PAPER

Discrimination of Diptera order insects based on their saturated cuticular hydrocarbon content using a new microextraction procedure and chromatographic analysis†

L. O. León-Morán, [a] M. Pastor-Belda, [ac] P. Viñas, [ac] N. Arroyo-Manzanares, [ac] M. D. García, [bc] M. I. Arnaldos [bc] and N. Campillo *[ac]

The nature and proportions of hydrocarbons in the cuticle of insects are characteristic of the species and age. Chemical analysis of cuticular hydrocarbons allows species discrimination, which is of great interest in the forensic field, where insects play a crucial role in estimating the minimum post-mortem interval. The objective of this work was the differentiation of Diptera order insects through their saturated cuticular hydrocarbon compositions (SCHCs). For this, specimens fixed in 70 : 30 ethanol : water, as recommended by the European Association for Forensic Entomology, were submitted to solid–liquid extraction followed by dispersive liquid–liquid microextraction, providing preconcentration factors up to 76 for the SCHCs. The final organic extract was analysed by gas chromatography coupled with flame ionization detection (GC-FID), and GC coupled with mass spectrometry was applied to confirm the identity of the SCHCs. The analysed samples contained linear alkanes with the number of carbon atoms in the C9–C15 and C18–C36 ranges with concentrations between 0.1 and 125 ng g^{-1}. *Chrysomya albiceps* (in its larval stage) showed the highest number of analytes detected, with 21 compounds, while *Lucilia sericata* and *Calliphora vicina* the lowest, with only 3 alkanes. Non-supervised principal component analysis and supervised orthogonal partial least squares discriminant analysis were performed and an optimal model to differentiate specimens according to their species was obtained. In addition, statistically significant differences were observed in the concentrations of certain SCHCs within the same species depending on the stage of development or the growth pattern of the insect.

Introduction

Cuticular hydrocarbons (CHCs) are long chains of saturated, unsaturated, and branched hydrocarbons found in the exoskeleton of insects. The composition of CHCs enables discrimination between and within species based on the number of compounds and proportions or chain lengths, which are associated with the stage and nature of the specimen.[1] This discriminatory capacity is of great interest in forensic practice, where identifying a specimen is often limited by the lack of adequate identification keys, the state of the collected specimen (insect remains) and time required for the identification of larval specimens. The correct identification of the fauna that colonises a corpse is crucial in forensic entomology. Each species has a distinct colonisation behaviour and life cycle, which must be considered when estimating the minimum post-mortem interval (mPMI).[2] Fauna associated with a corpse is mainly composed of necrophagous arthropods, with Diptera (such as flies or mosquitoes) being the most investigated order in the forensic field.[3–8] Within Diptera order, Calliphoridae is the most widely studied family regarding CHCs[4,7,9–14] followed by Sarcophagidae[10,15] and Phoridae.[3,10] The species that have attracted the most interest are *Lucilia sericata*,[10,13,16] *Calliphora vicina*,[6,12,16] *Calliphora vomitoria*[6,16,17] and *Chrysomya rufifacies*.[7,10,18] The colonising behaviour of the species is related to the predatory or non-predatory nature of the specimens. Thus, species such as *L. sericata*, *C. vicina* or *C. vomitoria* are classified as primary colonisers. In contrast, *Ch. albiceps* or *Synthesiomyia nudiseta* are known for their predatory behaviour towards other larvae and, consequently, are considered secondary colonisers.[19,20]

aDepartment of Analytical Chemistry, Faculty of Chemistry, Regional Campus of International Excellence "Campus Mare Nostrum", University of Murcia, E-30100 Murcia, Spain. E-mail: ncampi@um.es; Tel: +34 868887320

bDepartment of Zoology and Physical Anthropology, Faculty de Biology, Regional Campus of International Excellence "Campus Mare Nostrum", University of Murcia, E-30100 Murcia, Spain

cExternal Service of Forensic Sciences and Techniques (SECyTeF), Regional Campus of International Excellence "Campus Mare Nostrum", University of Murcia, E-30100 Murcia, Spain

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4ay00214h

The estimation of mPMI relies on the characterization of specimens from the cadaveric fauna. The determination of CHCs is a key factor in species identification and interspecies discrimination, which is widely carried out considering alkanes, methyl-branched alkanes and alkenes.[7,11,15,17,21,22] However, in some cases the discriminatory capacity has been addressed by considering only the saturated cuticular hydrocarbon composition (SCHC).[12,13,23]

The sample treatment commonly used for the determination of CHCs in insects consists of solid–liquid extraction (SLE) in hexane,[3,4,6,7,9,17,21,22,24] followed by a purification step using a silica gel column[6,9,17,21] or activated magnesium silicate.[3,22] Preconcentration is generally carried out by solvent evaporation and reconstitution of the dry residue in a lower volume of the solvent.[15,22] Other solvents used for SLE have been petroleum ether followed by dichloromethane[13] and acetonitrile (ACN).[25]

With the aim of enhancing method sensitivity, SLE has been combined with solid phase microextraction (SPME) by direct immersion of the fiber into the SLE extract.[25] SPME was also applied by directly contacting the SPME fiber to the insect body,[26] or by fiber exposure to the sample headspace over the heated specimen.[27] In all cases, analysis by gas chromatography (GC) is carried out after fiber thermal desorption. Although the literature does not show the use of microextraction techniques other than SPME for CHC determination in insects, other miniaturized extraction methods are likely to be applied. Dispersive liquid–liquid microextraction (DLLME)[28] has been demonstrated to be efficient for alkane extraction in matrices such as petroleum-contaminated water,[29] apple peel[30] and plants[31,32] samples. DLLME is based on a ternary solvent system where an organic solvent (extractant phase) breaks into tiny droplets through the aqueous phase with the help of a dispersant solvent. This methodology is characterized by its simplicity, low cost, great solvent and sample saving, and minimal waste generation.

GC is the most widely employed instrumental technique for CHC determination combined with flame ionization detection (FID)[8,23] and mass spectrometry (MS),[4,9,16,25] and in some cases both (GC-FID and GC-MS) are used in a complementary way.[5,11,33,34]

The application of statistical analyses is also shown in the literature. Principal component analysis (PCA) has been used to differentiate larval age[7,21] and the first instar larvae species of Calliphoridae.[17] PCA in combination with a support vector machine for pupae species classification[15] and an unsupervised artificial neural network for larvae[6] and adult[16] age discrimination has also given satisfactory results. Other tools such as Spearman correlation with cluster analysis and PERMANOVA have been used for adult age differentiation[18,22] and orthogonal partial least squares discriminant analysis (OPLS-DA) for larvae age differentiation.[11]

The analysis of CHCs in forensic specimens is typically performed on fresh or cold-preserved samples.[6,11,14,25] However, preservation in ethanol (EtOH) is a standard practice imposed by the European Association for Forensic Entomology (EAFE).[35] Recent studies have demonstrated that the profile of CHCs is preserved over long storage periods, even though a minimum portion of CHCs is extracted by the fixed medium.[36] Additionally, the study of samples preserved for extended periods may be useful in forensic counter-examinations. The aim of this work is to distinguish specimens of forensic interest through a targeted analysis specifically focused on SCHC determination in EtOH-fixed samples. For this purpose, a new and sensitive analytical method based on a combined SLE-DLLME procedure with GC-FID analysis is proposed for the quantification of SCHCs at trace levels. Identity confirmation of the targeted compounds was carried out by GC-MS. A combination of non-supervised PCA and supervised OPLS-DA was applied using the SCHC concentration in each sample to classify specimens according to their species.

Experimental

Chemicals and reagents

Chromatographic quality organic solvents ACN, methanol (MeOH), EtOH and chloroform were obtained from Sigma (St. Louis, MO, USA). Distilled water was purified using a Milli-Q system (Millipore, Bedford, MA, USA). Sodium chloride purchased from Sigma was used. Method optimization was performed using a standard mixture solution of alkanes, containing 35 alkanes which included linear (from n-octane (C8) to n-tetracontane (C40)), and the two branched (pristane and phytane), dissolved in dichloromethane at a concentration level of 500 mg L^{-1} (Sigma). The purity percentages ranged from 94 to 100%, corresponding to nonatriacontane (C39) and n-octadecane (C18), respectively.

Instrumentation

GC analyses were carried out using an Agilent 7820 instrument (Agilent Technologies, Santa Clara, CA, USA) equipped with an Agilent 7683 autosampler and coupled to an Agilent FID 7820 (Agilent Technologies). A non-polar analytical column HP-5 ((5% phenyl) methylpolysiloxane) with dimensions of 0.32 mm × 30 m and 0.25 μm film thickness (Agilent Technologies, Diegem, Belgium) was used. The oven programme started with an isothermal step of 1 min at 35 °C, followed by a linear gradient of 10 °C min^{-1} to reach a temperature of 320 °C, being maintained for 10.5 min. The total time of the chromatographic run was 40 min. The injection (1 μL) was carried out in splitless mode at 350 °C, with samples contained in vials of 2 mL capacity equipped with 250 μL micro-inserts. A glass liner with no filler with a capacity of 400 μL was used. Nitrogen (purity > 99.999%) provided by Messer Ibérica (Tarragona, Spain) was employed as the mobile phase with a constant column flow of 3 mL min^{-1}. The FID detector temperature was set at 300 °C. Flame ignition conditions were adjusted to 55 mL min^{-1} for hydrogen (used as fuel, purity > 99.999%), 450 mL min^{-1} for synthetic air (used as oxidizer, 20.5% O$_2$, 79.5% N$_2$) and 50 mL min^{-1} for nitrogen (as "make up" gas) applied for analytical signal improvement. All gases involved in the flame ignition were provided by Messer Ibérica. Under the specified conditions, the compounds eluted between 3.4 and 32.5 min, corresponding to C8 and C40, respectively.

GC-MS analyses were carried out with an Agilent 6890A gas chromatograph equipped with an Agilent 7693 autosampler and coupled to an Agilent 5975C MSD Triple-Axis mass spectrometer (Agilent Technologies). An analytical column Zebron ZB-5PLUs (0.25 mm × 30 m, 0.25 μm film thickness) from Phenomenex (Madrid, Spain) was used. Carrier gas (He) was kept at a constant flow rate of 1 mL min⁻¹. Temperatures of the transfer line, ionization source and quadrupole were 300, 230 and 150 °C, respectively. MS used an electron impact ionization source (EI, 70 eV) and the analyses were carried out in scan mode at 40–600 m/z. Oven program conditions, as well as temperature and injection mode employed were identical to those used for GC-FID analyses. Retention times of alkanes in GC-MS were between 3.9 and 37.6 min, for C8 and C40, respectively.

For sample treatment an EBA 20 centrifuge (Hettich, Tuttlingen, Germany) and an Heidolph vortex mixer (Schwabach, Germany) were used. An AE Adam analytical balance model SAB 124i (SOLIS) from Adam Equipment Inc. (USA) was also used.

Chemometric analysis of results was carried out by applying PCA and OPLS-DA using SIMCA version 14.1 software (Umetrics, Sartorius Stedim Biotech AS, Umea, Sweden). In the PCA study, the capacity to separate the first two principal components was considered, as shown in the literature.[37] Meanwhile, the selection of the OPLS-DA model was based on its sensitivity and its predictive (Q_2) and reproducibility (R_2) parameters, the appropriate values being $R_2 > 0.7$ and $Q_2 > 0.5$.[38] In addition, the validity of the model was tested by the ANOVA of the cross-validated residuals (CV-ANOVA) and permutation studies.[39] Prior to construction of models, a data alignment was carried out by means of OpenLab CDS Chemstation Edition software (revision C.01.08, Agilent Technologies), which was also used for FID data acquisition. Other software packages used were the MSD Chemstation Data Analysis (version G1701EA, revision E.02.02.1431, Agilent Technologies) for MS data acquisition, Statgraphics Centurion XV (Version 15.1.02), and Sigmaplot 12.5 (Systat, Software Inc., San Jose, CA).

Samples and analytical procedure

The entomological fauna samples were provided by the Unit of Forensic Entomology and Evidence Microscopic Analysis of the University of Murcia (Spain). After collection, the specimens were immediately killed in near-boiling water for a few seconds and then fixed in a 70 : 30 EtOH : water solution, in accordance with the legally enforceable standard imposed by the EAFE.[35] The studied samples had been stored for just over six months before their analysis. All specimens received belonged to the Diptera order, with 5 species from two different families: *Chrysomya albiceps* (Wiedemann, 1819), *Lucilia sericata* (Meigen, 1826), *Calliphora vicina* (Robineau-Desvoidy, 1830) and *Calliphora vomitoria* (Linnaeus, 1758) belong to the Calliphoridae family, while *Synthesiomyia nudiseta* (van der Wulp, 1883) belongs to the Muscidae family.

Specimens of different developmental stages were available for some species, including adults, larvae (third larval instar), and pupae. Two different growth patterns of *C. vomitoria* species were studied: wild adult specimens from Sierra Espuña (SE Spain) and adults and pupae laboratory-reared from commercial asticot. As regards the specimens raised in the laboratory: larvae were fed with pig liver and adults with sugar and water. Pig liver was used as the laying medium. Laboratory incubator temperature was set at 25 °C and relative humidity at 60%. Three specimens of each category were available, resulting in a total of 33 individuals. In this way, a variety of specimens representative of different stages of dipteran development were used in order to test the usefulness of the method under different assumptions. A summary of the available information for each specimen appears in Table 1. Prior to analysis, the samples were dried and weighed. Drying was carried out by keeping samples in a hood at room temperature for 24 h.

Sample treatment consisted of a SLE step by adding 2 mL of ACN to each specimen and application of mild agitation by vortexing at 100 rpm for 5 min. Next, the insect was removed and 75 μL of chloroform was added to the organic extract. This mixture of extractant and dispersant solvents was then used for the DLLME step, being rapidly injected into 10 mL of water previously placed in a conical-bottomed glass tube. The characteristic DLLME turbidity, due to the formation of microdroplets of chloroform dispersed within the aqueous phase, thanks to the dispersing action of ACN, was immediately observed. The ternary mixture was manually shaken for a few seconds and then centrifuged at 3000 rpm for 5 min. The sedimented phase (recovered volume around 25 μL) was collected and 1 μL was automatically injected into the GC-FID and GC-MS systems.

Table 1 Details of specimens

Type	Sex	Stage of development	Family	Subfamily	Species
1[a]	Female	Adults	Calliphoridae	Chrysomyinae	*Ch. albiceps*
2[a]	Male	Adults	Calliphoridae	Chrysomyinae	*Ch. albiceps*
3[a]		Larvae[c]	Calliphoridae	Chrysomyinae	*Ch. albiceps*
4[a]	—	Adults	Calliphoridae	Calliphorinae	*C. vomitoria*
5[a]		Pupae	Calliphoridae	Calliphorinae	*C. vomitoria*
6[b]	—	Adults	Calliphoridae	Calliphorinae	*C. vomitoria*
7[a]	—	Adults	Calliphoridae	Calliphorinae	*C. vicina*
8[a]	—	Adults	Calliphoridae	Luciliinae	*L. sericata*
9[a]	Female	Adults	Muscidae	Azeliinae	*S. nudiseta*
10[a]	Male	Adults	Muscidae	Azeliinae	*S. nudiseta*
11[a]		Pupae	Muscidae	Azeliinae	*S. nudiseta*

[a] Laboratory-reared. [b] Wild. [c] Third larval instar.

Results and discussion

Optimization of GC-FID instrumental conditions

A standard solution containing the alkanes at 10 µg mL^{-1} prepared in chloroform was used to optimize separation and detection conditions. The GC inlet temperature was studied in a range of 280–380 °C, and 350 °C was adopted considering the higher sensitivity provided for most compounds. Different GC oven programmes with diverse combinations of initial (35–50 °C) and final (300–320 °C) temperatures and gradients (8–15 °C min^{-1}) were tested using a mobile phase flow rate of 3 mL min^{-1}. The selected programme consisted of an initial temperature of 35 °C for 1 min, followed by a linear gradient at 10 °C min^{-1} up to 320 °C which was maintained for 10.5 min. Under these conditions, the targeted C8–40 alkanes eluted between 3.4 and 32.5 min.

The influence of hydrogen gas flow-rate for the FID system was studied at four levels (40, 45, 50 and 55 mL min^{-1}), while maintaining a fixed air flow-rate of 450 mL min^{-1}. This ensured that the air : hydrogen ratio remained within the recommended 8 : 1 to 12 : 1 interval for proper operation of the FID detector. The highest signals were obtained with the most hydrogen-enriched flame, so a flow rate of 55 mL min^{-1} was adopted for this gas.

Optimization of the sample procedure

To achieve the goal of increasing method sensitivity by preconcentrating sample extracts, the solvent for the SLE step was selected based on its ability to isolate hydrocarbons from the matrix and its suitability as a dispersant solvent in the subsequent DLLME stage. Although hexane[4,7,9] and the combined use of petroleum ether–dichloromethane[14] have demonstrated their efficiency for SLE, their characteristics prevented them from being considered as DLLME dispersant solvents. Consequently, ACN, MeOH and EtOH were tested for the SLE step, and the results were compared with those obtained using hexane, which is the solvent most commonly used. Hexane and ACN showed no significant differences in the extraction efficiencies, and therefore ACN was selected for further experiments.

The ACN volume was limited by specimen size to ensure complete coverage in the extraction step, so a 2 mL volume was selected. The SLE was assisted by gentle vortexing at 100 rpm, and the extraction time was studied in the range of 2–20 min. No significant differences were observed for times longer than 5 min, so this time was adopted.

The sensitivity achieved by evaporating the SLE extract and reconstituting the dried residue in 0.2 mL of ACN was compared to that obtained by using the SLE extract as the disperser solvent for a subsequent DLLME step with 200 µL of chloroform as the extractant solvent. Promising results were obtained by applying DLLME, so the optimization of the microextraction procedure was addressed.

Chloroform and hexane were evaluated as DLLME extractant solvents. For this, a mixture containing 200 µL of hexane or chloroform and the ACN extract was rapidly injected into 10 mL of water. The recovered volumes of hexane and chloroform were around 50 and 250 µL, respectively. As shown in Fig. 1A, lower signals were obtained for short chain alkanes when using chloroform, meanwhile this solvent provided similar or better results for heavier alkanes. The possibility of reducing the volume of chloroform, thus favouring the preconcentration, as well as the ease of recovery of the drop, features that hexane does not provide, led us to select chloroform as the extractant.

The efficiency of DLLME extraction is highly dependent on the volume ratio of the three phases. As the dispersant volume was set to 2 mL in the previous SLE step to ensure specimen coverage, only the volumes of the aqueous phase and extractant solvent were considered in the optimization. Thus, a multivariate study based on a 2^2+star central compound design (CCD), face-centered with three axial points and spaced, was applied. Three levels for each factor were studied obtaining a design with a total of 11 experiments. Volumes of 5–10 mL for the aqueous phase and 75–150 µL for the extractant solvent were tested. Both factors showed a significant effect on C8–C34 hydrocarbons but not on long chain alkanes (C35–C40). Fig. 1B shows the response surface graph obtained, indicating 10 mL of water and 75 µL of chloroform as optimal volumes for C8–C34 alkanes.

Univariate studies were carried out for studying the effect of NaCl concentration in the aqueous phase on the DLLME efficiency. The experiments were performed using 10 mL of water with NaCl contents ranging from 0 to 10% m/v, 2 mL ACN and 75 µL of chloroform. A signal decrease was observed for all analytes with increasing NaCl percentage, so salt addition was discarded. Other variables such as temperature, extraction time and centrifugation were not considered because distribution equilibrium is instantaneously reached in DLLME. Centrifugation speed was fixed at 3000 rpm for 5 min.

The optimized SLE-DLLME methodology was compared with a simple SLE step with 2 mL of hexane. As expected, considering that similar extraction efficiencies obtained with ACN and hexane, the increase of sensitivity observed when DLLME was included was up to 76-fold, depending on the compound.

Method validation

The SLE-DLLME with GC-FID method was validated by evaluating the linearity range, limits of detection (LOD) and quantification (LOQ) and precision according to international guidelines.[40] Linearity was assessed in the 1–2500 ng mL^{-1} range, depending on the compound, with regression coefficients greater than 0.99 in all cases. Preconcentration factors (PFs) were calculated by comparing the calibration slopes obtained by GC-FID analysis of a standard solution of the hydrocarbons submitted to DLLME under the optimized conditions and in the absence of preconcentration. PF values up to 76 were obtained.

LODs and LOQs were calculated following the most widely accepted criterion, which considers signals of three and ten times greater than the noise (S/N = 3 and S/N = 10), respectively. LODs ranged from 0.26–74 ng mL^{-1}, with C17 and C40 being the most and least sensitive compounds, respectively. Additionally, LODs were calculated taking into account the mass of the specimens, varying between 0.26 and 74 ng mg^{-1} (for pupae specimens) and 0.008–2.31 ng mg^{-1} (for the rest of the specimens).

Repeatability studies were carried out to determine the precision of the method. Quality control experiments were
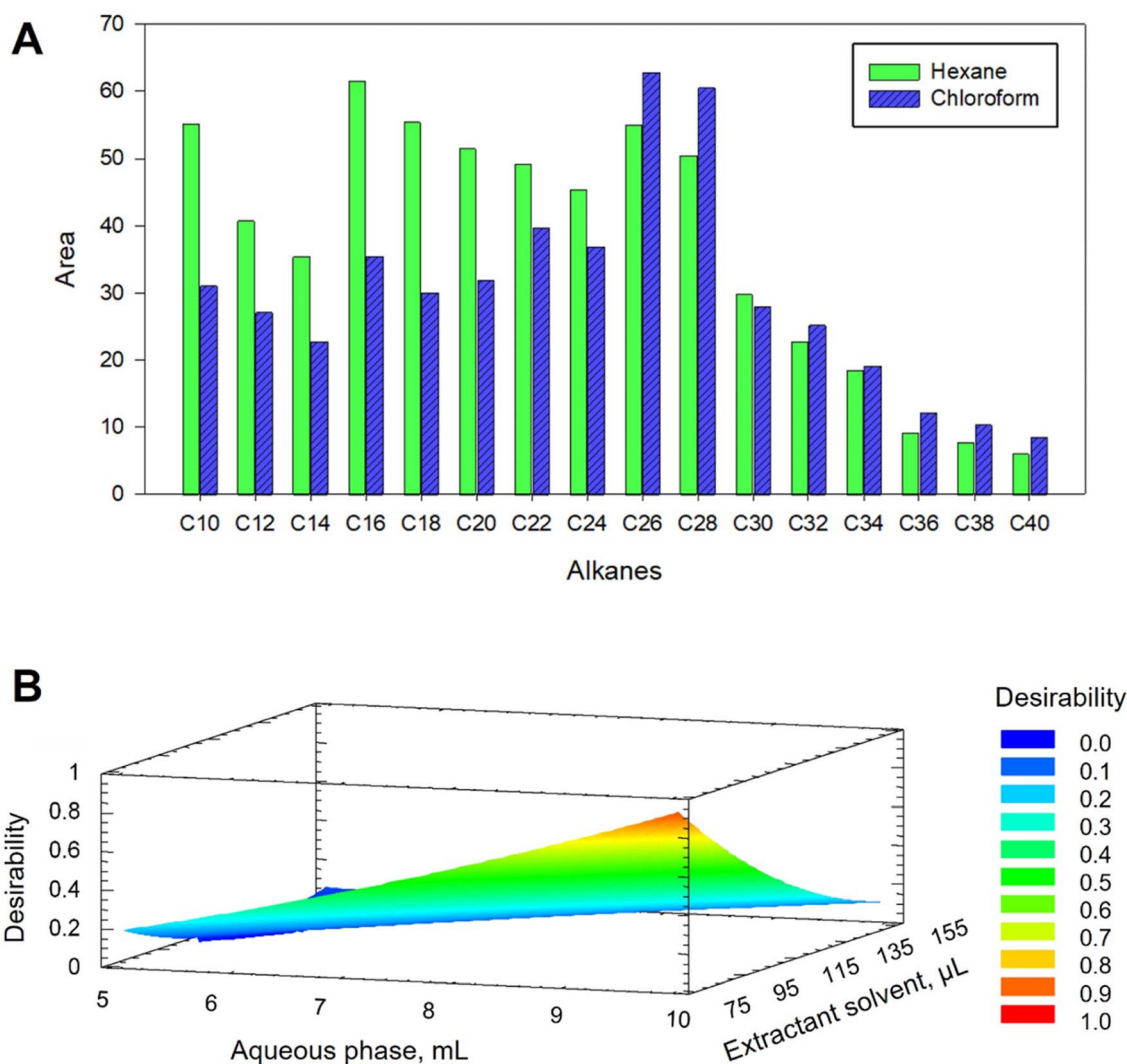
Fig. 1 (A) Effect of the nature of the extractant solvent on DLLME extraction efficiency and (B) response surface obtained for the optimization of aqueous phase and chloroform volumes using a multivariate study based on a 2^2+star central composite design.

established using concentration values of 300, 800 and 1200 ng mL^{-1}, which were within the linear range of all n-alkanes studied. For each concentration, four standard solutions prepared in ACN were subjected to DLLME and the collected drop was injected into the GC-FID system by triplicate ($n = 12$) on the same day. The relative standard deviations were in the 1.1–13.5% range (Table S1†).

Fig. S1† shows a chromatogram obtained by DLLME with GC-FID analysis of a standard solution at 1 μg mL^{-1}. On the other hand, the SLE-DLLME with GC-MS method was validated, and linearity was established through calibration curves obtained at five concentration levels in the 5–2500 ng mL^{-1} range. R_2 values greater than 0.99 were obtained for all compounds. LOD values, obtained using the same criteria described above, were found in the range of 0.33–161 ng mL^{-1}, with C11 and C38 being the highest and lowest sensitive compounds, respectively. LODs for larvae and pupae specimens were in the range 0.33–161 ng mg^{-1}, meanwhile for adult specimens 0.010–5 ng mg^{-1}. Precision studies were carried out by using four ACN standard

solutions of 500 ng mL^{-1}, which were subjected to DLLME and injected in triplicate ($n = 12$) into the GC-MS system. The RSD values were below 12% in all cases.

Comparison of the analytical characteristics obtained by means of the GC-FID and GC-MS equipment used showed that the former had slightly higher sensitivity. Considering the ease of management and interpretation of FID data compared to MS, GC-FID was selected for sample quantification, relegating GC-MS when it was necessary to confirm the identity of SCHCs.

Analysis of Diptera specimens

The proposed SLE-DLLME with GC-FID method was applied for the quantification of SCHCs in the 33 Diptera samples. Table 2 shows the mean SCHC concentrations for each specimen, categorized by species and maturity developmental stage. The identity of the detected SCHCs in the samples by GC-FID was corroborated by GC-MS analysis through comparison of the retention times, MS spectra and fragmentation patterns by monitorization of the characteristic fragment ions for the

Table 2 Mean concentration ± SD (ng mg^{-1}) for detected alkanes in mature and non-mature specimens on the different species by SLE-DLLME with the GC-FID method

| Analyte | S. nudiseta | | C. vicina | L. sericata | Ch. albiceps | | C. vomitoria | | |
| | | | | | | | Laboratory | | Wild |
	Adult[a]	Pupae[b]	Adult[b]	Adult[b]	Adult[a]	Larvae[b]	Adult[b]	Pupae[b]	Adult[b]
C9	4 ± 1	ND	ND	ND	ND	ND	2.5 ± 0.4	12 ± 3	2 ± 1
C11	45 ± 10	ND	4 ± 2	7 ± 5	ND	ND	25 ± 3	125 ± 32	14 ± 9
C12	9 ± 4	7 ± 4	0.8 ± 0.3	1.0 ± 0.3	0.6 ± 0.2	0.10 ± 0.02	6 ± 1	18 ± 8	3 ± 2
C13	1 ± 1	ND	ND	ND	ND	ND	ND	ND	0.5 ± 0.2
C14	4 ± 3	6 ± 4	0.3 ± 0.1	0.4 ± 0.1	0.2 ± 0.1	0.05 ± 0.01	2.6 ± 0.4	5 ± 3	1 ± 1
C15	10 ± 5	19 ± 8	ND	ND	0.3 ± 0.3	0.6 ± 0.1	10 ± 1	27 ± 7	3 ± 2
C18	ND	ND	ND	ND	ND	0.04 ± 0.01	ND	ND	ND
C20	ND	13 ± 8	ND	ND	ND	1.2 ± 0.2	5 ± 1	ND	2 ± 1
C21	ND	ND	ND	ND	ND	0.12 ± 0.03	ND	ND	ND
C22	2 ± 2	10 ± 2	ND	ND	1 ± 1	0.6 ± 0.2	2.4 ± 0.4	ND	2 ± 1
C23	ND	ND	ND	ND	ND	0.3 ± 0.2	ND	ND	0.2 ± 0.1
C24	4 ± 5	5 ± 3	ND	ND	ND	1 ± 1	4 ± 2	ND	2 ± 1
C25	ND	ND	ND	ND	1 ± 1	2 ± 1	ND	ND	ND
C26	1 ± 1	ND	ND	ND	0.1 ± 0.1	1 ± 1	ND	ND	ND
C27	40 ± 44	20 ± 8	ND	ND	0.1 ± 0.2	1 ± 1	ND	ND	3 ± 4
C28	6 ± 9	ND	ND	ND	ND	4 ± 3	ND	ND	ND
C29	51 ± 45	52 ± 11	ND	ND	0.3 ± 0.3	2 ± 1	ND	13 ± 3	0.7 ± 0.1
C30	5 ± 7	ND	ND	ND	1 ± 1	3 ± 2	ND	ND	ND
C31	16 ± 9	24 ± 3	ND	ND	0.2 ± 0.4	2 ± 1	ND	ND	ND
C32	ND	ND	ND	ND	0.1 ± 0.3	2 ± 1	ND	ND	ND
C33	ND	ND	ND	ND	0.4 ± 0.9	3 ± 2	ND	ND	ND
C34	ND	ND	ND	ND	4 ± 4	8 ± 4	ND	ND	ND
C35	ND	ND	ND	ND	0.4 ± 1.0	6 ± 4	ND	ND	ND
C36	ND	ND	ND	ND	1 ± 2	4 ± 6	ND	ND	ND

Experimental replicates: [a]$n = 6$, [b]$n = 3$; ND means not detected.

alkanes such as m/z 57, as well as ions 43 and 71, between samples and standards.

A total of twenty-four SCHCs were quantified, in the C9–C36 range. Only two alkanes (C12 and C14) were found in all specimens, regardless of their nature or developmental stage. ANOVA tests were applied to compare the SCHC concentrations within the groups of mature and non-mature specimens shown in Table 2. Statistical analyses revealed significantly higher concentrations of C29 in *S. nudiseta* pupae, of C34 in *Ch. albiceps* adult specimens and of C11 in *C. vomitoria*, *C. vicina* and *L. sericata*, whatever the stage of development. These results could be attributed to the evolutionary proximity of *C. vicina*, *C. vomitoria* and *L. sericata* species.[41]

Ch. albiceps had the highest number of alkanes with 21 and 15 detected in their larvae and adult forms, respectively. The third larval instar showed an increase in detected alkanes (21 *vs.* 17 hydrocarbons) compared to the work of Alotaibi *et al.*[9] Furthermore, this species showed the widest range of hydrocarbons, from C12 to C36, followed by *S. nudiseta* (C9–C31) and *C. vomitoria* (C9–C29).

On the other hand, previous studies have detected a greater number of alkanes in *L. sericata* specimens (ranging from C12 to C31)[13,16] and *C. vicina* specimens (between C21 and C31),[14,16] detecting up to 18 CHCs compared to the three found in this study (C11, C12 and C14). In the case of *C. vomitoria* laboratory-reared specimens, both mature and immature, other studies have found a content of SCHCs rich in alkanes C21–C31,[14,16] while in this study the lightest SCHCs were predominant. It should be noted that the studies carried out by the authors mentioned above were based on fresh or cold-preserved specimens. In contrast, the specimens used in this work were preserved in an EtOH : water mixture.

Discrimination of specimens by species

Non-supervised PCA and supervised OPLS-DA were applied to classify specimens based on their species using SCHC concentrations in each sample. The classification was performed using a data matrix with 33 samples in rows and 24 SCHCs in columns. A data distribution study was carried out using the unit variance (UV) scale. As a normal data distribution was not observed, the logarithmic scale was applied.

The non-supervised PCA technique allows the study of the principal structure of the data ignoring any class information, as shown in the literature.[38] When PCA was applied to the data, the first two principal components were able to explain 65.7% of the variability of the set, with 44% attributed to the first component and 21.7% to the second component. As can be seen in Fig. 2, two principal sets are clearly differentiated, one formed by the species *C. vomitoria* and *S. nudiseta* and the other formed by the species *L. sericata*, *C. vicina* and *Ch. albiceps*.
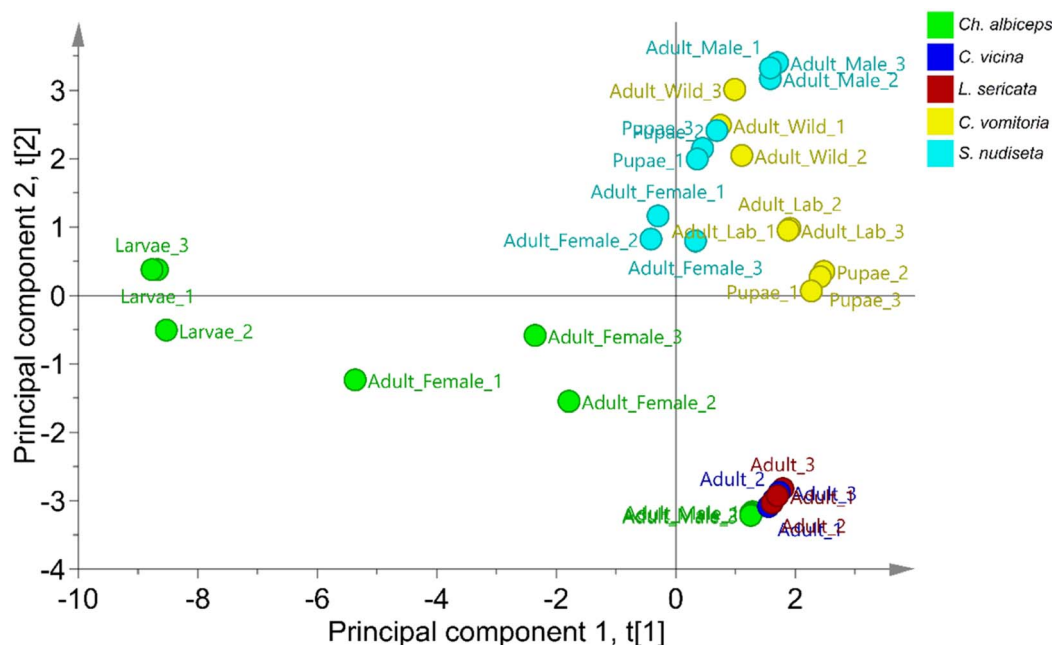
Fig. 2 Score scatter plot of PCA for the differentiation of Diptera order species.

Previous studies have observed chemical similarities in the cuticle of the latter three species.[2] However, separations between species can be also observed within the sets. Thus, the species *S. nudiseta* and *C. vomitoria* species are separated with a slight overlap between the two groups. Likewise, the species *Ch. albiceps* can be differentiated from the species *C. vicina* and *L. sericata* which form an indistinguishable group. Thus, it is possible to build a no overly forced OPLS-DA model for species discrimination.

An OPLS-DA model was built to accurately classify specimens of Diptera order. This model can be used to identify the most significant SCHCs in the species differentiation. The optimal model consists of: UV scale with logarithmic transformation; $3 + 2 + 1$ components, $R_2X(\text{cum}) = 0.866$, $R_2Y(\text{cum}) = 0.664$, and $Q_2(\text{cum}) = 0.522$. The validity of the model was confirmed by performing 100 random permutations (Fig. S2†) and confirmed by the sub-zero crossing of the regression line Q_2 with the vertical axis. Furthermore, a CV-ANOVA was performed to assess the reliability of the model and a p-value < 0.05 was obtained, showing the statistical significance of the model at a 95% confidence level. Additionally, the cross-validation study demonstrated a 90.91% classification rate (CR) for the training set.

The results of the OPLS-DA model are summarised in the score and loading scatter plots (Fig. 3), as well as in the Variable Importance for the Projection (VIP) score plot (Fig. S3†). As shown in Fig. 3A, the first two principal components allow a clear differentiation of the species *Ch. albiceps*, *C. vomitoria* and *S. nudiseta* species, whereas the species *C. vicina* and *L. sericata* are indistinguishable. The loading scatter plot (Fig. 3B) also shows the relationship between species and SCHC content for the principal components. The differentiation of *C.*

vomitoria is mainly determined by the presence of C9, C11, C12 and C14 compounds in the insect cuticle. The presence of hydrocarbons of short and long chains, especially C27, C29 and C31, as well as the C15, determines the differentiation of the species *S. nudiseta*. The species *Ch. albiceps* is associated with longer chain compounds such as C25, C32, C33, C34 and C35. Most of the compounds described above are shown in Fig. S3† with VIP values higher than 1, indicating their importance in the classification.

To extract as much information as possible from the sample set, a hierarchical clustering analysis was performed using single-linkage clustering and size sorting (Fig. 4). The dendrogram shows the two main groups previously mentioned, along with subdivisions within the *Ch. albiceps*, *S. nudiseta*, and *C. vomitoria* species, which can be attributed to either the insect's developmental stage or the specimen's growth pattern. Sex differentiation was also provided by the dendrogram, but no discussion about this topic was carried out, because it is not interesting for forensic applications. Thus, ANOVA tests were applied to compare the SCHC concentrations between mature and immature specimens of the same species.

Statistical analyses according to *Ch. albiceps* development stage revealed significant differences (p-values < 0.05) in several SCHCs. Adult specimens had a significantly higher concentration of short chain SCHCs C12 and C14, while immatures specimens had higher levels of long chain SCHCs C26, C31, C32, C33 and C35. Likewise, compounds C18, C20, C21, C23 and C28 were identified as SCHCs characteristic of immature specimens, as they were only found in larvae specimens.

The samples of *S. nudiseta*, a species that to our knowledge has not been previously studied, also showed differences according to development stages. Significant differences (p-
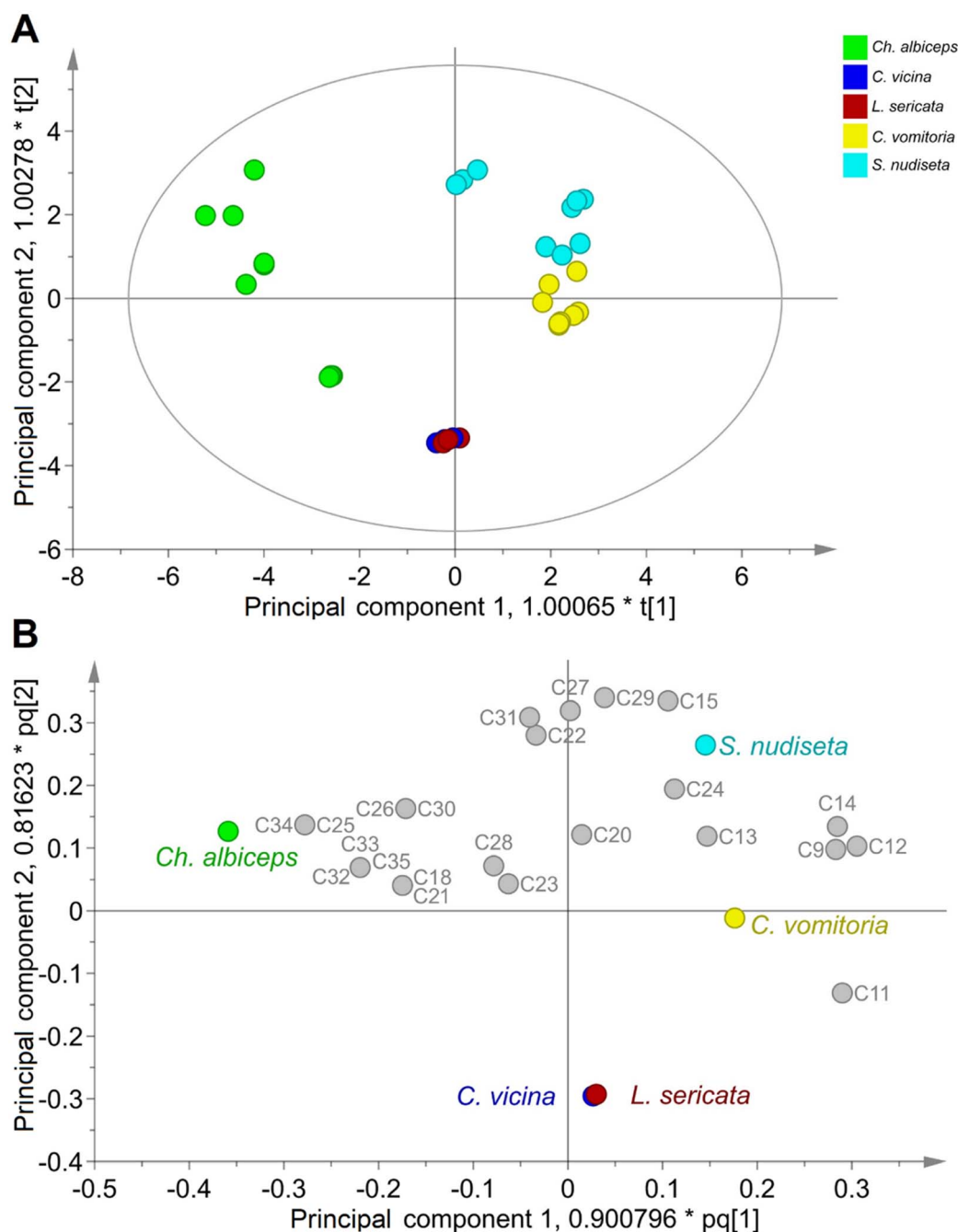
Fig. 3 Score (A) and loading (B) scatter plots of OPLS-DA using logarithmic UV scale for the differentiation of Diptera order species. The first two principal components cover 27.5% and 15.1% of the variance. The plot A shows a tolerance ellipse at 95% based on Hotelling's T^2.

values < 0.05) were observed for C9 and C11 which were present only in mature specimens, for C20 which was specific to immature specimens, and for C22 which was found in a significantly higher concentration in pupae specimens.

In *C. vomitoria* specimens, the significant presence of C20, C22 and C24 showed the maturity of the specimens as they were absent in the pupae, unlike C9, C11, C12, C14, C15 and C29 that have been found in a significantly higher concentration in immature specimens. Another important segregation observed in *C. vomitoria* was the growth pattern of the specimens. Significant

differences were found in the cuticular contents of C13, C14, C15 and C23 (p-values < 0.05). Thus, C13 and C23 appear as characteristic compounds in the cuticle of wild specimens, whereas C14 and C15 are present in higher concentrations in laboratory-reared specimens (Table 2). The differences in cuticular content between specimens of the same species with different growth patterns have been reported by Moore *et al.*[15] as empty puparia from different geographical regions were differentiated.

Very interesting results were obtained for intra- and interspecies differentiation by focusing only on linear alkanes, in
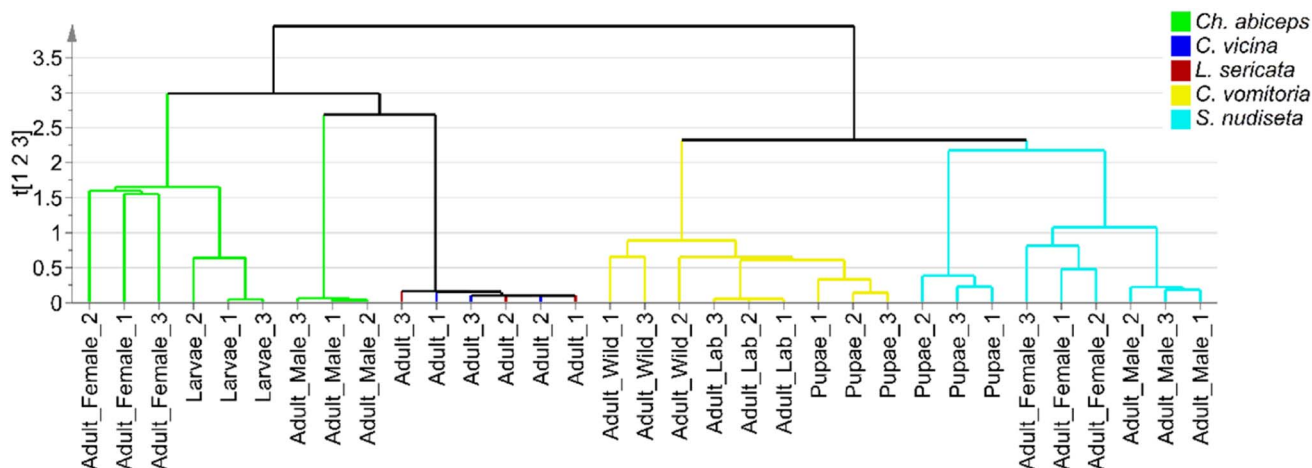
Fig. 4 Dendrogram based on single-linkage clustering for differentiation of species of Diptera order.

line with previous work[12,13,23] and in contrast to other studies considering other groups of compounds such as alkenes.[7,15,17,21,22] The statistical treatment of the SCHC levels in the Diptera order specimens allowed the differentiation of several species, as well as the observation of segregation groups according to their development stage or growth pattern.

Conclusions

The developed sample preparation procedure based on SLE-DLLME has demonstrated a high efficiency for the isolation of SCHCs from insects preserved in EtOH : water solution, following the standard imposed by the EAFE. Thus, this method can be applied to samples that have been preserved for a long time, and is of great use in forensic investigations where ratification is required for identification, counter-expertise or even when the specialist has not collected the samples and needs to analyse stored evidence. The high pre-concentration factor achieved with this miniaturised methodology has allowed the range of found SCHCs to be extended, by analysing only one specimen, compared to previous studies. The analysis of the obtained extracts by GC-FID provided a very sensitive quantification of SCHCs at trace levels, while GC-MS analysis is proposed as an identification tool, thus ensuring the reliability of the proposed method. The high sensitivity provided by the presented method will allow the identification of insects even when part of them is collected instead of the entire specimen. So, when traditional identification methods cannot be applied, the proposed method appears as an interesting alternative.

The concentration of SCHCs found in each sample allowed the performance of non-supervised and supervised chemometric methods, such as PCA and OPLS-DA, to build a model for distinguishing insects according to the species. An OPLS-DA model based only on linear alkane concentrations is proposed to discriminate specimens of Diptera order according to their species, with a CR of 90.91%. A non-supervised hierarchical clustering dendrogram allowed the observation of segregations within each species according to the insect development stage

and growth pattern, based on their SCHC concentration levels. The chemometric models developed may become excellent tools in the field of forensic entomology, given the good results in the classification of species that can help to establish the mPMI.

Author contributions

Lixy Olinda León-Morán: investigation, formal analysis, validation, software, writing – original draft preparation. Marta Pastor-Belda: conceptualization, formal analysis, investigation, resources, methodology, data curation, writing – original draft preparation. Natalia Arroyo-Manzanares: visualization, methodology, software, data curation. Natalia Campillo: conceptualization, investigation, visualization, resources, supervision, writing – review & editing. María Dolores García: resources, writing – review & editing. María Isabel Arnaldos: resources, writing – review & editing. Pilar Viñas: resources, writing – reviewing and editing, project administration, funding acquisition.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 H. Holze, L. Schrader and J. Buellesbach, *Heredity*, 2020, **126**, 219–234.
2 M. I. Arnaldos and M. D. García, *Insects*, 2021, **12**, 429.
3 M. V. Braga, P. M. Mendonça, R. R. Barbosa, G. J. Blomquist, S. P. C. Novo, J. d. M. F. Dutra, S. M. d. Souza and M. M. d. C. Queiroz, *J. Nat. Hist.*, 2015, **50**, 1381–1388.
4 N. J. Butterworth, P. G. Byrne, P. A. Keller and J. F. Wallman, *J. Chem. Ecol.*, 2018, **44**, 248–256.

5 F. Claudio-Piedras, B. Recio-Tótoro, J. Cime-Castillo, R. Condé, M. Maffei and H. Lanz-Mendoza, *Sci. Rep.*, 2021, **11**, 11258.

6 H. Moore, J. B. Butcher, C. D. Adam, C. R. Day and F. P. Drijfhout, *Forensic Sci. Int.*, 2016, **268**, 81–91.

7 A. Sharma, F. P. Drijfhout, J. K. Tomberlin and M. Bala, *J. Forensic Sci.*, 2021, **66**, 236–244.

8 M. Rasoolian, J. Sadrai and M. R. Nikbakhtzadeh, *Anim. Cells Syst.*, 2008, **12**, 165–170.

9 F. Alotaibi, M. Alkuriji, S. Alreshaidan, R. Alajmi, D. M. Metwally, B. Almutairi, M. Alorf, R. Haddadi and A. Ahmed, *J. Med. Entomol.*, 2020, **58**, 1048–1055.

10 S. Beyramysoltan, J. E. Giffen, J. Y. Rosati and R. A. Musah, *Anal. Chem.*, 2018, **90**, 9206–9217.

11 H. Xu, G.-Y. Ye, Y. Xu, C. Hu and G.-H. Zhu, *Forensic Sci. Int.*, 2014, **242**, 236–241.

12 S. Sharif, C. Wunder, M. K. Khan, A. Qamar and J. Amendt, *Forensic Sci. Int.*, 2023, **349**, 111748.

13 M. Gołębiowski, M. Paszkiewicz, A. Grubba, D. Gąsiewska, M. I. Boguś, E. Włóka, W. Wieloch and P. Stepnowski, *Bull. Entomol. Res.*, 2012, **102**, 453–460.

14 O. Roux, C. Gers and L. Legal, *Med. Vet. Entomol.*, 2008, **22**, 309–317.

15 H. Moore, L. Lutz, V. Bernhardt, F. P. Drijfhout, R. B. Cody and J. Amendt, *Int. J. Leg. Med.*, 2022, **136**, 1791–1800.

16 H. Moore, J. B. Butcher, C. R. Day and F. P. Drijfhout, *Forensic Sci. Int.*, 2017, **280**, 233–244.

17 H. Moore, C. D. Adam and F. P. Drijfhout, *Forensic Sci. Int.*, 2014, **240**, 48–53.

18 J. L. Pechal, H. Moore, F. Drijfhout and M. E. Benbow, *Forensic Sci. Int.*, 2014, **245**, 65–71.

19 M. I. Arnaldos, E. L. Gallego and M.-D. García, *Ciencia Forense. Revista Aragonesa de Medicina Legal*, 2015, vol. 12, pp. 153–174.

20 T. Ivorra, A. Martínez-Sánchez and S. Rojo, *Int. J. Leg. Med.*, 2019, **133**, 651–660.

21 H. Moore, C. D. Adam and F. P. Drijfhout, *J. Forensic Sci.*, 2013, **58**, 404–412.

22 M. V. Braga, Z. T. Pinto, M. M. d. C. Queiroz and G. J. Blomquist, *Forensic Sci. Int.*, 2016, **259**, e37–e47.

23 J. H. d. S. Brito, W. F. Antonialli-Junior, T. D. S. Montagna, A. Mendonça, D. Sguarizi-Antonio, Y. R. Súarez, S. M. Lima, L. H. d. C. Andrade and C. A. L. Cardoso, *Sociobiology*, 2017, **64**, 327–333.

24 M. C. Paula, W. F. Antonialli-Junior, A. Mendonça, K. B. Michelutti, A. D. M. M. Eulalio, C. A. L. Cardoso, T. d. Lima and C. J. V. Zuben, *J. Med. Entomol.*, 2017, **54**, 14–23.

25 I. Alnajim, M. Agarwal, T. Liu, B. Li, X. Du and Y. Ren, *Molecules*, 2020, **25**, 1565.

26 M. J. Ferreira-Caliman, A. C. R. R. Andrade-Silva, M. C. Guidetti-Campos, I. C. C. Turatti, F. S. d. Nascimento and N. P. Lopes, *Anal. Methods*, 2014, **6**, 8823–8828.

27 C. D. Pasquale, S. Guarino, E. Peri, G. Alonzo and S. Colazza, *Anal. Bioanal. Chem.*, 2007, **389**, 1259–1265.

28 M. Cruz-Vera, R. Lucena, S. Cárdenas and M. Valcárcel, *Anal. Methods*, 2011, **3**, 1719–1728.

29 K. S. Hasheminasab, A. R. Fakhari and M. Baghdadi, *Clean: Soil, Air, Water*, 2014, **42**, 1106–1114.

30 B. Klein, F. R. Thewes, A. R. d. Oliveira, A. Brackmann, J. S. Barin, A. J. Cichoski and R. Wagner, *Food Res. Int.*, 2019, **116**, 611–619.

31 Y. Wen, J. Nie, Z. G. Li, X. Y. Xu, D. Wei and M. R. Lee, *Anal. Methods*, 2014, **6**, 3345–3352.

32 H. Sereshti, A. Ghiasi, M. Naderloo, M. Taghizadehb and S. D. A. Astaneh, *Anal. Methods*, 2014, **6**, 6695–6701.

33 U. Savković, I. Vučković and B. Stojković, *J. Stored Prod. Res.*, 2012, **50**, 66–72.

34 G.-H. Zhu, X.-J. Yu, L.-X. Xie, H. Luo, D. Wang, J.-Y. Lv and X.-H. Xu, *PLoS One*, 2013, **8**, e73043.

35 J. Amendt, C. P. Campobasso, E. Gaudry, C. Reiter, H. N. LeBlanc and M. J. R. Hall, *Int. J. Leg. Med.*, 2007, **121**, 90–104.

36 D. A. D. S. Cunha, R. S. T. Menezes, C. A. L. Cardoso and W. F. A. Junior, *Environ. Entomol.*, 2021, **50**, 580–588.

37 M. J. Li, Z. M. Zhang, F. Fan, P. Ma, Y. Wang and H. M. Lu, *Anal. Methods*, 2019, **11**, 2895.

38 M. García-Nicolás, N. Arroyo-Manzanares, J. d. D. Hernández, I. Guillén, P. Vizcaíno, M. Sánchez-Rubio, I. López-García, M. Hernández-Córdoba and P. Viñas, *Anal. Chim. Acta*, 2020, **1128**, 52–61.

39 N. Arroyo-Manzanares, M. García-Nicolás, F. Abellán-Alfocea, L. Prieto-Baeza, N. Campillo, B. d. V. Oliver, J. Zarauz-García, L. Sáenz and P. Viñas, *Anal. Bioanal. Chem.*, 2023, **415**, 3571.

40 European Commission, *Off. J. Eur. Communities*, 2002, 8–36.

41 L. M. McDonagh and J. R. Stevens, *Parasitology*, 2011, **138**, 1760–1777.