimensions, 7.8 ± 1.4 mg	*D. coccus* Specimens 34–36 in *D. coccus* region.
55338	—	1855	Australia (NSW), from International exhibition, Paris	Dried naturally, rusted red colour, little dimensions, 2.1 ± 0.3 mg	Not cochineal insect, unidentified yellow compounds (main peak: rt = 19.86 min, λ_{max} = 421 nm)
55340	—	1862	Australia (Victoria), from International exhibition, London	Dried naturally, rusted yellow colour, little dimensions, 1.8 ± 0.7 mg	Not cochineal insect, unidentified yellow compounds (main peak: rt = 19.86 min, λ_{max} = 421 nm)
58236.1	37–39	Late 19th century	Ripley, Roberts & Co. 3. Mincing lane[b]	Dry, dark shiny red, medium dimensions, 20.7 ± 1.7 mg	*D. coccus* Specimens 37–39 in *D. coccus* region
58236.2	40–42	Late 19th century	Ripley, Roberts & Co. 3. Mincing lane[b]	Dry, dark shiny red, medium dimensions, 15.9 ± 2.0 mg	*D. coccus* Specimen 42 not valid Specimens 40 and 41 assigned as *D. coccus*
58236.3	43–45	Late 19th century	Beazley & Co. Dunster House, Mincing Lane[b]	Dry, dark shiny red, medium dimensions, 18.7 ± 2.5 mg	*D. coccus* Specimens 43–45 in *D. coccus* region;l
58236.4	46–48	Late 19th century	Beazley & Co. Dunster House, Mincing Lane[b]	Dry, black, medium dimensions, 16.3 ± 6.9 mg	*D. coccus* Specimens 46–48 in *D. coccus* region

Table 3 (*Contd.*)

EBC sample sets classification[a]	Chemometric analyses samples nr.	Collection date	Source, donor	Samples description	Species attribution/ observations
58236.5	49–51	Late 19th century	Ripley, Roberts & Co. 3. Mincing lane[b]	Dry, salmon light colour, medium dimensions, 19.5 ± 4.7 mg	*D. coccus* Specimens 49–51 in *D. coccus* region
58236.6	52–54	Late 19th century	Ripley, Roberts & Co. 3. Mincing lane[b]	Dry, black, little dimensions, 0.7 ± 0.3 mg	*D. coccus* Specimens 42–54 in *D. coccus* region
58236.7	55–57	Late 19th century	Ripley, Roberts & Co. 3. Mincing lane[b]	Dry, dark brown, medium dimensions, 10.6 ± 3.5 mg	*D. coccus* Specimens 55–57 in *D. coccus* region
58236.8	58–60	Late 19th century	Beazley & Co. Dunster House, Mincing Lane[b]	Dry, dark red, medium dimensions, 5.0 ± 1.0 mg	*D. coccus* Specimens 58–60 in *D. coccus* region
73057	—	1800–1857	Royal Pharmaceutical Society of Great Britain (Museum)[b]	Naturally dried, orange colour, big dimensions, 7.20 mg	Not cochineal insect. No coloured compounds.
73237	61–63	1851	Mexico, Oaxaca, from J. Sadler, probably International exhibition, London	Dry, black, medium dimensions, 11.6 ± 2.4 mg	*D. coccus* Specimens 61–63 in *D. coccus* region.

[a] Classification of the insects, given by the donor. [b] Unidentified source.

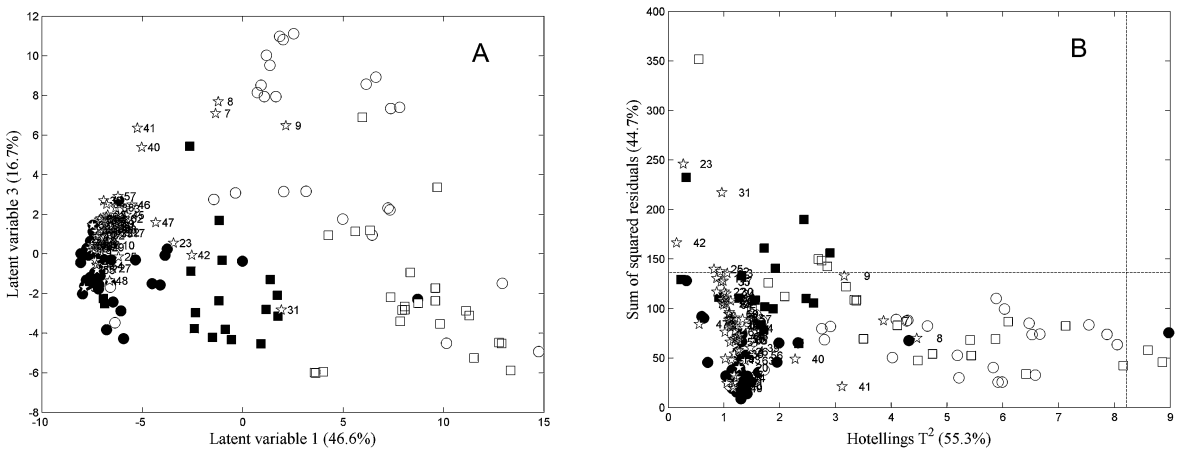

Fig. 5 (A) PCA scores representing 38.9% of total data variance of HPLC-DAD chromatograms (19.8 to 24.15 min) acquired at 275 nm for cochineal reference insects together with the projected scores from the Kew EBC specimens and (B) Hotelling's T^2 and SPE statistics for the same replicates with indication of 95% confidence limits. Caption: *D. coccus* (black circles), *D. opuntiae* (white circles), *D. confusus* (white squares), *D. ceylonicus* (black squares) and Kew EBC replicates (☆).

method) but as a mean to observe the degree of fitting of their chromatograms to the reference database specimen chromatograms.

To obtain species predictions, the historical cochineal samples were projected onto the PLSDA model built with the reference database samples, excluding the results with lower reproducibility. The projections not only revealed a class assignment for each sample, but also an associated probability (Table 4).

Only samples 23, 31 and 40–42 were found to hold associated probabilities lower than 95%. Therefore, predictions for these samples should not be considered. In fact, chromatograms from set EBC 58236.2 (composed by samples 40–42) were uncertainly attributed, with a very low probability to *D. coccus* (samples 40 and 41) and to *D. ceylonicus* (sample 42). As for the remaining samples, all were predicted as *D. coccus* except samples 7–9 (EBC 54389) which were attributed to *D. opuntiae*.

These predictions show that the majority of the historical insect specimens are *D. coccus* and that samples in EBC 54389 are most likely *D. opuntiae*. These samples were projected onto the PLSDA model built for discriminating between *D. opuntiae* and *D. confusus*, and all three samples were shown to belong to *D. opuntiae*, with a probability higher than 99%. This confirms the results from the first model. In fact, this set belongs to a

Table 4 Class assignments for the historical cochineal specimens, based on the PLSDA model estimated using the reference database

Historical replicates	Species assignment	Species membership assignment probability
1–6, 10–22, 24–30, 32–39, 43–65	*D. coccus*	>0.99
7–8	*D. opuntiae*	>0.98
9	*D. opuntiae*	0.95
40, 41	*D. coccus*	>0.85
31	*D. coccus*	0.62
23, 42	*D. ceylonicus*	<0.61

cake, in which the insects were probably mixed with other unknown components.[4,23] The preparation method of this cake is typical for the indigenous people of the Americas and it is very likely that they often used species of *Dactylopius* that were available locally,[23,25] which is in agreement with the information provided by Kew's EBC.

It is noteworthy that samples 25–27 (EBC 54402) and 52–54 (EBC 58236.6) were not expected to be identified as domesticated cochineal, due to their low weight (Table 3) and small appearance (Table 5). However, the results confirm that they were *D. coccus*, illustrating that individual specimens can vary in size, which might reflect differences in culture conditions. This could result in incorrect labelling. Indeed, products sold as "granilla" cochineal were usually composed of fragments of *D. coccus* or small individuals, which were separated from the good quality ones and sold at a lower price.[27]

The final appearance and value of cochineal depended on the environmental conditions in which the insects were grown, and on the dyestuff preparation treatments. Hence, it was a common assumption that these factors could influence the yield of colorant in the dyeing procedure. Table 5 depicts several EBC insects which were treated using different preparation procedures and which strongly influenced the final colour, size and weight of the specimens.[4,23,27,28] For instance, the best method for producing "silver" cochineal consists of the slow drying of the insects under the sun and it can be exemplified by set EBC 54404 (samples 31–33). The method involving the use of hot metal sheets or pans that produce "black" cochineal is shown by set EBC 73237 (samples 61–63). On the other hand, the immersion of the cochineal insects in boiling water with subsequent drying with sun exposure and mixing with grains of sand produces the glossy varieties of "dull red" or "rosy black" cochineal, which are well exemplified by set EBC 58236.1 (samples 37–39). In spite of their very heterogeneous appearance, these samples were all identified as *D. coccus*. Sets EBC 58236.1 to 58236.8 (samples 37–60, all identified as *D. coccus*) belong to a collection of eight cochineal varieties, dated from the late 19th and early 20th centuries, which were used for dyeing purposes. The corresponding HPLC-DAD chromatograms and chemometric models have shown that the cochineal composition is very similar among the specimens and, for this reason, the dyestuff content does not significantly change between insects, evidencing different types of dyestuff treatment. For this reason, the preparation procedure did not seem to affect the identification of the species. These observations are important for understanding past traditions of preparing American cochineal insects for dyeing purposes, and moreover, their significance for human history, historical value in museum collections, and the necessity of preserving this knowledge for future generations.

Table 5 Heterogeneous appearance of *D. coccus* cochineal insects which have been submitted to different preparation treatments for commercial purposes

Classification	Photo (amplification)
54404 silver cochineal	 200×
73237 black cochineal	 200×
58236.1 rosy black cochineal	 200×
58236.6 wild cochineal	 500×
54402 inferior cochineal	 400×

The successful and pioneering discrimination of *Dactylopius* cochineal species by allying HPLC-DAD with such statistical methods, like PCA and PLSDA, constitutes a very precise and important contribution for characterizing historical works of art,[9] especially those dated up to the 19th century and which may have been dyed with wild cochineal species, and also, for solving current entomological issues, namely the origins of the culture of *D. coccus*, among early pre-Colombian cultures.[6] Therefore, it is highly recommended that the remaining *Dactylopius* cochineal species are analysed, in order to obtain more accurate and definitive responses to questions raised by the various disciplines interested in this subject.

Conclusions

This study could successfully discriminate for the first time species of *Dactylopius* cochineal insects using a non-supervised PCA method, along with a supervised PLSDA method based on HPLC-DAD chromatograms. The construction of a reference library with such an accurate analytical system and with such precise statistical methods allowed the subsequent identification and characterization of Kew's Economic Botany Collection of historical cochineal insect specimens of unidentified species. The majority of the historical cochineal specimens were identified as *D. coccus*, while samples belonging to the set EBC 54389 were identified as *D. opuntiae*. Owing to their macroscopic

appearance some samples did not seem initially to be *D. coccus*, although the chemometric methods applied here have successfully shown that it is possible to analyse samples that vary depending on raising and preparation treatment parameters.

The development of a chromatographic reference database with the HPLC-DAD analysis of taxonomically verified species of *Dactylopius*, along with the use of PCA and PLSDA methods, provides a powerful platform to analyse cochineal insects, and it is therefore advisable for further identification and characterization of *Dactylopius* cochineal species in objects of cultural heritage in order to confirm their date and provenance.

Acknowledgements

We are very grateful for the financial support given by the Fundação para a Ciência e Tecnologia (grants PEst-C/EQB/LA0006/2011 and SFRH/BD/73409/2010), the FEDER (under contract POCI/QUI/099388/2008) and the Access to Research Infrastructures activity in the 7th Framework Programme of the EU (CHARISMA Grant Agreement no. 228330). We also thank Douglas Miller (Agricultural Research Service, Systematic Entomology Laboratory, Maryland, U.S.A), Liberato Portillo (Botanical and Zoology Department, University of Guadalajara, Mexico), and Mónica González (Instituto Canario de Investigaciones Agrarias, Tenerife, Canary Islands), for the cochineal insect specimens.

References

1 P. Anastas and N. Eghbali, *Chem. Soc. Rev.*, 2010, **39**, 301.
2 D. Cardon, *Proceedings of the 12th Biennial Symposium Textile Society of America*, Lincoln, 2010.
3 C. K. Chávez-Moreno, A. Tecante, M. Fragoso-Serrano and R. Pereda-Miranda, *Biochem. Syst. Ecol.*, 2010, **38**, 671.
4 D. Cardon, *Natural Dyes – Sources, Tradition, Technology and Science*, Archetype, London, 2010.
5 C. K. Chávez-Moreno, A. Tecante and A. Casas, *Biodivers. Conserv.*, 2009, **18**, 3337.
6 L. Portillo, *Dugesiana*, 2005, **12**, 1.
7 F. E. Lancaster and J. F. Lawrence, *J. Chromatogr., A*, 1996, **732**, 394.
8 M. Gonzaléz, *Food Control*, 2005, **16**, 105.
9 A. Serrano, M. M. Sousa, J. Hallett, J. A. Lopes and M. C. Oliveira, *Anal. Bioanal. Chem.*, 2011, **401**, 735.
10 J. Méndez, M. González, M. G. Lobo and A. Carnero, *J. Agric. Food Chem.*, 2004, **52**, 1331.
11 N. Tormod, I. Tomas, T. Fearn and D. Tony, *A User-Friendly Guide to Multivariate Calibration and Classification*, NIR Publications, Chichester, 2002.
12 P. Vandenabeele and L. Moens, *Analyst*, 2003, **128**, 187.
13 R. G. Brereton, *Chemometrics: Data analysis for the laboratory and chemical plant*, John Wiley & Sons Ltd, Chicester, 2003.
14 P. Geladi and B. Kowalski, *Anal. Chim. Acta*, 1986, **185**, 1.
15 B. K. Alsberg, D. B. Kell and R. Goodacre, *Anal. Chem.*, 1998, **70**, 4126.
16 K. V. Casteele, H. Geiger, R. de Loose and C. F. van Sumere, *J. Chromatogr.*, 1983, **259**, 291.
17 V. Golikov, *Dyes in history and archaeology*, 1998, **16**/**17**, 21.
18 R. Marques, M. M. Sousa, M. C. Oliveira and M. J. Melo, *J. Chromatogr., A*, 2009, **1216**, 1395.
19 M. J. Melo, M. M. Sousa, A. J. Parola, J. S. S. Melo, F. Catarino, J. Marçalo and F. Pina, *Chem.–Eur. J.*, 2007, **13**, 1417.
20 M. Bylesjo, M. Rantalainen, O. Cloarec, J. K. Nicholson, E. Holmes and J. Trygg, *J. Chemom.*, 2006, **20**, 341.
21 O. E. Preisner, R. Guiomar, J. Machado, J. C. Menezes and J. A. Lopes, *Appl. Environ. Microbiol.*, 2010, **76**, 3538.
22 J. Wouters and A. Verhecken, *Ann. Soc. Entomol. Fr.*, 1989, **25**, 393.
23 R. A. Donkin, *Trans. Am. Philos. Soc.*, 1977, **67**, 1.
24 C. S. Silva and M. S. Bosa, *Rev. Indias*, 2006, **66**, 473.
25 C. K. Chávez-Moreno, A. Tecante and A. Casas, *Biodivers. Conserv.*, 2009, **18**, 3337.
26 A. Roquero, *Tintes y Tintoreros de América – Catálogo de materias primas y registro etnográfico de México, Centroamérica, Andes Centrales y Selva Amazónica*, Instituto del Patrimonio Historico Español, Madrid, 2006.
27 W. Born, *Ciba Rev.*, 1938, **1**, 206.
28 R. L. Lee, *The Americas*, 1948, **4**, 449.