DIFFERENTIATION OF BEE POLLEN SAMPLES ACCORDING TO THEIR INTACT-GLUCOSINOLATE CONTENT USING CANONICAL DISCRIMINANT ANALYSIS

Ana M. Ares[1*], Montserrat Redondo[1], Jesús Tapia[1,2], Amelia V. González-Porto[3], Mariano Higes[3], Raquel Martín-Hernández[3,4], José Bernal[1,*]

[1]I. U. CINQUIMA, Analytical Chemistry Group, Faculty of Sciences, University of Valladolid, Valladolid, Spain.

[2]Department of Statistics and Operations Research, Faculty of Sciences, University of Valladolid, Valladolid, Spain.

[3]Instituto Regional de Investigación y Desarrollo Agroalimentario y Forestal de Castilla La Mancha (IRIAF), Centro de Investigación Apícola y Agroambiental (CIAPA) de Marchamalo, Guadalajara, Spain.

[4]Instituto de Recursos Humanos para la Ciencia y la Tecnología (INCRECYT-FEDER), Fundación Parque Científico y Tecnológico de Castilla – La Mancha, Guadalajara, Spain.

***Corresponding authors:** Tel# 34-983186347; ana.maria.ares@uva.es; jose.bernal@uva.es; web: tesea.uva.es

Abstract

A study is presented of the real possibilities of glucosinolate content and chemometrics (canonical discriminant analysis) to differentiate bee pollen samples from four different apiaries (Fuentelahiguera, Monte, Pistacho, Tío Natalio) located in the same geographical area. Fifteen intact-glucosinolates were quantified by means of ultra-performance liquid chromatography coupled to a quadrupole time-of-flight mass detector in forty-nine bee pollen samples. Glucosinolate residues were detected in most of the samples, and these differed in number and concentration. It was possible to directly differentiate one of the apiaries (Fuentelahiguera) from the other three (Monte, Pistacho y Tío Natalio) by comparing glucosinolate content. These three apiaries were differentiated by means of the first two canonical variables obtained from a canonical discriminant analysis. Following this analysis, more than 88% of the samples could be assigned correctly to the Pistacho and Monte apiaries, and 100% to the Tío Natalio apiary.

Keywords: Authenticity; Bee pollen; Canonical discriminant analysis; Glucosinolates; UPLC-Q-TOF/MS.

1. Introduction

Bee pollen is a natural product that has been used in the human diet for many centuries due to its nutritional and therapeutic/pharmacological properties (antioxidant, antimicrobial or anti-inflammatory; Anjos et al., 2019), and it is currently one of the most widely consumed food supplements (Ares, Valverde, Bernal, Nozal, & Bernal, 2018). The presence of bioactive compounds (proteins, lipids, amino acids, vitamins, minerals and phenolic compounds) in bee pollen depends on several factors; not only the plant species from which it originates, but also the type of soil, climatic conditions, agriculture, the harvesting season, apicultural practices and even the treatment of the product during storage or processing prior to its commercialization (Negrão & Orsi, 2018). It can be seen, then, that the composition of bee pollen is highly variable, making the labelling of this product difficult (Gonçalves, Estevinho, Pereira, Sousa, & Anjos, 2018), especially in terms of specifying its origin (botanical and/or geographical) and composition. Indeed, determining origin is a particularly relevant issue for the beekeeping industry, especially if we consider that consumers' preference is influenced by this parameter (Truchado, Tourn, Gallez, Moreno, Ferreres, & Tomás Barberán, 2010), and that this could be employed to detect potential bee pollen fraud (Chica & Campoy, 2012; Zhou et al., 2015).

Over the past years, several studies have been published in which different procedures have been employed to distinguish the origin of bee pollen due to its composition. Most of these works (Gonçalves et al., 2018; Kaškonienė, Ruočkuvienė, Kaškonas, Akuneca, & Maruška, 2015; Sattler et al., 2015), attempted to correlate the botanical origin of the pollen with some of its most relevant parameters (moisture, pH, lipids, proteins, sugars, vitamins, volatile and phenolic compounds). Meanwhile, Zhou et al. (2015) demonstrated the usefulness of flavonoid glycosides as floral origin markers to discriminate unifloral bee pollen. Considering the latter study, in which a single family of compounds was successfully employed to discriminate bee pollen, we decided to check the suitability as origin markers of a different family of bioactive compounds present in this substance,

namely, glucosinolates (GSLs); these had previously been proposed as botanical biomarkers in honeys (Truchado et al., 2010), but to our knowledge this is the first time that they have been employed to perform this task in bee pollen. GSLs are secondary plant metabolites in the order of the Brassicales, and their quality and number differ among plant species. In the last few years they have attracted researchers' attention due to the potential health benefits (chemopreventive and antimicrobial) associated with their breakdown products, especially isothiocyanates (Rossetto, Shiga, Vianello, & Lima, 2013).

However, GSLs have been the object of scant investigation in bee pollen, as we know of only three previously published related studies (Ares, Nozal & Bernal, 2015; Bernal, González, Valverde, Toribio, & Ares, 2019; Dungey et al., 1988). In the light of the above-mentioned publications, we decided to employ the procedure recently developed and validated by our research group (Bernal et al., 2019; see subsections 2.3, 2.4) to analyze forty-nine samples were obtained from four different apiaries in a province in Spain (Guadalajara) famous for beekeeping. Following this sample treatment, GSLs maintained the sulfate group in their structure, and for this reason they are specifically known as intact-GSLs (Glauser, Schweizer, Turlings & Reymond, 2012). Therefore, the main goal of this study was to verify whether, by means of chemometrics, and more specifically, canonical discriminant analysis (CDA), intact-GSL content permits the determining of the origin of bee pollen samples. To our knowledge, this is the first attempt to pinpoint the original apiary from which bee pollen samples have been taken by examining their intact-GSL composition.

2. Materials and methods

2.1. Reagents and materials

GSLs standards (Det. Purity > 95%; see Table 1) were obtained from Phytoplan Diehm & und Neuberger GmbH (Heidelberg, Germany). All reagents were of LC grade (Bernal et al., 2019). Syringe filters (17mm, Nylon 0.45 μm) were purchased from Nalgene (Rochester, NY, USA), and

ultrapure water was obtained from Millipore Milli-RO plus and Milli-Q systems (Bedford, MA, USA). An Eppendorf Centrifuge 5810R (Hamburg, Germany), an R-210/215 rotary evaporator 109 (Buchi, Flawil, Switzerland), Bond Elut NH_2 (3mL with 500 mg of sorbent) SPE cartridges from Agilent Technologies (Palo Alto, CA, USA), a 10-port Visiprep vacuum manifold (Supelco, St. Louis, MO, USA), a Moulinette chopper device from Moulinex (Paris, France), and a Vibromatic mechanical shaker and a drying oven both from J.P. Selecta S.A. (Barcelona, Spain) were used for the extractions.

2.2. Standards

Individual standard stock (≈ 100 mg/L) and working solutions were prepared with ultrapure water. Bee pollen samples (1 g), which had shown to contain no GSLs (Bernal et al., 2019), were spiked before sample treatment with different amounts of the GSL, in order to prepare the matrix-matched standard calibration curves (limit of quantification (LOQ, see Table 1), 100, 200, 500, 750, 1000, 1500 μg/kg). These samples were obtained from local markets (Valladolid, Spain).

2.3. Sample procurement and treatment

2.3.1. Samples

An analysis was made of a total of forty-nine corbicular bee pollen samples, collected in April and May of 2018, from four experimental apiaries of the *Centro de Investigación Apícola y Agroambiental* (CIAPA; Marchamalo, Guadalajara, Spain), with homogeneous *Apis mellifera iberiensis* colonies. Three of the apiaries (*Pistacho*, PI; *Monte*, MO; *Tío Natalio*, TN) came from the CIAPA estate in Marchamalo, and the other (*Fuentelahiguera*, FH) was 10.8 km away in the municipality of Fuentelahiguera de Albatages; all were in the province of Guadalajara (see Supplementary Material, Figure 1S). The PI apiary was located close to a plot where plant species sown in spring and autumn, such as rapeseed (*Brassica napus* L.), were cultivated. Samples of bee pollen were collected by means of pollen-traps set at the entrance to the hive. Every 15 days, the

pollen-trap grid was closed for a period of 24 hours in all the hives involved in the study. Pollen stored in the collection drawer during this period was collected, immediately sealed, identified (date of collection, apiary and colony) and taken to the laboratory, where it was frozen until palynological analysis. It should be mentioned that, although it would be possible for a few bees to change their own hive for another, a phenomenon called drifting, this occurs only once per individual bee and does not represent general behavior. Consequently, a small number of pollen balls collected in the whole sampling, which might correspond to those bees, would be mixed with the hundreds of balls collected from the other bees in this hive, minimizing a potential cross-interference of the samples.

2.3.2. Palynological analysis

First, each bee pollen sample was separated by color (Hidalgo & Bootello, 1990), in accordance with the Pantone 747 XR Universal Color Guide. For this study, the pollen balls of the predominant color were selected for palynological analysis to determine their botanical origin. Twenty balls (from each predominant color per sample) were punched and a small amount was added to a drop (≈ 10 µL) of water; this was then placed on a glycerin jelly slide and examined under an optical microscope (× 250 magnification; Wild Leitz GmbH, Wetzlar, Germany). The species of plant was identified by means of a photographic atlas (Faegri & Iversen, 1989) together with the reference collection of pollen slides from CIAPA (Cepero et al., 2014). As can be seen in Table 2, taxa are grouped into pollen types (t.) because different types of plants share common pollen shapes and characteristics, and with optical microscopy further differentiation is difficult. This occurs in the case of Brassica t., which includes different species of the same genus and some species of other genera (albeit of the same family), both wild and cultivated rapeseed, wild radish (Raphanus t.) and wild rocket (Diplotaxis t.).

2.3.3. Sample treatment

Bee pollen samples were mixed, ground and pooled for optimum sample homogeneity. Next, the pollen was dried until the mass stabilized (humidity was between 9% and 12%), and subsequently it

was stored in the dark at −20ºC until analysis (Ares et al., 2015). Figure 2S (see Supplementary Material) outlines the steps of the procedure used during the present study.

2.4. UPLC-Q-TOF/MS system

An Acquity™ UPLC system (ACQUITY, Waters, Milford, MA, USA) and a Q-TOF/MS spectrometer (maXis impact, Bruker Daltonik, Bremen, Germany) were coupled through an electrospray (ESI) interface, which was operated in the negative ionization mode. It must be specified that the UPLC-Q-TOF/MS method was optimized and validated in a previous study (Bernal et al., 2019). Chromatogram of a bee pollen sample (TN-02) obtained using the selected UPLC-Q-TOF/MS conditions is shown in Figure 3S (see Supplementary Material).

2.4.1. UPLC conditions

A Luna® Omega C_{18} (50 × 2.1 mm, 1.6 µm) and a Luna® Omega C_{18} guard column (Phenomenex; Torrance, CA, USA) were employed for all UPLC analyses. Mobile phase composition was described in Table 1S (see Supplementary Material); meanwhile, injection volume and column temperature were set at 5 µL and 30ºC, respectively.

2.4.2. Q-TOF/MS conditions

The values of the most relevant parameters were: capillary voltage, -2250 V; drying gas flow, 12 L/min; drying gas temperature, 220 °C; nebulizer pressure, 2 bar. GSLs were quantified by generating extracted ion chromatograms with the precursor ions; meanwhile the most relevant fragments for each precursor ion were also used to confirm the presence of each intact-GSL. These MS/MS experiments were carried out by using an isolation width of 10 m/z and a collision energy of 40 eV. The ions monitored for each compound are summarized in Table 1.

2.5. Canonical discriminant analysis

Statistical data analysis was performed using the CANDISC procedure of SAS software (version 9.4; SAS Institute Inc., Cary, NC, USA). Hastie, Tibshirani and Friedman (2008) refer to CDA as

the analogue for grouped data of principal component analysis for ungrouped data. When there are two or more groups of observations and the same information is measured with quantitative variables, CDA determines the linear combination of the variables that provide maximal separation between the groups. The combination of variables with the greatest possible multiple correlation with the groups is called the first canonical variable. The second canonical variable is obtained by finding the linear combination uncorrelated with the first canonical variable that has the greatest possible multiple correlation with the groups. The CDA process establishes the same number of canonical variables as the number of original variables. The proportion of variance explained by the canonical variables determines how many canonical variables are used in the discriminant analysis to classify each sample in one of the groups. The data base used in the present study comprised the response of each sample to the qualitative variable (apiary of origin) and the three analyses of each individual sample for each intact-GSL (quantitative variables).

3. Results and discussion

3.1. Determination of glucosinolate content

Several bee pollen samples (n= 49; see Subsection 2.3), from four different apiaries (FH, MO, PI and TN), were analyzed in triplicate. The intact-GSL content for each one is summarized in Tables 2S-5S (see Supplementary Material). As can be observed, intact-GSL concentrations were outside the linear range in several samples, and as a result, dilutions were made with ultrapure water (1:200, 1:300 and 1:400 (v/v) depending on the sample and GSL) for accurate measurement. In order to facilitate comparison and discussion of the results, the GSL data obtained were presented in a simpler format, attention being focused on the frequency (the number of samples in which a specific GSL residue was detected/the total number of samples) and concentration range (see Table 3 and Tables 6S-9S in the Supplementary Material). The samples from the different apiaries contained mainly pollen from rapeseed (Brassica t.), radish (Raphanus t.), wild rocket (Diplotaxis t.), and unidentified *Brassica* species (Brassicaceae; see Table 2), GSL content being different in each

plant. For example, for most of the rapeseed varieties, the predominant GSL is PRO, although GBN, GBC, NAS, GNA and 4-OH are also commonly detected. On the other hand, rapeseed does not usually contain ALY, NEO and GTL (European Commission, 2001). In the case of radish, albeit depending on the variety, the biggest part of GSL might be GBN or GRA, although GBN is usually found; GER, NAS, GIB and SIN, meanwhile, are not commonly detected when this vegetable is analyzed (Yi et al., 2016). Finally, GRA is the most commonly detected GSL in the different wild rocket varieties, but 4-OH, NAS, 4-ME, GTL and GER are also usually present, whilst NEO, GNA, SIN, ALY and GIB are not common in this matrix (Bell, Oruna-Concha, & Wagstaff, 2015).

By referring to Table 3, in which the overall results (frequency and concentration range of the intact-GSLs) for the four apiaries are summarized, it can be observed that thirteen GSLs were detected in some of the samples, and that NAS was the only GSL found in all the samples; residues of GIB and EPI, meanwhile, were not detected in any of them. As for differing GSL content according to the plant species, the absence of EPI, GIB and GTL in most of the samples could be explained by the fact that their presence in rapeseed, radish and wild rocket is very unusual. On the other hand, the presence of NAS in all the samples might be related with its reported presence in the three main sources of *Brassica* pollen (radish, rapeseed and wild rocket), which are the predominant and secondary pollen in most cases. However, the presence of NAS and other GSLs at relatively lower concentrations (< 500 μg/kg) in samples from the FH apiary, in which no species of *Brassica* was seen to be one of the main pollen types (FH-03 to FH-08; see Table 2), could be tentatively explained by its potential presence as a minor pollen type. Moreover, it has been reported that PRO is the predominant GSL in rapeseed and, as it is the main pollen type in most of the samples, this could be directly related to the high frequency value obtained (see Table 3). These conclusions could be also extended to the individual analysis of intact-GSL content for each apiary, although some differences can be observed (see Supplementary Material, Tables 6S-9S). For example, GIB,

EPI and GTL were not detected in FH, MO or PI apiaries, and GTL was found only in the TN apiary; this might be explained by the presence of an unknown *Brassica* source in the samples from the latter. Therefore, the difference in the types of pollen may account for the fact that intact-GSL composition and concentration observed for the bee pollen samples from the FH apiary are different from those of the other three apiaries. The GSLs with the highest levels of concentration in all the apiaries were PRO, SIN, ALY and GBN.

The detection of GSL residues in bee pollen is consistent with the scant published data relating to their analysis in pollen (Ares et al., 2015; Bernal et al., 2019; Dungey et al., 1988). In the latter study, certain GSLs were found in rapeseed (11) and Indian mustard pollen (6), but the authors provided only the standard GSL composition of these samples rather than the concentration of each GSL, and only one sample of each matrix was analyzed. In both cases, the predominant GSL was not identified/named, but, like in our study, PRO was found at a high level of concentration in rapeseed pollen. Meanwhile, Ares et al. (2015) reported the presence of nine GSLs in most of the bee pollen samples analyzed (12 out of 14) over a wide concentration range (3-2226 μg/kg); consequently, it was necessary to dilute the sample in order to measure SIN and GNA in two of the samples. Finally, in the most recent study, residues of eight GSLs in a concentration range of 34-9806 μg/kg were found in the three bee pollen samples examined. GNA and GTL displayed the highest rate of concentration and two of the samples needed to be diluted for correct quantification of GNA.

3.2. Canonical discriminant analysis

The FH apiary was not included in this analysis, as the intact-GSLs detected in its bee pollen samples make it easily distinguishable from the other apiaries, a finding apparent in Tables 2S-5S (see Supplementary Material). This might be due to the different location of this apiary from that of the others (see subsection 2.3), and also to the types of pollen in the samples from this apiary, since

some of them were not directly derived from plants belonging to the *Brassica* family (see Table 2). Quantitative variables of GIB, EPI, NEO and GTL were likewise excluded from the CDA for different reasons: **i**) GIB and EPI were not detected in any sample; **ii**) GTL residues were found only in samples from the TN apiary; **iii**) in all apiaries NEO frequency was quite low in comparison with other GSLs. It must be specified that based on the presence of GTL, the accuracy of discriminating TN from the other apiaries would be 100%, and the use of CDA might be unnecessary. However, we did not discard this apiary from the CDA in order to compare the accuracy with the obtained with the direct approach (GTL). In addition, the use of all the information provided by the rest of the measured variables guarantees a more complete statistical classification procedure. The discriminant statistical technique used in this study determines the weight of the different variables by assigning a practically zero weight to those variables that are not very useful in the classification, which does not occur with the presence of TN (see Table 4). Thus, the concentrations of eleven GSLs in the forty-one bee pollen samples from three apiaries (28 PI; 9 MO; 4 TN) were employed to perform the CDA.

Only the first and second canonical variables were used, as it was found that the former variable obtained from the CDA provided 93.53% variability of the original data, and 100% if the first two were employed. The weights of the variables were then obtained (see Supplementary Material, Table 10S), the linear combination of the intact-GSLs (quantitative variables) that used these weights determining the canonical variables. It should be noted, for example, that Pro_1, Pro_2 and Pro_3 in Table 10S refer to the values obtained in each of the three analyses performed for each bee pollen sample. From these weights and the concentrations obtained for the different GSLs, in addition to the samples and analyses (data not shown), the values of the canonical variables for each sample were calculated (see Table 4) by means of the formulas described in Supplementary Material (Table 11S). It can be observed that the first canonical variable (Can1) has positive weights in nine GSLs (PRO, GRA, GNA, 4-OH, GBN, GER GBC, NAS and 4-ME) and negative

ones in the others (SIN, ALY). Therefore, samples with a positive value in Can1 (second column of Table 4) have a higher response in PRO, GRA, GNA, 4-OH, GBN, GER GBC, NAS and 4-ME than in SIN and ALY. Samples with a negative Can1 have a higher response in SIN and ALY than in PRO, GRA, GNA, 4-OH, GBN, GER GBC, NAS and 4-ME. In this canonical variable, GRA and 4-ME have very little relevance, since their weights are practically zero. The second canonical variable (Can2) has positive weights in six GSLs (PRO, GNA, 4-OH, GBN, GBC, 4-ME) and negative ones in the remaining ones (SIN, GRA, ALY, GER, NAS). A positive value of a sample in Can2 (third column of Table 4) indicates a higher value in PRO, GNA, 4-OH, GBN, GBC and 4-ME than in SIN, GRA, ALY, GER and NAS. Samples with a negative Can2 have a higher response in SIN, GRA, ALY, GER and NAS than in PRO, GNA, 4-OH, GBN, GBC and 4-ME. In this canonical variable, SIN, GRA, ALY, 4-OH and GER are the relevant GSLs because their weights are not close to zero.

Figures 1 and 2 give a graphic representation of, respectively, the measurements of the canonical variables obtained from the apiaries and the samples. Figure 1 represents the mean values of the response to the first two canonical variables in each apiary, whilst Figure 2 displays the responses to the first two canonical variables of the 41 samples. It can be seen how the first canonical variable distinguishes between the TN apiary (positive value of Can1) and the PI and MO apiaries (negative value of Can 1). The second canonical variable discriminates between the PI hive (negative response of Can2) and that of MO (positive response of Can2). Furthermore, this CDA makes it possible to appreciate, as shown in Table 12S (see Supplementary Material), that most of the samples were correctly classified: (88.89% (8 of 9), MO apiary; 89.29% (25 of 28), PI apiary; and 100% (4 of 4), TN apiary). As can be observed, the accuracy of discriminating TN from the other two apiaries is 100%, which matched perfectly with that obtained by considering only the presence of GTL. This implies that both approaches are useful in discriminating TN samples, although in this specific case, it would be recommended the use of the simpler and faster option based on the

presence of GTL. This high rate of successful classification of bee pollen samples is important, if we consider the distance separating these apiaries (< 900 m) and the fact that the type of pollen was quite similar in many cases. In that of the MO hive, the MO-07 sample was not correctly classified (see Figure 2). As can be seen in Table 2, this sample did not contain pollen from plants of the Brassica family, which might explain why it cannot be classified in the same way as the other samples from this apiary. However, this explanation does not apply to the three incorrectly assigned samples from the PI apiary (PI-01, PI-26 and PI-27; see Figure 2), since, according to Table 1, there are no significant differences between the plant of origin and the other samples. Nevertheless, as previously mentioned, it was not possible to identify the specific variety of each plant, and this factor could explain the difference in intact-GSL content and, subsequently, in the canonical variables.

4. Conclusions

The intact-GSL content of forty-nine bee pollen samples, from 4 different apiaries (FH, MO, PI and TN) located in Marchamalo (Guadalajara, Spain), was determined by using UPLC-Q-TOF/MS. GSL residues were found in most of the samples analyzed, although these differed in number and concentration. It has been possible to relate most of the samples to their corresponding plants of origin as a result of their intact-GSL content, which has made it possible to distinguish one of the apiaries (FH) from the other three (MO, PI, and TN). Moreover, these three apiaries have been differentiated by means of a CDA based on the intact-GSL content of the pollen samples from each apiary; the first two canonical variables were used, as between them 100% variability of the original data was explained. By means of the proposed CDA, more than 88% of the samples could be assigned correctly for the PI and MO apiaries, and 100% for the TN apiary. It must be specified that the accuracy of discriminating TN from the other apiaries by considering only the presence of GTL was also 100%, which in this case was a simpler and faster alternative to CDA. To sum up, it has been demonstrated for the first time that, depending on intact-GSL content, it is possible to

distinguish and classify bee pollen samples from different apiaries located in the same geographical area and with similar plants of origin. Therefore, the potential of GSLs has been shown to pinpoint the origin of bee pollen, which will consequently facilitate its labeling and prevent possible adulteration or fraudulent practice. However, this study should be considered as a starting point for new studies, in which the proposed methodology could be applied to analyzing a greater number and diversity (origin) of samples.

Funding

This work was supported by the Spanish Ministry "*Economía y Competitividad*" and the "*Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria*" (project numbers RTA2015-00013-C03-01 and 03).

Acknowledgements

Authors thank David Rixham (White Rose English School, Valladolid, Spain) for performing the performing the revision of the English language/grammar of the manuscript. In addition, authors thank the Laboratory of Instrumental Techniques (University of Valladolid, Spain) for using their UPLC-Q-TOF/MS system.

Declaration of interest

None.

Abbreviations:

ALY, glucoalyssin; **Can1,** first canonical variable; **Can2,** second canonical variable; **CDA,** canonical discriminant analysis; **CIAPA,** *Centro de Investigación Apícola y Agroambiental*; **EPI,** epiprogoitrin; **ESI,** electrospray ionization; **FH,** *Fuentelahiguera;* **GBC,** glucobrassicin; **GBN,** glucobrassicanapin; **GER,** glucoerucin; **GIB,** glucoiberin; **GSL**, glucosinolate; **GNA,** gluconapin;

GRA, glucoraphanin; **GTL,** glucotropaeolin; **LC,** liquid chromatography; **LOD,** limit of detection; **LOQ,** limit of quantification; **4-ME,** 4-methoxyglucobrassicin; **MO,** *Monte*; **NAS,** gluconasturtiin; **NEO,** neoglucobrassicin; **4-OH,** 4-hydroxyglucobrassicin; **PI,** *pistacho*; **PRO,** progoitrin; **Q-TOF/MS**, quadrupole time-of-flight mass detector**; SIN,** sinigrin; **SPE,** solid phase extraction; **TN,** *Tío Natalio*; **UPLC,** ultra-performance liquid chromatography.

References

Anjos, O., Fernandes, R., Cardoso, S. M., Delgado, T., Farinha, N., Paula, V., Estevinho, L. M., & Carpes, S. T. (2019). Bee pollen as a natural antioxidant source to prevent lipid oxidation in black pudding. *LWT-Food Science and Technology*, *111*, 869-875. https://doi.org/10.1016/j.lwt.2019.05.105.

Ares, A. M., Nozal, M. J., & Bernal, J. (2015). Development and validation of a liquid chromatography-tandem mass spectrometry method to determine intact glucosinolates in bee pollen. *Journal of Chromatography B*, *1000*, 49-56. https://doi.org/10.1016/j.jchromb.2015.07.017.

Ares, A. M., Valverde, S., Bernal, J. L., Nozal, M. J., & Bernal, J. (2018). Extraction and determination of bioactive compounds from bee pollen. *Journal of Pharmaceutical and Biomedical Analysis*, *147*, 110-124. https://doi.org/10.1016/j.jpba.2017.08.009.

Bell, L., Oruna-Concha, M. J., & Wagstaff, C. (2015). Identification and quantification of glucosinolate and flavonol compounds in rocket salad (*Eruca sativa*, *Eruca vesicaria* and *Diplotaxis tenuifolia*) by LC-MS: highlighting the potential for improving nutritional value of rocket crops. *Food Chemistry*, *172*, 852-861. https://doi.org/10.1016/j.foodchem.2014.09.116.

Bernal, J., González, D., Valverde, S., Toribio, L.,& Ares, A. M. (2019). Improved separation of intact glucosinolates in bee pollen by using ultra-high-performance liquid chromatography coupled to quadrupole time-of-flight mass spectrometry. *Food Analytical Methods*, *12*, 1170-1178. https://doi.org/10.1007/s12161-019-01446-2.

Cepero, A., Ravoet, J., Gómez-Moracho, T., Bernal, J. L., Del Nozal, M. J., Bartolomé, C., Maside, X., Meana, A., González-Porto, A. V., de Graaf, D. C., Martín-Hernández, R., Higes, M. (2014). Holistic screening of collapsing honey bee colonies in Spain: a case study. *BMC Research Notes*, *7*, 649. https://doi.org/10.1186/1756-0500-7-649.

Chica, M., & Campoy, P. (2012). Discernment of bee pollen loads using computer vision and one-class classification techniques. *Journal of Food Engineering*, *112*, 50-59. https://doi.org/10.1016/j.jfoodeng.2012.03.028.

Dungey, S. G., Sang, J. P., Rothnie, N. E., Palmer, M. V., Burke, D. G., Knox, R. B., Williams, E. G., Hilliard, E. P., & Salisbury, P. A. (1988). Glucosinolates in the pollen of rapeseed and Indian mustard. *Phytochemistry*, *27*, 815-817. https://doi.org/10.1016/0031-9422(88)84098-6.

European Commission. (2001). The certification of the total glucosinolate content and Sulphur contents of three rapeseed (colza) materials BCR-190R, -366r and -367R. http://publications.jrc.ec.europa.eu/repository/bitstream/JRC21284/EUR%2019764%20EN.pdf Accessed 4 May 2020.

Faegri, K., & Iversen, J. (1989). *Textbook of pollen analysis*. (4th ed.). New York: Wiley.

Glauser, G., Schweizer, F., T. C. J. Turlings, & Reymond, P. (2012). Rapid profiling of intact glucosinolates in *Arabidopsis* leaves by UHPLC- QTOFMS using a charged surface hybrid column. *Phytochemical Analysis*, *23*, 520-528. https://doi.org/10.1002/pca.2350.

Gonçalves, P. J. S., Estevinho, L. M., Pereira, A. P., Sousa, J. M. C., & Anjos, O. (2018). Computational intelligence applied to discriminate bee pollen quality and botanical origin. *Food Chemistry*, *267*, 36-42. https://doi.org/10.1016/j.foodchem.2017.06.014.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference and prediction*. (2nd ed.). New York: Springer.

Hidalgo, M. I., & Bootello, M. L. (1990). About some physical characteristics of the pollen loads collected by *Apis mellifera* L. *Apicoltura*, *6*, 179-191.

Kaškonienė, V., Ruočkuvienė, Kaškonas, P., Akuneca, I., & Maruška, A. (2015). Chemometric analysis of bee pollen based on volatile and phenolic compound compositions and antioxidant properties. *Food Analytical Methods*, *8*, 1150-1163. https://doi.org/10.1007/s12161-014-9996-2.

Negrão, A. F., & Orsi, R. O. (2018). Harvesting season and botanical origin interferes in production and nutritional composition of bee pollen. *Annals of the Brazilian Academy of Sciences*, *90*, 325-332. https://doi.org/10.1590/0001-3765201720150192.

Rossetto, M. R. M., Shiga, T. M., Vianello, F., & Lima, G. P. P. (2013). Analysis of total glucosinolates and chromatographically purified benzylglucosinolate in organic and conventional

vegetables. *LWT-Food Science and Technology*, *50*, 247-252. https://doi.org/10.1016/j.lwt.2012.05.22.

Sattler, J. A. G., de Melo, I. L. P., Granato, D., Araújo, E., de Freitas, A. S., Barth, O. M., Sattler, A., & de Almeida-Muradian, L. B. (2015). Impact of origin on bioactive compounds and nutritional composition of bee pollen from southern Brazil: a screening study. *Food Research International*, *77*, 82-91. https://doi.org/10.1016/j.foodres.2015.09.013.

Truchado, P., Tourn, E., Gallez, L. M., Moreno, D. A., Ferreres, F., & Tomás-Barberán, F. A. (2010). Identification of botanical biomarkers in Argentinean *Diplotaxis* honeys: flavonoids and glucosinolates. *Journal of Agricultural and Food Chemistry*, *58*, 12678-12685. https://doi.org/10.1021/jf103589c.

Yi, G., Lim, S., Chae, W. B., Park, J. E., Park, H. R., Lee, E. J., & Huh, J. H. (2016). Root glucosinolate profiles for screening of Radish (*Raphanus sativus* L.) genetic resources. *Journal of Agricultural and Food Chemistry*, *64*, 61-70. https://doi.org/10.1021/acs.jafc.5b04575.

Zhou, J., Qi, Y., Ritho, J., Zhang, Y., Zheng, X., Wu, L., Li, Y., & Sun, L. (2015). Flavonoid glycosides as floral origin markers to discriminate unifloral bee pollen by LC-MS/MS. *Food Control*, *57*, 54-61. https://doi.org/10.1016/j.foodcont.2015.03.035.

Figure captions

Figure 1.- Representation of apiaries (*Pistacho*, PI; *Monte*, MO; *Tío Natalio*, TN) as function of the first two canonical variables.

Figure 2.- Representation of individual bee pollen samples (see Table 2) from *Pistacho* (PI), *Monte* (MO) and *Tío Natalio* (TN) apiaries as function of the first two canonical variables.

Figure 1

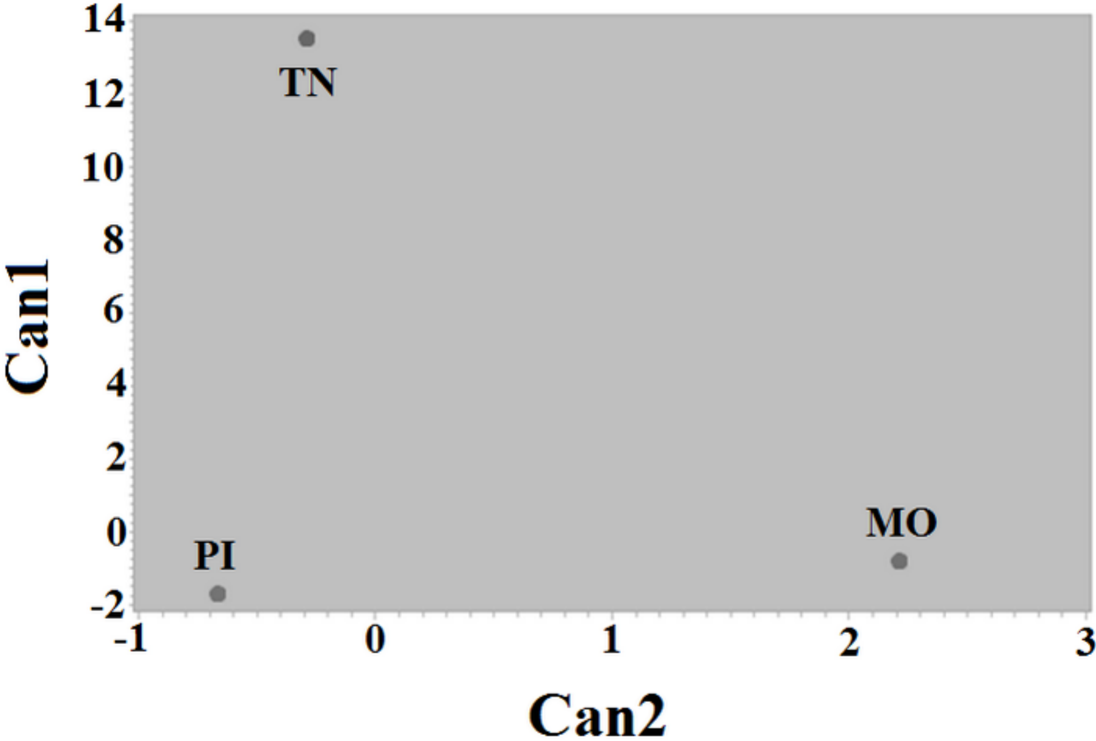


Figure 2

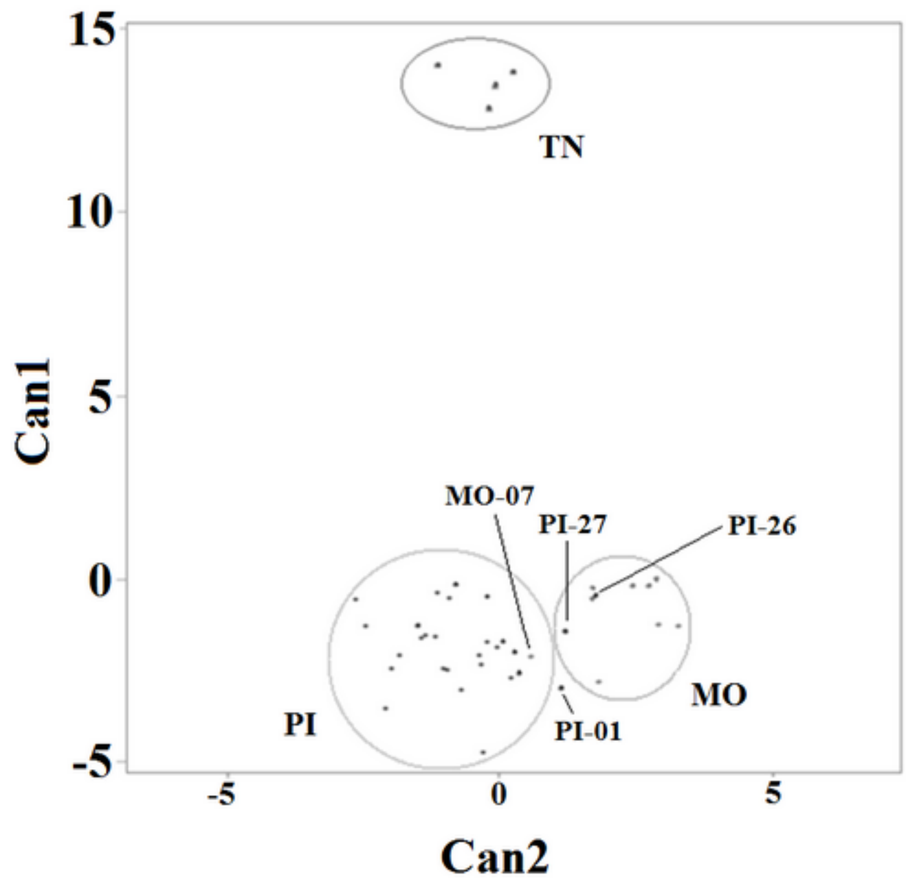


Table 1.- UPLC-Q-TOF/MS data and limit of detection/quantification (LOD/LOQ) values of each intact-GSLs.

Compound name (abbreviation)	Molecular weight	Retention time (min)	Precursor ions[A] (m/z)	Product ions[B] (m/z)	LOD[C] (µg/kg)	LOQ[C] (µg/kg)
Glucoiberin (GIB)	423	1.1	422.0044	96.9540	18	60
Progoitrin (PRO)	389	1.3	388.0203	96.9543	18	60
Sinigrin (SIN)	359	1.5	358.0120	96.9544	16	52
Epiprogoitrin (EPI)	389	1.6	388.0283	96.9539	12	38
Glucoraphanin (GRA)	437	1.8	436.0291	96.9537	14	44
Gluconapin (GNA)	373	3.3	372.0269	96.9542	7	23
Glucoalyssin (ALY)	451	3.6	450.0500	96.9541	16	55
4-hydroxyglucobrassicin (4-OH)	464	3.8	463.0337	96.9540	25	80
Glucobrassicanapin (GBN)	387	4.0	386.0379	96.9544	8	26
Glucotropaeolin (GTL)	409	4.2	408.0224	96.9542	7	21
Glucoerucin (GER)	421	4.3	420.0400	96.9536	10	32
Glucobrassicin (GBC)	448	4.8	447.0348	96.9542	9	34
Gluconasturtiin (NAS)	423	5.6	422.0401	96.9543	6	19
4-Metoxyglucobrassicin (4-ME)	478	6.8	477.0403	96.9538	28	88
Neoglucobrassicin (NEO)	478	7.3	477.0446	96.9540	8	25

[A]Quantification and confirmation; [B]Confirmation.[C]Data obtained from Bernal et al., (2019).

Table 2.- Plant families given the predominant and secondary pollen in case of non-homogenous pollen balls.

Sample	Predominant pollen	Secondary pollen
FH-01	Brassica t.	NP
FH-02	Brassica t.	NP
FH-03	Brassicaceae	NP
FH-04	Prunus t.	NP
FH-05	Cistus t.	NP
FH-06	Cistus t.	NP
FH-07	Leguminosae	NP
FH-08	Leguminosae	Rosaceae
MO-01	Brassica t.	NP
MO-02	Brassica t.	NP
MO-03	Brassica t.	NP
MO-04	Brassicaceae	NP
MO-05	Brassica t.	NP
MO-06	Brassicaceae	NP
MO-07	Cistus t.	Hypecoum + Brassicaceae
MO-08	Brassicaceae	Papaver
MO-09	Quercus sp.	Brassicaceae
PI-01	Brassica t.	NP
PI-02	Brassica t.	NP
PI-03	Brassica t.	NP
PI-04	Brassica t.	NP
PI-05	Brassica t.	Diplotaxis t.
PI-06	Brassica t.	NP
PI-07	Brassica t.	Diplotaxis t.
PI-08	Brassica t.	Diplotaxis t.
PI-09	Diplotaxis t.	Brassica t.
PI-10	Brassica t.	Diplotaxis t.
PI-11	Brassica t.	NP
PI-12	Brassica t.	NP
PI-13	Brassica t.	NP
PI-14	Brassica t.	NP
PI-15	Brassica t.	NP
PI-16	Brassicaceae	NP
PI-17	Brassicaceae	Raphanus t.
PI-18	Brassicaceae	NP
PI-19	Brassicaceae	NP
PI-20	Brassica t.	NP

Table 2.- Continued.

Sample	Predominant pollen	Secondary pollen
PI-21	Brassica t.	NP
PI-22	Brassica t.	NP
PI-23	Brassicaceae	NP
PI-24	Brassica t.	NP
PI-25	Brassica t.	NP
PI-26	Brassicaceae	NP
PI-27	Brassica t.	NP
PI-28	Brassica t.	NP
TN-01	Brassica t.	NP
TN-02	Raphanus t.	Brassica t.
TN-03	Brassicaceae	NP
TN-04	Papaver	Brassica t

NP: not present

Table 3.- Overall frequency and concentration range data of each intact-GSL.

	Frequency[A] (%)	Concentration range (µg/kg)
GIB	0	<LOD
PRO	94	377-272871
SIN	78	<LOD-581786
EPI	0	<LOD
GRA	92	89-6311
GNA	90	915-15014
ALY	94	63-163318
4-OH	53	<LOQ-1585
GBN	96	126-145695
GTL	8	<LOD-4479
GER	88	107-2090
GBC	96	40-11323
NAS	100	89-24986
4-ME	63	95-595
NEO	24	1186-2827

[A]number of samples in which a GSL residue was detected/total number of samples (n=49)

Table 4.- Values of the first two canonical variables for each bee pollen sample.

Sample	Can1	Can2
MO-01	-1,126	3,249
MO-02	-2,637	1,810
MO-03	-0,038	2,413
MO-04	0,188	2,859
MO-05	-0,032	2,704
MO-06	-0,358	1,703
MO-07	-1,935	0,553
MO-08	-0,060	1,676
MO-09	-1,084	2,887
PI-01	-2,275	-1,035
PI-02	-0,319	-0,224
PI-03	-2,159	-0,329
PI-04	-0,342	-0,923
PI-05	-1,379	-1,378
PI-06	-1,449	-1,429
PI-07	-2,790	1,121
PI-08	-2,845	-0,693
PI-09	-1,692	-0,051
PI-10	-0,211	-1,153
PI-11	-1,897	-0,370
PI-12	-1,550	-0,231
PI-13	-2,247	-1,994
PI-14	-1,096	-1,489
PI-15	-1,390	-1,182
PI-16	-1,899	-1,820
PI-17	-4,542	-0,312
PI-18	0,007	-0,832
PI-19	-1,098	-2,448
PI-20	-3,332	-2,085
PI-21	-2,532	0,215
PI-22	-0,376	-2,631
PI-23	-2,411	0,361
PI-24	-1,542	0,039
PI-25	-2,291	-0,970
PI-26	-0,338	1,708
PI-27	-1,246	1,180
PI-28	-1,833	0,270
TN-01	13,836	0,240
TN-02	13,459	-0,088
TN-03	14,019	-1,131
TN-04	12,839	-0,187

Supplementary Material

DIFFERENTIATION OF BEE POLLEN SAMPLES ACCORDING TO THEIR INTACT-GLUCOSINOLATE CONTENT USING CANONICAL DISCRIMINANT ANALYSIS

Ana M. Ares[1]*, Montserrat Redondo[1], Jesús Tapia[1,2], Amelia V. González-Porto[3], Mariano Higes[3], Raquel Martín-Hernández[3,4], José Bernal[1,*]

[1]I. U. CINQUIMA, Analytical Chemistry Group, Faculty of Sciences, University of Valladolid, Valladolid, Spain.

[2]Department of Statistics and Operations Research, Faculty of Sciences, University of Valladolid, Valladolid, Spain.

[3]Instituto Regional de Investigación y Desarrollo Agroalimentario y Forestal de Castilla La Mancha (IRIAF), Centro de Investigación Apícola y Agroambiental (CIAPA) de Marchamalo, Guadalajara, Spain.

[4]Instituto de Recursos Humanos para la Ciencia y la Tecnología (INCRECYT-FEDER), Fundación Parque Científico y Tecnológico de Castilla – La Mancha, Guadalajara, Spain.

***Corresponding author:** Tel# 34-983186347; ana.maria.ares@uva.es; jose.bernal@uva.es; web: tesea.uva.es

Table 1S.- UPLC elution program.

Time (min)	% 0.1% (v/v) formic acid in water	% 0.1% (v/v) formic acid in acetonitrile
0.0	100	0
2.0	100	0
2.1	90	10
6.0	90	10
6.1	50	50
7.5	50	50
7.6	10	90
9.5	10	90
11.0	35	65
12.5	100	0
15.0	100	0

Table 2S.- Mean and confident interval (95%) values in µg/kg obtained for the intact-GSLs in the samples from MO apiary. The intact-GSLs not detected in those samples (GIB, EPI and GTL) were not included.

	PRO		SIN		GRA		GNA		ALY		4-OH		GBN		GER		GBC		NAS		4-ME		NEO	
MO-01	197281	±18473	581786	±19825	4520	±321	15014	±418	116210	±6036	1295	±222	126931	±5537	1267	±17	638	±10	19211	±417	237	±36	1485	±23
MO-02	76938	±2048	90659	±30122	4602	±285	5476	±10	124515	±2258	854	±256	121086	±4017	864	±4	1085	±56	12703	±487	218	±5	2506	±23
MO-03	145003	±804	158459	±4320	6311	±18	7122	±69	150087	±8334	1048	±199	145695	±6392	1514	±7	650	±10	12976	±206	224	±5	1186	±69
MO-04	68883	±1576	324822	±8295	3671	±124	6908	±588	5172	±4	489	±153	99549	±2707	476	±20	304	±17	10322	±149	156	±26	1875	±174
MO-05	66587	±283	26673	±2110	3999	±10	4919	±109	125416	±2799	949	±72	106655	±6446	618	±12	987	±70	10840	±525	230	±16	2827	±335
MO-06	6364	±153	9600	±319	1099	±48	915	±67	18940	±1067	<LOD		18867	±968	107	±9	69	±3	3124	±27	95	±1	<LOD	
MO-07	1251	±95	<LOD		469	±3	<LOD		1020	±172	<LOD		1488	±350	<LOD		50	±2	157	±96	<LOD		<LOD	
MO-08	20849	±1158	<LOD		1804	±51	1422	±114	27036	±3212	<LOD		42605	±820	175	±30	430	±6	2219	±55	<LOD		<LOD	
MO-09	112631	±4990	<LOD		1972	±246	2639	±147	33071	±38	490	±117	61933	±1780	710	±136	689	±120	5244	±234	<LOD		<LOD	

Table 3S.- Mean and confident interval (95%) values in µg/kg obtained for the intact-GSLs in the samples from PI apiary. The intact-GSLs not detected in those samples (GIB, EPI and GTL) were not included.

	PRO		SIN		GRA		GNA		ALY		4-OH		GBN		GER		GBC		NAS		4-ME		NEO	
PI-01	19315	±1827	67291	±1522	2000	±137	3958	±38	86696	±3025	291	±45	39150	±465	338	±40	664	±35	7211	±896	198	±11	< LOD	
PI-02	59023	±35	96813	±1255	3384	±283	5126	±193	148811	±6469	512	±9	110127	±1250	503	±22	670	±85	11651	±260	186	±1	1190	±182
PI-03	55518	±6525	49781	±28288	3843	±217	4368	±90	143064	±6673	825	±40	89507	±6109	323	±21	421	±7	10079	±425	195	±4	<LOD	
PI-04	32801	±4530	483734	±17724	2636	±131	4883	±52	89146	±5076	<LOQ		72541	±105	199	±8	311	±1	5734	±327	157	±9	<LOD	
PI-05	117526	±2859	535901	±11861	2983	±279	4814	±495	75219	±9615	609	±76	101664	±3928	898	±111	498	±91	14987	±3076	183	±26	<LOD	
PI-06	252223	±9498	310099	±34259	4039	±324	8996	±1256	92917	±2184	1585	±536	129772	±2499	2033	±350	1132	±213	24986	±1121	245	±34	<LOD	
PI-07	44604	±1421	43632	±665	3258	±278	3162	±226	103747	±1670	178	±118	94982	±13	693	±3	524	±23	9453	±639	157	±5	<LOD	
PI-08	73874	±5213	196265	±1016	3916	±181	3921	±4	126316	±9606	421	±5	96278	±3234	359	±5	439	±2	10954	±955	132	±6	<LOD	
PI-09	248746	±7455	359082	±10120	3898	±304	6024	±775	30638	±911	1094	±167	112161	±2370	1223	±47	932	±18	16761	±1546	241	±19	1787	±29
PI-10	46999	±1633	211559	±9923	3232	±25	2834	±319	96809	±10813	<LOD		79891	±1513	427	±30	430	±32	7559	±276	224	±28	<LOD	
PI-11	47673	±3733	87538	±12239	897	±70	5009	±474	137330	±8150	288	±174	77381	±5872	364	±27	371	±38	9683	±634	182	±3	<LOD	
PI-12	11851	±970	123217	±7000	1420	±20	2500	±172	67485	±532	259	±35	35901	±2209	290	±15	408	±36	5572	±327	167	±1	<LOD	
PI-13	35587	±1311	260924	±2040	2662	±204	2617	±219	98926	±1667	<LOD		75455	±692	435	±31	327	±41	7155	±194	146	±10	<LOD	
PI-14	53954	±7402	148589	±1671	4090	±215	4460	±369	163318	±5371	<LOD		102407	±10270	533	±6	705	±65	12585	±299	139	±5	<LOD	
PI-15	186111	±310	156024	±5652	2718	±222	6585	±398	82012	±472	839	±122	107219	±2112	1021	±46	555	±49	15784	±1132	152	22	<LOD	
PI-16	28701	±4527	396146	±21752	2833	±111	3693	±457	95055	±1142	<LOD		64717	±1883	236	±20	236	±39	6758	±197	149	±5	<LOD	
PI-17	24669	±53	394610	±12200	2928	±84	3850	±126	81589	±1556	<LOD		75934	±1839	345	±8	263	±3	6691	±246	191	±7	<LOD	
PI-18	36598	±1251	95278	±3849	4775	±104	4651	±144	54252	±4994	<LOD		48449	±3711	409	±38	726	±16	3068	±150	<LOD		<LOD	
PI-19	59750	±4051	244052	±1998	3575	±266	5043	±816	50489	±4819	541	±112	89186	±2423	1586	±173	113	±15	8491	±189	<LOD		<LOD	
PI-20	53185	±276	192459	±1957	4434	±88	4824	±43	59648	±771	967	±245	84640	±2803	1357	±51	278	±1	5358	±42	<LOD		<LOD	
PI-21	37433	±3909	166662	±8186	4567	±185	4762	±139	72850	±2269	<LOD		61191	±200	1032	±63	115	±13	4725	±274	<LOD		<LOD	
PI-22	65270	±1735	111308	±17205	6247	±199	5583	±67	115616	±6619	<LOD		51266	±2172	1066	±11	516	±12	5629	±243	<LOD		<LOD	
PI-23	35995	±6161	317518	±5804	4427	±60	3807	±243	72052	±1350	<LOD		68920	±3469	580	±43	54	±5	4641	±482	<LOD		<LOD	
PI-24	214850	±5256	162051	±554	5606	±105	7781	±196	74228	±1042	1518	±67	86665	±528	2090	±174	106	±3	10765	±565	<LOD		<LOD	
PI-25	60436	±5247	172077	±6853	5180	±325	5466	±174	96299	±3764	<LOD		85058	±670	1591	±68	110	±24	6321	±231	<LOD		<LOD	
PI-26	4622	±368	82991	±1229	1090	±10	3028	±21	18946	±4939	458	±71	65271	±1631	761	±26	157	±34	4618	±109	<LOD		<LOD	
PI-27	8813	±156	3663	±494	1846	±15	931	±28	12170	±632	356	±22	14067	±699	308	±21	965	±21	1385	±9	111	±1	<LOD	
PI-28	18582	±712	23143	±2914	4059	±74	1203	±103	46990	±2041	<LOD		30299	±1209	200	±2	346	±58	3963	±348	<LOD		<LOD	

Table 4S.- Mean and confident interval (95%) values in μg/kg obtained for the intact-GSLs in the samples from TN apiary. The intact-GSLs not detected in those samples (GIB and EPI) were not included.

	PRO		SIN		GRA		GNA		ALY		4-OH		GBN		GTL		GER		GBC		NAS		4-ME		NEO	
TN-01	202115	±2545	41990	±1848	3015	±118	6001	±1117	23233	±409	942	±115	111435	±8082	266	±63	1267	±164	230	±7	16586	±38	156	±17	1227	±127
TN-02	272871	±8929	37603	±3814	4320	±189	13505	±391	104488	±1324	747	±167	119068	±662	1452	±182	1370	±98	226	±41	18519	±453	171	±7	1649	±316
TN-03	218459	±2771	162310	±21799	3222	±37	8105	±1153	70905	±3931	690	±20	118977	±5859	600	±66	1887	±199	325	±23	20236	±843	199	±19	<LOD	
TN-04	8373	±224	20811	±209	3983	±7	3164	±202	18787	±202	<LOD		36111	±1061	4479	±152	121	±15	1791	±239	2099	±111	<LOD		<LOD	

Table 5S.- Mean and confident interval (95%) values in µg/kg obtained for the intact-GSLs in the samples from FH apiary. The intact-GSLs not detected in those samples (GIB, SIN, EPI and GTL) were not included.

	PRO		GRA		GNA		ALY		4-OH		GBN		GER		GBC		NAS		4-ME		NEO	
FH-01	75452	±1448	4091	±32	5399	±123	98370	±2351	802	±161	131448	±829	1172	±91	135	±17	10429	±366	140	±8	2044	±391
FH-02	77386	±10179	4106	±56	5439	±351	128357	±8264	<LOD		130211	±557	782	±48	182	±1	13338	±227	160	±21	2411	±44
FH-03	75945	±1670	4061	±208	4758	±97	72888	±9482	<LOD		91465	±5885	354	±1	121	±6	4882	±137	<LOD		<LOD	
FH-04	470	±13	<LOD		<LOD		361	±42	<LOD		353	±15	<LOD		<LOD		190	±12	<LOD		312	±15
FH-05	377	±8	312	±3	<LOD		63	±7	425	61	158	±2	<LOD		40	±6	110	±2	395	±12	<LOD	
FH-06	<LOD		<LOD		<LOD		159	±15	110	39	<LOD		<LOD		<LOD		89	±4	270	±8	<LOD	
FH-07	<LOD		<LOD		<LOD		<LOD		<LOD		126	±7	<LOD		480	±30	122	±10	<LOD		<LOD	
FH-08	<LOD		<LOD		<LOD		<LOD		<LOD		<LOD		<LOD		323	±25	258	±47	<LOD		<LOD	

Table 6S.- Frequency and concentration range data of each intact-GSL from MO apiary.

	Frequency[A] (%)	Concentration range (µg/kg)
GIB	0	<LOD
PRO	100	1251-197281
SIN	67	9600-581786
EPI	0	<LOD
GRA	100	469-6311
GNA	89	915-15014
ALY	100	1020-150087
4-OH	67	489-1295
GBN	100	1488-145695
GTL	0	<LOD
GER	89	107-1514
GBC	100	50-1085
NAS	100	157-19211
4-ME	67	95-237
NEO	56	1186-2827

[A]number of samples in which a GSL residue was detected/total number of samples (n=9)

Table 7S.- Frequency and concentration range data of each intact-GSL from PI apiary.

	Frequency[A] (%)	Concentration range (µg/kg)
GIB	0	<LOD
PRO	100	4622-252223
SIN	100	3663-535901
EPI	0	<LOD
GRA	100	89-6247
GNA	100	931-8996
ALY	100	12170-163318
4-OH	61	<LOQ -1585
GBN	100	14067-129772
GTL	0	<LOD
GER	100	199-2090
GBC	100	54-1132
NAS	100	1385-24986
4-ME	64	111-245
NEO	7	1190-1787

[A]number of samples in which a GSL residue was detected/total number of samples (n=28)

Table 8S.- Frequency and concentration range data of each intact-GSL from TN apiary.

	Frequency[A] (%)	**Concentration range (µg/kg)**
GIB	0	<LOD
PRO	100	8373-272871
SIN	100	20811-162310
EPI	0	<LOD
GRA	100	3015-4320
GNA	100	3164-13505
ALY	100	18787-104488
4-OH	75	690-942
GBN	100	36111-119068
GTL	100	266-4479
GER	100	121-1887
GBC	100	226-1791
NAS	100	2099-20236
4-ME	75	156-199
NEO	50	1227-1649

[A]number of samples in which a GSL residue was detected/total number of samples (n=4)

Table 9S.- Frequency and concentration range data of each intact-GSL from FH apiary.

	Frequency[A] (%)	**Concentration range (µg/kg)**
GIB	0	<LOD
PRO	50	377-77386
SIN	0	<LOD
EPI	0	<LOD
GRA	50	312-4106
GNA	38	4758-5439
ALY	75	63-128357
4-OH	61	110-802
GBN	38	126-131448
GTL	0	<LOD
GER	38	354-1172
GBC	75	40-480
NAS	100	110-13338
4-ME	50	140-395
NEO	38	312-2411

[A]number of samples in which a GSL residue was detected/total number of samples (n=8)

Table 10S.- Weights of the first two canonical variables

Variable	Can1	Can2
PRO_1	0,408	0,015
PRO_2	0,418	0,004
PRO_3	0,410	0,022
SIN_1_	-0,248	-0,203
SIN_2	-0,240	-0,199
SIN_3	-0,248	-0,197
GRA_1	0,046	-0,109
GRA_2	0,060	-0,134
GRA_3	0,042	-0,104
GNA_1	0,333	0,057
GNA_2	0,356	0,057
GNA_3	0,311	0,061
ALY_1	-0,199	-0,211
ALY_2	-0,206	-0,202
ALY_3	-0,195	-0,207
4_OH_1	0,117	0,190
4_OH_2	0,087	0,155
4_OH_3	0,148	0,224
GBN_1	0,173	0,039
GBN_2	0,171	0,039
GBN_3	0,168	0,044
GER_1	0,228	-0,136
GER_2	0,238	-0,141
GER_3	0,212	-0,134
GBC_1	0,161	0,146
GBC_2	0,169	0,099
GBC_3	0,151	0,170
NAS_1	0,305	-0,043
NAS_2	0,300	-0,045
NAS_3	0,312	-0,044
4_ME_1	0,056	0,090
4_ME_2	0,055	0,109
4_ME_3	0,054	0,078

Table 11S.- Formulas employed to calculate the values of the first two canonical variables for each sample.

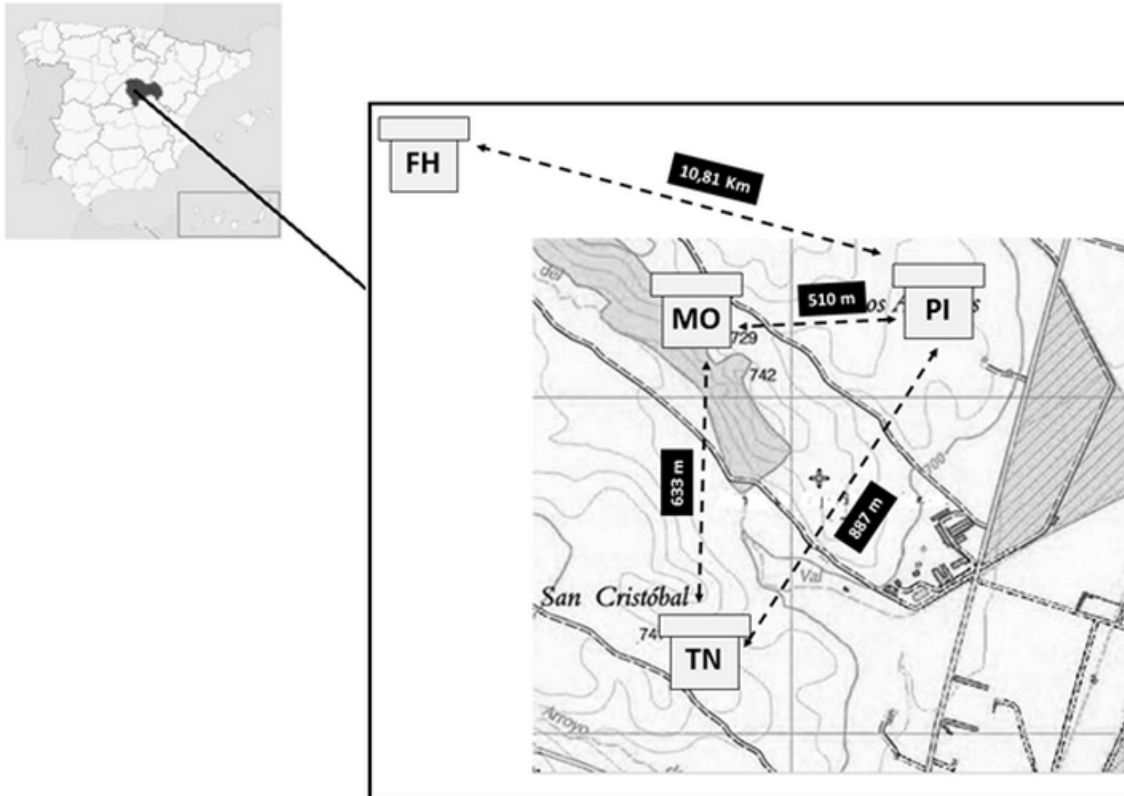
Canonical variable	Formula
Can1	0.408×PRO_1+0.418×PRO_2+0.410×PRO_3–0.248×SIN_3+0.046×GRA_1+0.06×GRA_2+0.042×GRA_3+0.333×GNA_1+0.356×GNA _2+0.311×GNA_3–0.199×ALY_1–0.206×ALY_2–0.195×ALY_3+0.117×4_OH_1+0.087×4_OH_2+0.148×4_OH_3+0.173×GBN_1+0.1 71×GBN_2+0.168×GBN_3+0.228×GER_1+0.238×GER_2+0.212×GER_3+0.161×GBC_1+0.169×GBC_2+0.151×GBC_3+0.305×NAS_ 1+0.300×NAS_2+0.312×NAS_3+0.056×4_ME_1+0.055 ME_2+0.054×4_ME_3
Can2	0.015×PRO_1+0.004×PRO_2+0.022×PRO_3–0.203×SIN_1–0.199×SIN_2–0.197×SIN_3–0.109×GRA_1–0.134×GRA_2–0.104×GRA_3+ 0.057×GNA_1+0.057×GNA:2+0.061×GNA_3+0.013×ALY_1+0.015×ALY_2+0.015×ALY_3+0.190×4_OH_1+0.155×4_OH_2+0.224×4_ OH_3+0.039×GBN_1+0.039×GBN_2+0.044×GBN_3–0.136×GER_1–0.141×GER_2–0.134×GER_3+0.146×GBC_1+0.099×GBC_2+0.170 ×GBC_3–0.043×NAS_1–0.045×NAS_2–0.044×NAS_3+0.090×4_ME_1+0.109×4_ME_2+0.078×4_ME_3

Numbers (0.408,…) correspond to the weights summarized in Table S9; Pro_1, Pro_2,…refers to concentrations detected for each intact-GSL in the three replicates.

Table 12S.- Number of observations and percentage classified in each apiary using canonical discriminant analysis

Apiary	MO	PI	TN	Total
MO	8	1	0	9
%	88.89	11.11	0.00	100.00
PI	3	25	0	28
%	10.71	89.29	0.00	100.00
TN	0	0	4	4
%	0.00	0,000	100.00	100.00
Total	11	26	4	41
	26.83	63.41	9.76	100.00

Figure 1S.- Location of the experimental apiaries (*Fuentelahiguera*, FH; *Pistacho*, PI; *Monte*, MO; *Tio Natalio*, TN).



Apiary	Latitude	Longitude
MO	40°41´13.38´´N	3°13´8.00´´W
TN	40°40´45.00´´N	3°13´1.01´´W
PI	40°41´8.19´´N	3°12´38.32´´W
FH	40°45´7.18´´N	3°18´15.54´´W

Figure 2S.- Analytical procedure work-up flow chart.

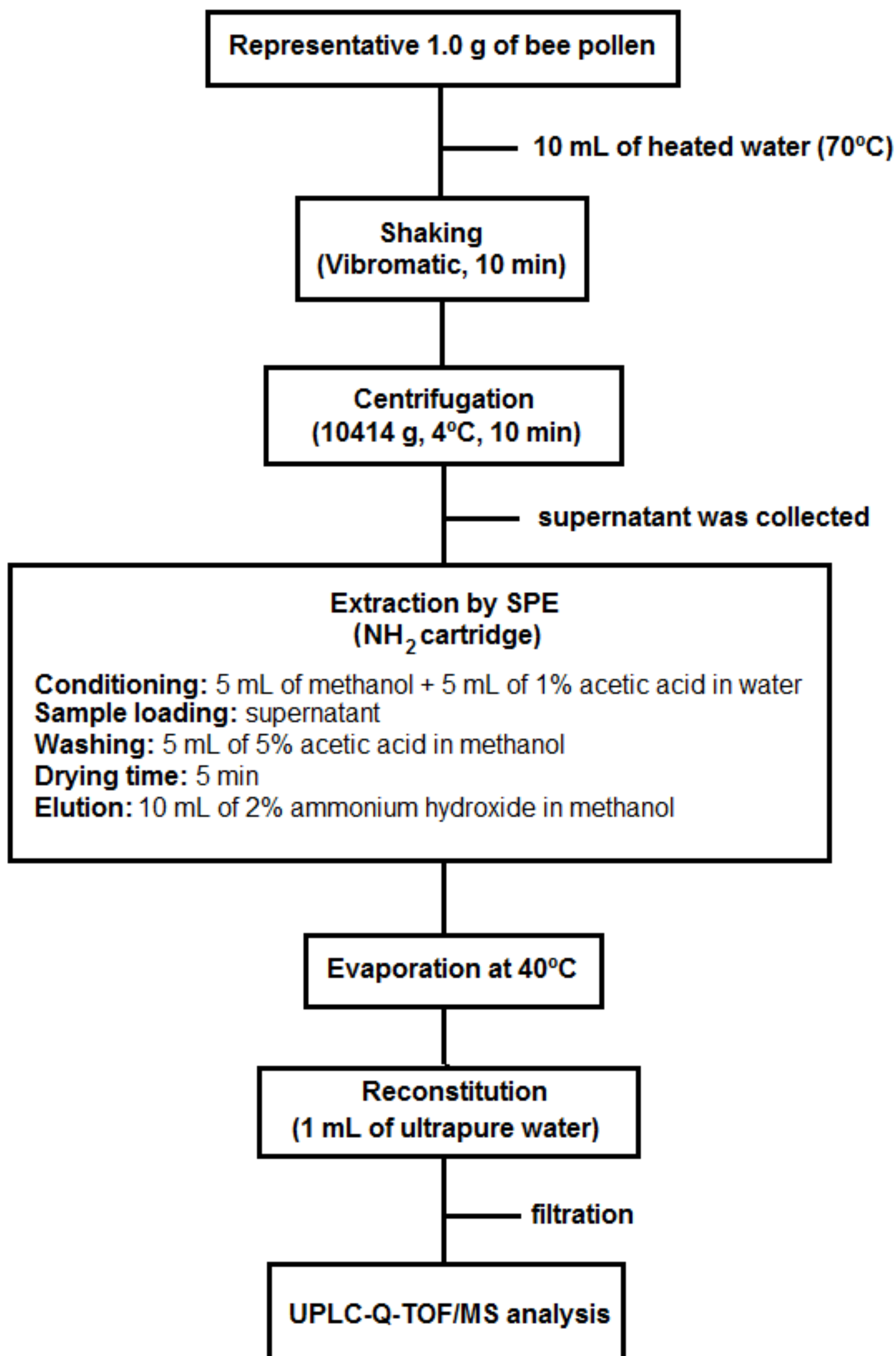


Figure 3S.- UPLC-Q-TOF/MS extracted ion chromatogram obtained in positive mode using the quantification ions (see Table 1) from TN-02 bee pollen sample in which thirteen intact-GSLs were detected: PRO (**2**); SIN (**3**); GRA (**5**); GNA (**6**); ALY (**7**); 4-OH (**8**); GBN (**9**); GTL (**10**); GER (**11**); GBC (**12**); NAS (**13**); 4-Me (**14**); NEO (**15**). The UPLC-Q-TOF/MS conditions are summarized in subsection 2.4 and Table 1.