Received: 30 September 2024 | Revised: 5 December 2024 | Accepted: 6 December 2024

DOI: 10.1002/jsf2.228

RESEARCH ARTICLE

JSFA Reports

Identification of pollen types of beekeeping interest by non-targeted mass spectrometry

Valeria Leoni[1,2,3,4] | Christina Kast[3] | Ueli Bütikofer[4] | Luca Giupponi[1,2] | Reto Portmann[4]

[1]Department of Agricultural and Environmental Sciences-Production, Landscape, Agroenergy (DISAA), University of Milan, Milan, Italy

[2]Centre of Applied Studies for the Sustainable Management and Protection of Mountain Areas (CRC Ge.S.Di.Mont.), University of Milan, Milan, Italy

[3]Agroscope, Swiss Bee Research Center, Bern, Switzerland

[4]Agroscope, Competence Division Methods Development and Analytics, Bern, Switzerland

Correspondence
Reto Portmann, Agroscope, Competence Division Methods Development and Analytics, Schwarzenburgstrasse 161, 3003 Bern, Switzerland.
Email: reto.portmann@agroscope.admin.ch

Funding information
Program Erasmus+/KA1 ISTRUZIONE SUPERIORE, Grant/Award Numbers: 2023-1-IT02-KA131-HED-000131307, CODICE CUP G41I23000110006

Abstract

Background: The identification of pollen is important in the field of beekeeping for the determination of the botanical origin of bee products and investigations of bee diet. Until now, it has been performed by melissopalynology, the microscopic examination of pollen grains. However, this technique has some limitations, such as the necessity of experienced analysts and identification restricted to the family level for some pollen types. Although many techniques have been proposed as alternatives or complements to melissopalynology and omics techniques have been explored to gather information on the botanical origin of honey, no study has yet been conducted on a large set of pollen types.

Results: The study dataset consisted of 34 different pollen types of pellets collected by honeybees in Switzerland and analyzed in multiple biological replications, leading to 150 observations. The pollen samples were analyzed after tryptic digestion using a non-targeted mass spectrometry-based method. Liquid chromatography coupled with mass spectrometry (LC–MS) was employed to identify pollen, and melissopalynology was used as a reference method for the identification. We built an OPLS-DA prediction model for the 34 pollen types. The model clearly identified new samples in their membership group (*Acer* sp., $n = 10$) and a new pollen type at the species-specific level for *Quercus* sp. Less predictable results were achieved for Composita H and pollen collected directly from the plant.

Conclusion: The use of a non-targeted mass spectrometry-based method and chemometrics resulted in a promising tool for pollen identification as a replacement/supplement method to traditional melissopalynology.

KEYWORDS
botanical identification, chemometrics, honeybee, mass spectrometry, melissopalynology, pollen

INTRODUCTION

Melissopalynology determines the botanical origins of honey, pollen, propolis, and royal jelly. It identifies the plants from which eusocial honeybees (*Apis mellifera* L.; *Hymenoptera: Apoidea*) forage for food sources. Melissopalynology is part of a broader field: palynology, the identification of pollen through visual microscopic examination. This technique has a wide application, for example, forensic sciences,[1] medicine,[2] paleoecology,[3] or ecology.[4]

Pollen identification is a crucial concept in the field of beekeeping. This analysis is routinely used to determine the botanical or geographical origins of bee products. In particular, for honey,

melissopalynology might reveal food fraud when pollen types do not correspond to the geographical environment where the honey is produced.[5] Pollen is the protein source for brood and adult bees, and it is also an important source of many other nutrients, such as vitamins and lipids. Thus, the identification of the pollen types collected by honeybees allows for investigating the diet and preferences of honeybees[6–8] and the attractiveness of wild plants and crop cultures.

Pollen is the haploid male gametophyte responsible for the production and transportation of sperm into the ovule.[9] Male gametophyte shapes and dimensions can vary greatly from one species to the other at microscopic examination, and the protein composition is very characteristic for each pollen species as well. This is due to the pollen grain's biological role (avoidance of hybrids, natural selection, biodiversity, etc.) and proteins act as "recognition substances" to avoid interspecies hybridization.[10]

Pollen grains adhere to the body of the honeybee while visiting flowers. The bee then brushes the pollen grains and mixes them with nectar and saliva to make the pollen stick as pollen loads on the specific anatomical points on the hind legs (*Corbiculae*). Throughout the day, honeybees of the same colony typically forage on a variety of plant species resulting in a heterogeneous daily pollen harvest. Pollen collected from different plant species may show a large number of color variations,[11] indicating the diversity of floral resources visited by honeybees. The differently colored pollen pellet groups can be identified by randomly selecting pellets from the various fractions of the harvest to define the botanical origin following the rules of melissopalynology.

Melissopalynology has been the routine analysis for the definition of the botanical origin of pollen for decades, both for scientific investigation and commercial labeling since its creation.[12] Later, the method was harmonized without major changes.[13] Although used for decades, this method has some limitations. The pollen identification is technically based on a visual estimation through microscopic observation,[13] and the success of the analyses depends heavily on the operator's competence and experience. Even proficient analysts may give different interpretations for less-known pollen types.[14] The geographical area (floristic district) where the specialist has been trained also has a strong impact on the precision of the analysis, since a melissopalynologist will know better the pollen types from a familiar flora. Melissopalynology also often leads to different levels of taxonomic identification, which can range from species-specific for unique pollen shapes, as for *Corylus avellana* L., to genus or even general family identification. At best, pollen is identified by the taxonomy of the botanical group (species, genus, or family), depending on how reliable the identification is at the microscopic examination.[15] If the species is not recognizable, a term such as group, form, or type is added to the scientific name to indicate a larger taxonomic group.[16]

In view of the above-mentioned critical aspects of melissopalynological analysis, there is a need for possible innovations to counterpart the traditional identification method. Since the late 1960s, the palynologist community has been interested in the automation of pollen grain recognition.[17] As a result, several methods have been proposed as alternatives to traditional melissopalynology, addressing specific questions such as botanical or geographical origins, particularly for honey analysis. While these methods have often proven valuable, especially in scientific contexts, they have never substituted melissopalynology as the routine analytical method.

In the last 10 years, foodomics techniques have been widely explored to obtain information on the botanical origins of honey,[18] due to the commercial values of especially monofloral honeys, examining many metabolites as volatile organic compounds or polyphenols, as well as with untargeted approaches, but without a particular focus on pollen or pollen sediments contained in honey. Metabolomics and proteomics through high-performance liquid chromatography coupled with mass spectrometry (LC–MS) have proven useful in recent work to study various aspects of pollen. These include the presence of specific allergens[19,20] nutritional properties of pollen[21] or even pollen physiological aspects.[9] However, to our knowledge, no study has yet utilized these techniques for pollen identification.

Our goal was to investigate whether a non-targeted mass spectrometry-based method obtained using state-of-the-art high-resolution mass spectrometry (MS) and chemometrics could be an alternative/complement to melissopalynology. Melissopalynology was initially used as a reference methodology for defining the botanical origin of pollen loads collected by honeybees. The identified pollen types were then analyzed by LC–MS. Finally, a regression model was trained on the measured data to perform some identification trials and test our hypothesis.

MATERIALS AND METHODS

Sample collection and identification through melissopalynological analysis

In recent years, pollen samples have been collected for various projects across Switzerland by the Swiss Bee Research Center at Agroscope. Pollen samples from 2012 to 2014 were collected from April to September at apiaries located in the cantons of Ticino and Basel (Southern and Northern part of Switzerland[22]). More recently, in 2022 and 2023, pollen samples were collected from March to August at apiaries situated in the cantons of Berne and Fribourg (Liebefeld, Bellechasse, and Witzwil, all north of the Alps in Switzerland). All pollen samples were collected from pollen traps installed at the entry of beehives. These samples were divided into several fractions according to the color, dimension, and shape of the pollen pellets in the best possible manner, as previously described.[11] The botanical origin of the pollen pellets in each fraction was determined using melissopalynology. The sorted pollen fractions were stored at −20°C, resulting in a pollen bank. The list of pollen types and biological repetitions that underwent further analytics are presented in Table 1.

In total, from 3 to 6 biological repetitions were prepared for further analyses for each pollen type. When possible, samples were taken from fractions from different place/time or at least from different corbicular pollen pellets of the same fraction. Each pollen pellet

TABLE 1 Pollen samples investigated in this study.

Species	Common name	Bush/ tree/ herb/vine	Family	Samples
Acer sp.	Maple		Sapindaceae	6
Aesculus hippocastanum L.	Horse chestnut		Sapindaceae	4
Brassica napus L.	Canola		Brassicaceae	6
Brassica forma (A)	N.D.		Brassicaceae	3
Brassica forma (B)	N.D.		Brassicaceae	3
Corylus avellana L.	Hazel		Betulaceae	4
Castanea sativa L.	Chestnut		Fagaceae	3
Centaurea cyanus L.	Cornflower		Compositae	4
Centaurea jacea L.	Cornflower		Compositae	4
Convolvulus sp.	Bindweed		Convolvulaceae	3
Cornus sanguinea L.	Dogwood		Cornaceae	4
Crataegus sp.	Hawthorn		Rosaceae	5
Daucus sp.	Carrot (wild)		Apiaceae	3
Echium sp.	Viper bugloss		Boraginaceae	3
Fagus sp.	Beech		Fagaceae	4
Hedera helix L.	Ivy		Araliaceae	5
Helianthus annus L.	Sunflower		Compositae	4
Knautia sp.	Ambrette		Caprifoliaceae	5
Lotus sp.	Birdsfoot trefoil		Fabaceae	3
Onobrichis sp.	Sainfoin		Fabaceae	4
Papaver sp.	Poppy		Papaveraceae	6
Phacelia sp.	Phacelia		Boraginaceae	4
Plantago sp.	Plantain		Plantaginaceae	3
Prunus avium L.	Cherry		Rosaceae	4
Quercus robur L.	Common oak		Fagaceae	3
Quercus ilex L.	Holm oak		Fagaceae	3
Rubus sp. (Rubus forma)	Raspberry/ blackberry		Rosaceae	3
Salix sp.	Willow		Salicaceae	6
Taraxacum sp.	Dandelion		Compositae	3
Tilia sp.	Linden		Malvaceae	4
Trifolium pratense L.	White clover		Fabaceae	3
Trifolium repens L.	Pink clover		Fabaceae	4
Verbascum sp.	Mullein		Scrophulariaceae	3
Zea mays L.	Corn		Poaceae	4

Note: herb (monocotyledon), herb (dicotyledon), bush, tree, vine.

chosen from the fractions was double-checked and identified through melissopalynological analyses to avoid mislabeling, since often pollen pellets from different plants can show the same color or pollen of the same plant species can sometimes show different colors.

To test the developed model based on the statistical elaboration of the data, we added pollen pellets from *Acer* sp. (10 biological repetitions), an unspecified oak (*Quercus* sp.) in 3 biological replications, and an unidentified pollen of the Compositae family (H type; 3 biological replications). We also added pollen collected directly from plants from five different *Corylus avellana* plants.

For microscopic analysis, part of a pellet that was chosen for subsequent high-performance liquid chromatography (HPLC) analysis was diluted in Millipore water without performing acetolysis and fixed in glycerol-gelatin, as previously described.[12] Next, a light microscope (Leica DM 2000 LED) at 10×, 20×, 40×, and 100× magnification was used to determine the external characteristics of the pollen grain (the overall shape and size, the surface sculpture, and the presence of pores, openings, etc.).[23] Where possible, the pollen type was identified at the species level, otherwise the genus or the family, or the pollen type ("forma") comparing the samples with reference preparations of pollen grains of the plants as PalDat[24] or PollenAtlas[25] if necessary.

Samples preparation for HPLC analysis

For each biological repetition, a small portion of the pollen pellet, about a quarter or half the pellet, depending on the pellet size (weight not defined, about 5 mg), was added to 50 µL of 10 mM ammonium bicarbonate and vortexed for 10 s (Vortex-Genie TM® at maximum speed) in a 1.5 mL Eppendorf tube. The solution was heated (as fast as possible) to 95°C while shaking at 1000 rpm for 10 min in an Eppendorf Thermo mixer C. This passage helped to improve the subsequent digestion step and completely inactivate the pollen cells to prevent the formation of living gametophytes, which would rapidly alter the metabolism (and consequently the protein profile) under certain conditions, such as in the presence of water.[26] Then, 1 mL of cold MeOH (−20°C) was added to the solution, which was left at −20°C for 1 hour to precipitate proteins and remove the oils that sometimes cover the upper layer of the pollen. The MeOH solution was centrifuged at 20,000 g for 10 min in an Eppendorf Centrifuge 5804 R. The supernatant was removed, and the pellet was dried by vacuum evaporation at room temperature (15 min) in a refrigerated CentriVap Concentrator LABCONCO. Then, 50 µL of 10 mM ammonium bicarbonate (pH 8) was added, and the solution was mixed well with a vortexer for 30 s. 50 µL of 19 ng/µL trypsin (Trypsin Gold, Promega) was added, and the solution was mixed well again. The solution was incubated overnight at 37°C with shaking at 1000 rpm in the same thermomixer. After centrifugation under the same conditions as before (10 min at 20,000 g), the supernatant was ready for the following LC–MS analysis.

Chromatographic/spectrometric analysis

Methanol and acetonitrile (both LC–MS grade) and formic acid were purchased from Sigma-Aldrich. A 2-µL aliquot of the trypsinized pollen solution was injected into an HPLC (Vanquish, Thermo Fisher Scientific, San Jose, CA, USA) coupled to a Q-Exactive Plus mass spectrometer (Orbitrap®, Thermo Fisher Scientific). The sample was

chromatographically separated on an Acquity C18 column (UPLC HSS T3, 1.8 µm, 2.1 × 150 mm, Waters Corporation, Milford, MA, USA, at 40°C) with a linear gradient of 5% acetonitrile in water to 95% acetonitrile in water (both with 0.1% formic acid) within 15 min and ionized with a heated electrospray ionization source. The mass spectrometer was operated in positive mode. A data-dependent full-scan MS2 acquisition (FS-dd-MS2) with resolving powers of 70,000 and 17,500 for FS and dd-MS2, respectively, was used to measure the parent ions and their fragments. The FS acquisition mode was selected with a scan range of 300–2000 m/z. Fragmentation of precursors was performed with a stepped, normalized collision energy of 10, 25, and 40 eV.

Statistical analyses

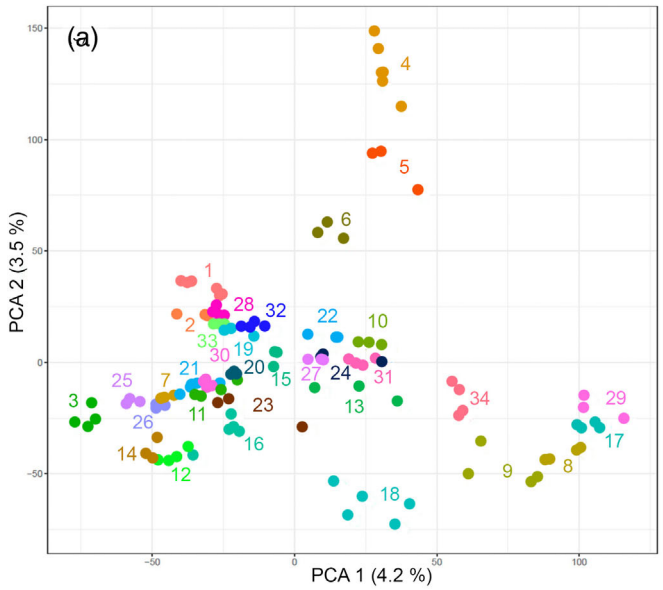
The raw MS data were analyzed in Progenesis QI software (Waters TM® Milford, MA, USA), which allowed retention time alignment of the data, detected the features, and provided the relative abundance of these compounds. The alignment was performed using a mixed pollen sample as reference, composed of 5 µL of the extracts of all pollen types analyzed. Progenesis automatic area normalization (compensation for any differences in injected samples) was used to account for the possible different magnitudes of the measured variables and to prevent differences in overall intensity from affecting subsequent analyses. After peak extraction, the samples were manually grouped based on melissopalynological identification to train the OPLS-DA (Orthogonal partial least squares discriminant analysis) models. The total information resulted in $n = 150$ observations (different samples, considering biological replicates), $n = 34$ pollen types, and $n = 48{,}444$ features (based on retention time and m/z).

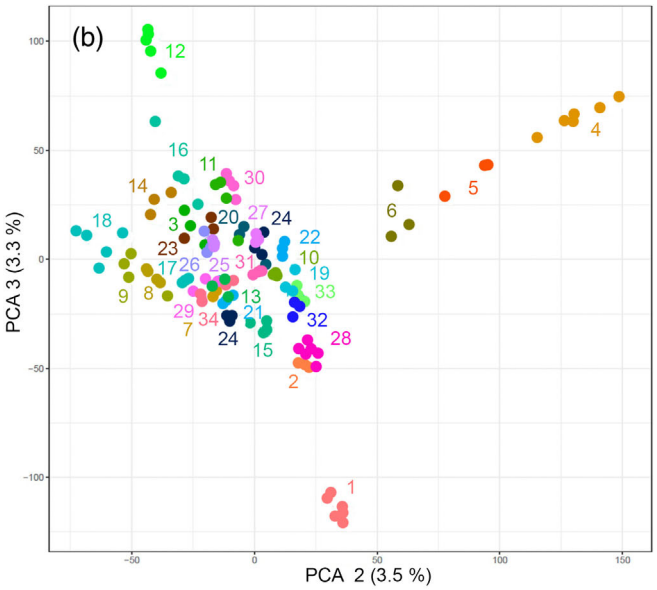
The general relationship between samples was revealed by principal component analysis (PCA). PCA was performed considering the 34 pollen types listed in Table 1. Each pollen type was then compared with all the others through the OPLS-DA statistical model to examine the differences between bee pollen species. The OPLS-DA evaluations showed the R2 and Q2 quality metrics, permutation diagnostics, scores, and orthogonal distances, as well as graphs (scores, loadings, predictions, diagnostics, etc.). The cross-validation segments were set to 7. With the predict option, unknown samples were checked for their affiliation to the individual pollen types using the 34 different models.

The statistical processing of the data was performed through the Package "ropls" in R (PCA and OPLS-DA) for multivariate analysis of the omics data. In total, 150 observations were clustered with the function hclust (dist. method = Euclidean with link = single) from the R package stats.

RESULTS

The total dataset consisted of 34 different pollen types analyzed in three to six biological replications, leading to 150 observations





1 Acer sp.
2 Aesculus hippocastanum L.
3 Corylus avellana L.
4 Brassica napus L.
5 Brassica forma (A)
6 Brassica forma (B)
7 Castanea sativa L.
8 Centaurea cyanus L.
9 Centaurea jacea L.
10 Convolvulus sp.
11 Cornus sanguinea L.
12 Crataegus sp.
13 Daucus sp.
14 Echium sp.
15 Fagus sp.
16. Hedera helix L.
17. Helianthus annuus L.
18. Knautia sp.
19. Lotus sp.
20. Onobrichis sp.
21. Papaver sp.
22. Phacelia sp.
23. Plantago sp.
24. Prunus avium L.
25. Quercus robur L.
26. Quercus ilex L.
27. Rubus forma
28. Salix sp.
29. Taraxacum sp.
30. Tilia sp.
31. Trifolium repens L.
32. Trifolium pratense L.
33. Verbascum sp.
34. Zea mays L.

FIGURE 1 PCA in dimensions 1 and 2 (a) and 2 and 3 (b) of the 150 observations.

(Table 1). The first step considered the total information for a data overview in an unsupervised model, performed using PCA in dimensions 1 and 2 and dimensions 2 and 3 (Figure 1). The first three principal components of the PCA analysis explained about 11% of the total variability. We observed many groups, some of which overlapped, although some samples tended to differentiate, such as the four biological replicates of *C. avellana* pollen in dimensions 2 and 3 or the group of *B. napus* and the two *Brassica* formae, both of which were observed in dimensions 1 and 2 and dimensions 3 and 4.

The PCA displayed the scores of the 150 observations. The result of this first stage was promising, since the biological replicates already showed good clustering in the unsupervised model. Similar subsets of samples/observations (mostly totally coherent with the pollen type) were grouped together to outline the setup of the different models for each pollen type.

A clustering was performed for the 150 observations obtained from the 34 different pollen types in form of a dendrogram (Figure 2). As in the PCA, the biological replicates of the same pollen types generally clustered correctly together (e.g., *B. napus*, *C. cyanus*, or *T. repens*), although some pollen samples did not correctly cluster into the group of biological repetitions (e.g., a sample of *Daucus* sp. or a sample of *Crataegus* sp.). In some cases, the clustering grouped species from the same families or genera together, such as oak (*Q. robur* and *Q. ilex*), but generally did not respect taxonomical groups, as previously noted.[27]

The previous steps through unsupervised models were preliminary to the supervised classification of the data. The results from the PCA informed the second stage of analysis, which was a supervised model that moved from a maximum variance model to a maximum separation model. The training dataset consisted of 34 different pollen types, which were used as a group membership/subset of samples. The horizontal component of the OPLS-DA score scatter plot shows variation between the groups, and the vertical dimension displays variation within the groups. Figure 3 shows five examples of the 34 pollen types analyzed by OPLS-DA. All the OPLS-DA models and the relative metrics are shown in the supplemental data (S1–S34), including the R2 and Q2 quality metrics. The Q2Y estimation of the predictive performance of the models by cross-validation was always positive and over 0.9, apart for *C. sanguinea*, *Daucus* sp., *Echium* sp., *Q. ilex*, and *T. repens*, which remained above 0.8.

This OPLS-DA classification was useful for determining whether the new samples featured in the prediction set were similar to one class or another or did not fit into any class. Therefore, the major target was to define borders around the subsets of pollen types and infer the membership for future samples in the prediction set. Figure 3 shows the pollen types that are relevant in the following model checks: (a) *Acer* sp., (b) *H. annuus*, (c) *C. avellana*, (d) *Q. robur*, and (e) *Q. ilex*.

In the last step, the OPLS-DA model was verified with some additional samples that were not used in the training sets. OPLS-DA uses a binary variable for Y that represents group membership. Predictions have a value of 0 and 1, depending on group membership. We chose the *Acer* sp. subset, for which we tested 10 samples of corbicular pellets. As shown in Figure 4, all 10 samples were correctly identified; thus, our predictive method correctly recognized the pollen type belonging clearly to a group on which the model was trained.

To verify whether our OPLS-DA model was also able to distinguish new pollens apart from the 34 types included in the initial training set, we tested three biological replicates of corbicular pollen of an oak species for which melissopalynological analysis suggested *Q. pubescens* or *Q. petrea* (while the model was trained on the two species, *Q. robur* and *Q. ilex*). According to our results, this new oak sample was not attributed to the two oak species included in the training set (Figure 4), suggesting that our OPLS-DA model could possibly distinguish different pollen types at the species level. Additionally, three unknown pollen pellets of the Composita (H forma) type were included. The results were less clear in this second case of a pollen type on which the model was not trained, since the three samples of Compositae H were identified as *H. annuus* (i.e., the pollen type more similar to the new one) but also as *Lotus* sp. and *T. repens* and one of them as *Daucus* sp. and *Papaver* sp. (Figure 4).

Finally, we tested pollen samples collected directly from five *Corylus avellana* plants to determine whether they fit the *C. avellana* group from corbicular pollen in the prediction model. All the samples of *C. avellana* collected from the plant correctly fit in the *C. avellana* set, but three of them also fit in the horse chestnut (*A. hippocastanum*) (Figure 4).

DISCUSSION

Our study demonstrates that a non-targeted mass spectrometry-based LC–MS method after tryptic digestion could be a suitable tool for pollen identification. Many samples can be analyzed in a single run, resulting in large datasets. Our procedure showed promise for integrating the identification of pollen by the traditional microscopic technique, sometimes leading to even more accurate information, such as species-specific identification or highlighting the difference between pollens of the same "forma."

The rationale for the extraction method and the biological particularities of the matrix analyzed

In this study, we decided to focus on proteins/peptides, the main components of pollen,[28] and to exclude other components from the non-targeted analysis by the extraction method. Penetration of the pollen wall with digestive enzymes is one of the strategies used to overcome the resistance of the very robust outer layer, which otherwise must be destroyed.[29] The pollen content can thus be extracted without having to break open the very hard exine layer. Chemical analyses using tryptic digestion and LC–MS were previously proposed as an advanced method in previous trials for identification with methods alternative to microscopy.[30]

Pollen grains share the same reproductive function but are morphologically and physiologically very different.[31] As previously mentioned, pollen is responsible for fecundation in the vegetal kingdom[9] and is therefore subject to great variability in terms of appearance (shapes and dimensions) and composition, especially with respect to proteins. A large proportion of the proteins located in the inner cellulose layer of the pollen grain wall (intine) are "recognition substances" to avoid interspecific hybridization. Among them are those with

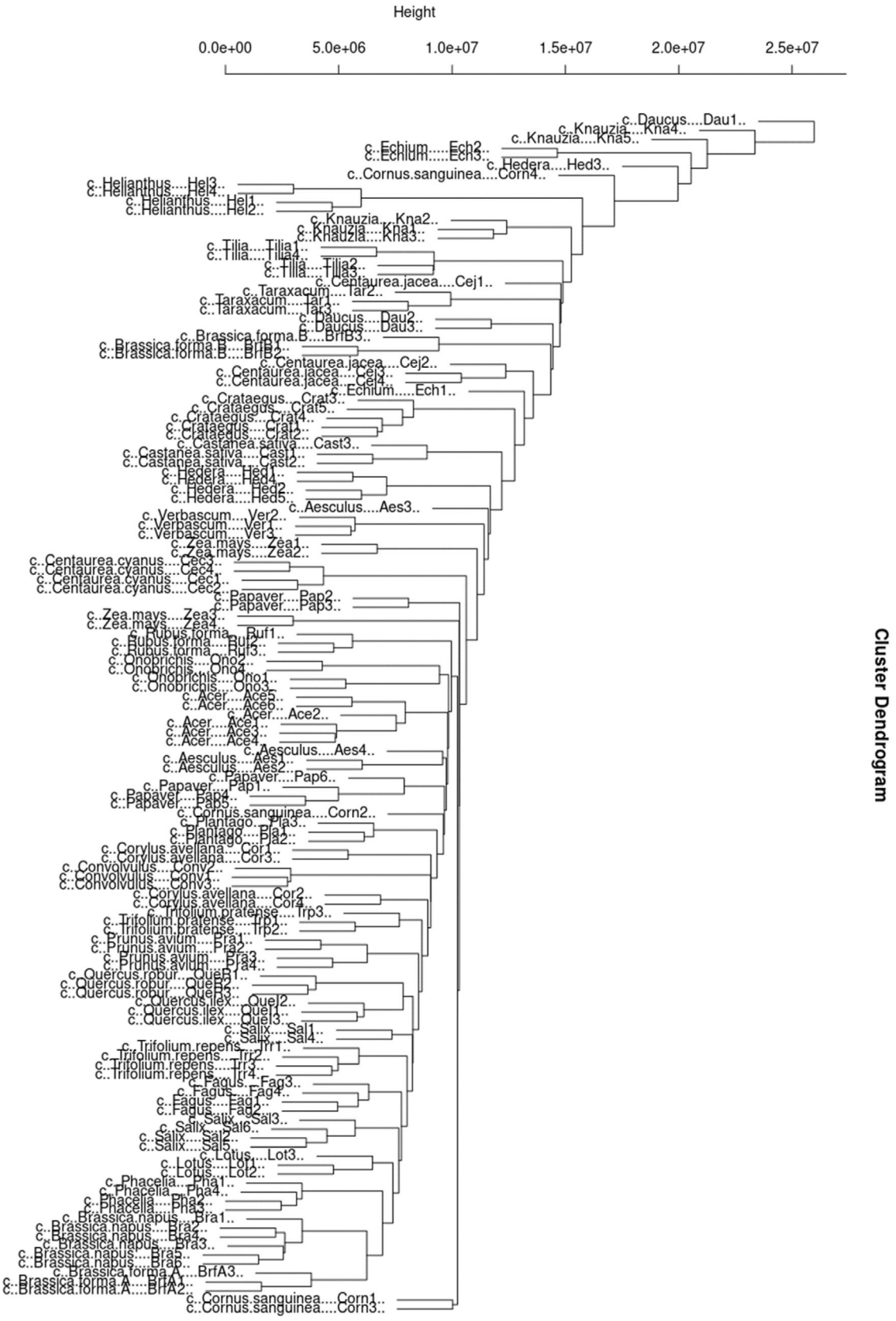


FIGURE 2 Dendrogram of the 150 biological replicates. The samples are labeled with the identification name obtained through microscopic examination.

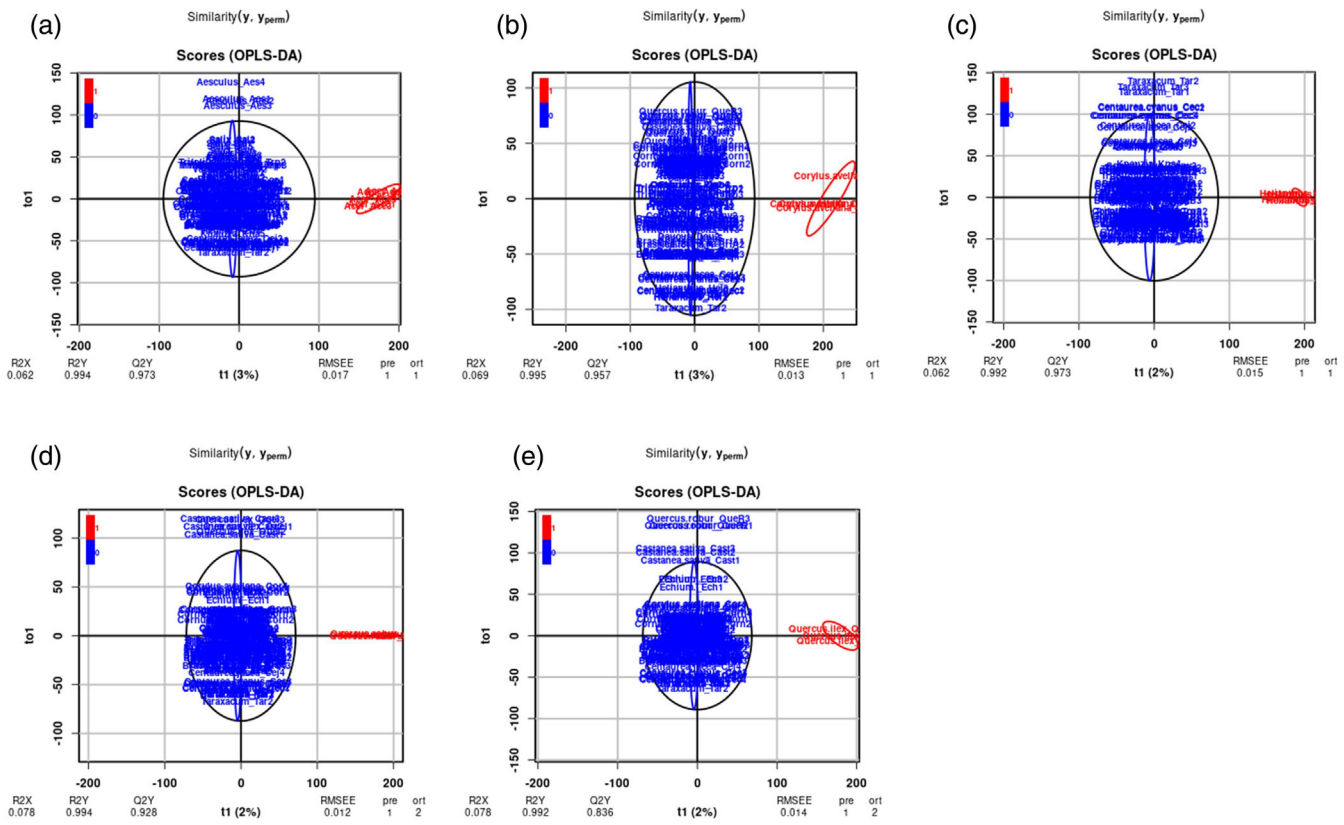


FIGURE 3 OPLS-DA model for the following pollen types: (a) *Acer* sp., (b) *C. avellana*, (c) *H. annuus*, (d) *Q. robur*, (e) *Q. ilex*. The figures show the x-score plot for each model: The number of components and the cumulative R2X, R2Y, and Q2Y are indicated below the plot.

antigenic and allergenic properties.[10] Pollen–pistil interactions are important prezygotic reproductive barriers that play a critical role in mate selection in plants and involve complex physical as well as molecular and biochemical mechanisms.[32]

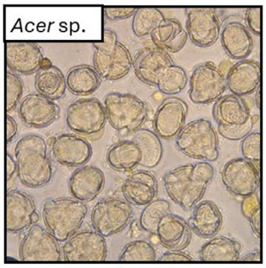
Pollen is composed of pre-synthesized proteins required for pollen germination, but proteins are also synthesized by the gametophytic cell for pollen tube growth and germination.[26] Pollen is activated by special conditions such as hydration or the presence of sugars and other substances that indicate to the gametophyte that it has completed its journey to the female reproductive organs. As this activation process occurs very easily and quickly in some species,[26] inactivation of this cascade with heat is necessary. Protein composition can be considered one of the most important discriminators of pollen diversity. Mature pollen is composed of all the proteins necessary for its functional specialization—that is, those involved in early fertilization events, including hydration, cohesion, pollen tube formation, polarity, and cell recognition of pollen-stigma—and the initiation of a hierarchical signaling flow.[33]

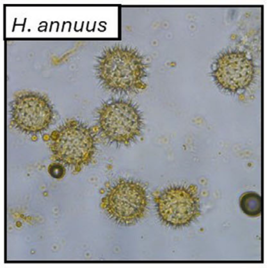
The promising results of the first step of data elaboration with an unsupervised model

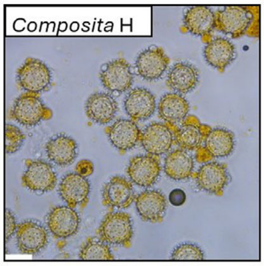
Our first step was to obtain an overall view of the groups that could cluster together. PCA (Figure 1) showed very good grouping of the samples, even before using specific models to assess the differences between each pollen and all the others. Indeed, the biological replicates of the same pollen types clustered clearly in the graph. PCA, an unsupervised model, is one of the first methods used in metabolomics research to observe whether there is a clustering trend without setting up specific criteria.[34] In our case, PCA was a useful tool to obtain a clear representation, starting from a matrix of experimental data that included dozens of samples (the pollen types and all their biological replicates) and thousands of variables. The overlay, as well as the clear clustering of samples of the same pollen type in the PCA, showed that the different pollen types share some common features and were strongly differentiated by many features, which is consistent with the biological rationale described previously.

Representation of relationships between datasets

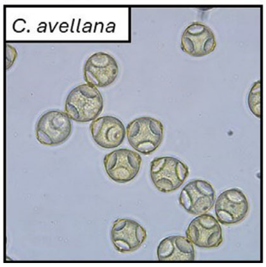
Almost all the different species (pollen types) were correctly clustered in the dendrogram (Figure 2). In some cases, as with the two species of oak, the genus was also correctly grouped. However, the separation of families was not easily obtained with this approach. This outcome is in line with previous studies.[27] We find that some plants of the same family correctly clustered, as was the case for the Brassicaceae group (the crop species *Brassica napus* and the *Brassica* forma A), while the grouping was less precise for, for example, Fabaceae, where

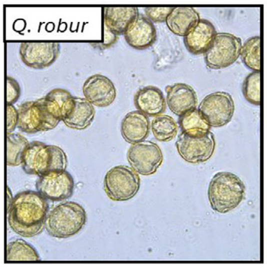


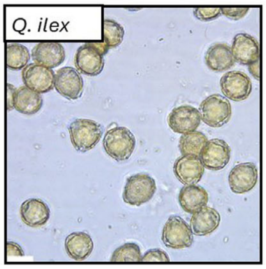




Identification of unknown samples with OPLS-DA model	**Acer sp.**	A. hippocastanum	**C. avellana**	B. napus	Brassica forma (A)	Brassica forma (B)	C. sativa	C. cyanus	C. jacea	Convolvulus sp.	C. sanguinea	Crataegus sp.	Daucus sp.	Echium sp.	Fagus sp.	H. helix	**H. annuus**	Knauzia sp.	Lotus sp.	Onobrichis sp.	Papaver sp.	Phacelia sp.	Plantago sp.	P. avium	**Q. robur**	**Q. ilex**	Rubus forma	Salix sp.	Taraxacum sp.	Tilia sp.	T. repens	T. pratense	Verbascum sp.	Z. mays
Acer sp.	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Acer sp.	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Acer sp.	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Acer sp.	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Acer sp.	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Acer sp.	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Acer sp.	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Acer sp.	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Acer sp.	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Acer sp.	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Composita H	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Composita H	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Composita H	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0
C. avellana	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C. avellana	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C. avellana	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C. avellana	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C. avellana	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Quercus sp.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Quercus sp.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Quercus sp.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0







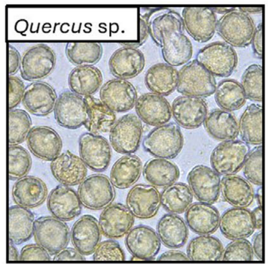


FIGURE 4 Identification of new samples with OPLS-DA models. OPLS-DA predictions use a value of 1 if it attributes a sample a group membership. The figure also shows a microscopic image (40× magnification) of the pollen types considered. Microscopic images of *H. annuus* and *Q. robur* and *Q. ilex* are shown for comparison with the Composita H pollen and the third *Quercus* species.

T. pratense, *T. repens*, and *Lotus* sp. were quite far apart, and for Compositae, where we find no similarities between *T. officinalis*, *H. annuus*, and the two *Centaurea* species. Some discrepancies could also be due to mislabeling at the microscopic step (as for *Brassica* forma B) or to analytical/sample conservation issues (as for *C. sanguinea* pollen from 2013). However, it is not necessarily expected to see similarities among members of different families, since plants could be related at the taxonomic level but have pollens with very different morphology[35] and, consequently, distinct metabolomes for their biological role of species-specific gametophytes.

Separation of individual pollen species in a supervised model and identification of new pollen analyses

Using the data from all pollen analyses as a training set, OPLS-DA (which is a supervised model, unlike PCA) was created (Figure 3). This made it possible to distinguish each pollen type from the others. An identification trial using the model was tested with additional new pollen samples (Figure 4). These identification tests yielded very promising results. The model was able to assign all 10 biological replicates of the "unknown" *Acer* sp. pollen samples to the correct group.

The results of the three samples of "unknown oak" pollen (*Quercus* sp.) are very interesting. In this case, the experience of the melissopalynologist was crucial and confirmed by our analytical method. The microscopic analysis allowed for the distinction of the *Q. robur* pollen from the *Q. ilex*, and for the three samples used for testing the OPLS-DA model, a third species (*Q. pubescens* or, more likely, *Q. petraea*) was proposed. This hypothesis was confirmed by the OPLS-DA model, which was trained on Q. *robur* and Q. *ilex*, and did not assign these last samples to either of the two species (Figure 4). However, as previously mentioned, identification through microscopic examination requires expert knowledge[36] and is therefore tied to the analyst's experience. Another melissopalynologist with less experience

might have simply assigned the samples to the Quercus group.[23] Usually, oak is determined as either Quercus deciduous or Quercus evergreen types, although some very experienced melissopalynologists can define the species.[37] In this case, our approach resulted not only in a possible substitution but also in a complementary technique to microscopic analysis and might help the traditional technique in discriminating at a more precise taxonomic level. The chemical variations in the pollen of oak species (*Quercus* sp.) were nevertheless found to be an effective tool for taxonomic identification in previous research, confirming our results. Fourier-transform infrared spectroscopy and chemometrics may return a more specific identification compared with melissopalynology, even in subfossil pollens.[37]

In the case of an unknown Compositae, however, the results were less clear. The identification of Compositae is also very challenging at the microscopic level, and Compositae cannot be precisely defined melissopalinologically. The taxonomic group of Compositae is mostly divided into "formae" A, C, H, J, S, and T based on the dimensions and especially the esine (the outer layer of the pollen cell) structure, without species-specific identification being possible in the absence of additional information, such as the time of the year when the pollen was collected. In our case, we assigned microscopically some pollen samples to Compositae forma H, knowing that the pollen was probably *Helianthus annuus* L. because the pollen was collected at the time and location where sunflowers were blooming. In the absence of information regarding the collection area, the identification of *H. annuus* at the species-specific level would have been difficult since Compositae H are all very similar when assessed visually, as we can also see in Figure 4. Our OPLS-DA model identified the unknown Compositae samples as *H. annuus* but also as other species that do not belong to the Compositae family. As shown in the dendrogram results and the OPLS-DA, the Compositae showed fewer similarities among themselves, which could be related to their high variability.

Some pollen samples were collected directly from the inflorescences (*amenti*) of *C. avellena*, in addition to the pollen loads collected by bees. All five biological replicates matched hazelnuts, but three also showed a match with horse chestnuts. There are several possible explanations. Differences could be due to salivary substances or nectar,[38] which the bees use to bundle the collected pollen grains into pollen loads.[28] Nectar also often contains pollen due to the primary enrichment in the flower organs, a concept that is at the base of honey floristic origin analysis.[39] Horse chestnut and hazelnut blooming are quite close to each other and often overlap, so the result could also be explained by some contamination coming from the nectar or from the bee body in the pollen loads. The substances added from the honeybee to the pollen collected in the early stages of foraging activity are still poorly known,[38] and this could be a further interesting application of foodomics. This result must be considered in future research, for example, in the search for markers for different pollen types. The direct harvesting of pollen from the plant could also allow for a more taxonomically precise identification of pollens that are not recognizable by microscopic examination, such as the Compositae previously mentioned. This could be achieved at the least for the agronomically or beekeeping important species. It is advisable to always compare pollen collected directly from plants in future metabolomics studies. Alternatively, if corbicular pollen is used, compounds present in high quantities and not only traces should be considered when researching markers.

Possible future applications of the method

A correct and more in-depth identification of pollen would be useful in many areas of beekeeping.

The study of honeybee and wild bee foraging behavior is important to counteract the decline in bee populations, which is due to many factors, including the lack of foraging resources for domestic and wild bees.[40,41] To properly propagate these resources, it is valuable to know exactly which species are more attractive, as species within the same genus can have very different target pollinators,[42] and pollen analysis is an essential part of bee ecology research.[43–45]

The application of this technique would also be interesting in the field of food fraud detection, which is still mainly conducted through melissopalynology. Microscopy-based pollen identification methods have been widely applied to determine the geographical provenance and floristic origin of bee products[46] and as a quality control protocol, particularly in the European Union (EU Directive Council, 2001). The identification of species-specific pollen is important in some cases to distinguish the geographical origin of pollen and honey, since some species that are indistinguishable by microscopic examination grow in very different environments, such as *Rubus fruticosus* L. and *Rubus idaeus* L.,[47–49] both reported as *Rubus* forma.

An analytical method complementary to melissopalynology would make the detection of plants contaminated with pesticides more precise, thereby improving the protection of pollinators. Additionally, it may facilitate the differentiation of pollen from plants containing natural contaminants (secondary metabolites) that may be harmful to human health, such as alkaloids. Pyrrolizidine alkaloids, for example, are present in plants belonging to the Compositae family, including *Senecio* spp., *Eupatorium* spp., *Petasites* spp., and *Tussilago farfara* L.[50] As indicated earlier, however, the Compositae family is subdivided into "formae" using the microscopic method, without allowing species-specific identification. A different analytical approach could perhaps overcome this limitation.

Comparison of our study with other methods proposed as alternatives to melissopalynology

Given the important role that pollen identification plays in many scientific questions, several other methods have been studied and proposed as substitute/complement to the traditional melissopalynology. Chemical fingerprints of elements[51] or organic compounds[52,53] have been proposed as markers for botanical identification in honey. In pollen, encouraging results have been obtained with Fourier-transform infrared spectroscopy[37,54] in some cases coupled with multivariate analysis.[16] The MALDI-TOF technique proved useful when

considering lipids[55] or using the full spectral range[27,30,56,57] and focused mainly on the identification of airborne pollen in mixtures.[30,56,57] More recently, visible spectroscopy has proven to be a useful method to effectively identify the botanical origin of honey,[58] as has Raman spectroscopy coupled with chemometric methods.[58] Recent trends are moving toward artificial intelligence based on visual estimation, basically reproducing melissopalynological analysis[18] and DNA metabarcoding.[8]

Visual AI estimation was particularly valuable for airborne pollen,[18,59–61] whereas only a few studies have concentrated on pollen of beekeeping interest. In other cases, such as pollen of paleoclimate and paleo-vegetation interest, identification ended at the family level,[3] although these methods were useful for identifying fossil-broken or incomplete pollen grains. Artificial intelligence was useful to discriminate 35 pollen types of a given botanical region, including some species of beekeeping interest, such as *T. repens* or *C. jacea*, but many images were needed for accurate identification, as misclassifications could occur due to the different orientation of the pollen on the images.[62]

Regarding molecular techniques, DNA analysis and metabarcoding have been shown to be promising techniques for identification, sometimes being more taxonomically accurate than melissopalynology, but very often not providing quantitative information.[7] Estimating the proportion of taxa present in a sample is challenging using expensive DNA metabarcoding[63,64] due to DNA extraction or PCR biases.[65,66] Biochemical analyzes (e.g., by tryptic digestion) and the use of LC–MS/MS techniques could in the future allow not only the identification but also a more detailed analysis of other molecular classifiers (e.g., glycans, lipids, peptides, and proteins) of the pollen set studied.[30]

Non-targeted spectrometry-based methods are a modern "omics" approach that provides a large amount of data on the molecules found in a matrix with high sensitivity and thus can be used to simultaneously detect a wide range of metabolites. To our knowledge, the Orbitrap technology and untargeted metabolomic approach have never been tested for pollen identification, although there is an increasing use of this technology to study the composition of pollen, for example, for the presence of contaminants,[67] allergens,[19] metabolites conferring particular bioactivity to this food,[21] or even for plant physiology and biology.[9,33] Regarding the latter topic, the molecular mechanism underlying the process of pollination is poorly understood due to the lack of information on pollen proteins,[68] and it would be advisable to use highly sensitive and up-to-date technologies, such as LC–MS/MS more frequently. Although often proposed as a superfood,[69] pollen digestion is still poorly known.

The limitations of our method, similar to the molecular (DNA metabarcoding) approach, are based on the need for a complete plant reference database and a more precise taxonomic resolution of some taxa, as it is for Compositae. A first attempt was made during our research to use protein databases, such as SwissProt and Trembl, which were feasible for plants of major agricultural or scientific interest, such as Canola, but these databases were extremely incomplete for other flowers and wild plants. Thus, we propose metabolomics' great potential as an authentication method for pollen as research moves toward identifying specific and multiple markers for each pollen type, which can be employed in a mixture of pollen from wild bees or pollen residues isolated from honey.

CONCLUSION

In our study, the use of a non-targeted mass spectrometry-based method and chemometrics was proposed as a promising tool for pollen identification, as substitute/complement method to traditional melissopalynology. Some identification experiments performed after training the model correctly clustered the samples, even at the specific level for the genus *Quercus*, and assigned the "unknown" pollen to the correct group (*Acer* sp.). The development of a database directly from plant pollen, overcoming some limitations of melissopalynology, would probably improve the taxonomic identification for some pollen types that are of great interest for beekeeping (i.e., bee diet or food fraud) and thus far mainly distinguished at the family level. It would also be useful to identify several markers for each pollen type, particularly in mixed samples, and to use them for quantification. As with other techniques, such as metabarcoding, the main limitation is that a pollen database is required for identification.

ACKNOWLEDGMENTS

We wish to thank Carla Gianoncelli of Fondazione Fojanini (Sondrio, Italy) for supervising the pollen identification of this study and the PhD School of Environmental Sciences of the University of Milano for the support.

FUNDING INFORMATION

The scholarship of the first author was supported by Program Erasmus+/KA1 ISTRUZIONE SUPERIORE, Convenzione n. 2023-1-IT02-KA131-HED-000131307, CODICE CUP G41I23000110006, a.a. 2023/2024.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest. The funders had no role in the design of the study, in the collection, analysis, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

DATA AVAILABILITY STATEMENT

All data generated or analyzed during this study are available upon reasonable request from the corresponding author.

ORCID

Valeria Leoni https://orcid.org/0000-0002-1468-2095
Christina Kast https://orcid.org/0000-0002-5673-6848
Ueli Bütikofer https://orcid.org/0000-0003-2261-7282
Luca Giupponi https://orcid.org/0000-0001-9618-1114
Reto Portmann https://orcid.org/0000-0001-8979-7439

REFERENCES

1. Ezegbogu MO. Identifying the scene of a crime through pollen analysis. Sci Justice. 2021;61:205–13.
2. Liu A, Sheng W, Tang X. Atmospheric pollen concentrations and chronic obstructive pulmonary disease (COPD) patients visits in Beijing: time series analysis using a generalized additive model. Sci Rep. 2024;14:3462.
3. Bourel B, Marchant R, de Garidel-Thoron T, Tetard M, Barboni D, Gally Y, et al. Automated recognition by multiple convolutional neural networks of modern, fossil, intact and damaged pollen grains. Comput Geosci. 2020;140:104498.
4. Bertrand C, Eckerter PW, Ammann L, Entling MH, Gobet E, Herzog F, et al. Seasonal shifts and complementary use of pollen sources by two bees, a lacewing and a ladybeetle species in European agricultural landscapes. J App Ecol. 2019;56:2431–42.
5. Martin P. Importance of melissopalynology for beekeeping and trade. Bee World. 2005;86:75–6.
6. De Vere N, Jones LE, Gilmore T, Moscrop J, Lowe A, Smith D, et al. Using DNA metabarcoding to investigate honeybee foraging reveals limited flower use despite high floral availability. Sci Rep. 2017;7(1):42838.
7. Hawkins J, De Vere N, Griffith A, Ford CR, Allainguillaume J, Hegarty MJ, et al. Using DNA metabarcoding to identify the floral composition of honey: a new tool for investigating honeybee foraging preferences. PLoS One. 2015;10(8):e0134735.
8. Leponiemi M, Freitak D, Moreno-Torres M, Pferschy-Wenzig EM, Becker-Scarpitta A, Tiusanen M, et al. Honeybees' foraging choices for nectar and pollen revealed by DNA metabarcoding. Sci Rep. 2023;13:14753.
9. Wang J, Kambhampati S, Allen DK, Chen LQ. Comparative metabolic analysis reveals a metabolic switch in mature, hydrated, and germinated pollen in Arabidopsis thaliana. Front Plant Sci. 2022;13:836665.
10. Knox RB, Willing RR, Ashford AE. Role of pollen-wall proteins as recognition substances in interspecific incompatibility in poplars. Nature. 1972;237:381–3.
11. Campos MG, Anjos O, Chica M, Campoy P, Nozkova J, Almaraz-Abarca N, et al. Standard methods for pollen research. J Apic Res. 2021;60:1–109.
12. Louveaux J, Maurizio A. Vorwohl G and methods of melissopalynology. Bee World. 1978;59:139–57.
13. Von Der Ohe W, Persano Oddo L, Piana ML, Morlot M, Martin P. Harmonized methods of melissopalynology. Apidologie. 2004;35:S18–25.
14. Airy H. Pollen-grains in the air. Nature. 1874;10:355.
15. Persano Oddo L, Ricciardelli d'Albore G. Nomenclatura melissopalinologica. Apicoltura. 1989;5:63–72.
16. Corvucci F, Nobili L, Melucci D, Grillenzoni FV. The discrimination of honey origin using melissopalynology and Raman spectroscopy techniques coupled with multivariate analysis. Food Chem. 2015;169:297–304.
17. Flenley JR. The problem of pollen recognition. In: Clowes MB, Penny JP, editors. Problems of picture interpretation. CSIRO, Canberra, ACT: Australia; 1968. p. 141–5.
18. Li N, Song M, Li H, Liu Z, Jiang A, Lang Y, et al. Advancement of foodomics techniques for honey botanical origins authentication: past decade (2013–2023) and future perspectives. Trends Food Sci Technol. 2024;147:104458.
19. Tao Y, Yin S, Fu L, Wang M, Meng L, Li F, et al. Identification of allergens and allergen hydrolysates by proteomics and metabolomics: a comparative study of natural and enzymolytic bee pollen. Food Res Int. 2022;158:111572.
20. Matuszewska E, Plewa S, Pietkiewicz D, Kossakowski K, Matysiak J, Rosiński G, et al. Mass spectrometry-based identification of bioactive bee pollen proteins: evaluation of allergy risk after bee pollen supplementation. Molecules. 2022;27(22):7733.
21. Zhang X, Yu M, Zhu X, Liu R, Lu Q. Metabolomics reveals that phenolamides are the main chemical components contributing to the anti-tyrosinase activity of bee pollen. Food Chem. 2022;389:133071.
22. Kast C, Kilchenmann V, Reinhard H, Bieri K, Zoller O. Pyrrolizidine alkaloids: the botanical origin of pollen collected during the flowering period of echium vulgare and the stability of pyrrolizidine alkaloids in bee bread. Molecules. 2019;24:2214.
23. Palmieri N, Grillenzoni FV, Corvucci F, Biondi C, Bedini G, Floris I, editors. Guida allo studio della Melissopalinologia. Sassary (Italy): Tipografia Gallizzi; 2017.
24. PalDat, palynological database. [cited 2024 May 14]. https://www.paldat.org/
25. Pollen Atlas, palynological database. [cited 2024 May 14]. https://pollenatlas.net/
26. Fíla J, Radau S, Matros A, Hartmann A, Scholz U, Fecikova J, et al. Phosphoproteomics profiling of tobacco mature pollen and pollen activated in vitro. Mol Cell Proteomics. 2016;15:1338–50.
27. Seifert S, Weidner SM, Panne U, Kneipp J. Taxonomic relationships of pollens from matrix-assisted laser desorption/ionization time-of-flight mass spectrometry data using multivariate statistics. Rapid Commun Mass Spectrom. 2015;29:1145–54.
28. Oroian M, Dranca F, Ursachi F. Characterization of Romanian bee pollen—an important nutritional source. Foods. 2022;11:2633.
29. Wu W, Qiao J, Xiao X, Kong L, Dong J, Zhang H. In vitro and in vivo digestion comparison of bee pollen with or without wall-disruption. J Sci Food Agric. 2021;101:2744–55.
30. Lauer F, Diehn S, Seifert S, Kneipp J, Sauerland V, Barahona C, et al. Multivariate analysis of MALDI imaging mass spectrometry data of mixtures of single pollen grains. J Am Soc Mass Spectrom. 2018;29:2237–47.
31. Pacini E, Franchi GG. Pollen biodiversity – why are pollen grains different despite having the same function? A review. Bot J Linn Soc. 2020;193:141–64.
32. Broz AK, Bedinger PA. Pollen-pistil interactions as reproductive barriers. Annu Rev Plant Biol. 2021;72:615–39.
33. Dai S, Li L, Chen T, Chong K, Xue Y, Wang T. Proteomic analyses of Oryza sativa mature pollen reveal novel proteins associated with pollen germination and tube growth. Proteomics. 2006;6:2504–29.
34. Li R, Sun Z, Zhao Y, Li L, Yang X, Cen J, et al. Application of UHPLC-Q-TOF-MS/MS metabolomics approach to investigate the taste and nutrition changes in tilapia fillets treated with different thermal processing methods. Food Chem. 2021;356:129737.
35. Lee S, Heo KI, Cho J, Lee C, Chen W, Kim SC. New insights into pollen morphology and its implications in the phylogeny of *Sanguisorba* L. (Rosaceae; Sanguisorbeae). Pl Syst Evol. 2011;291:227–42.
36. Keller A, Danner N, Grimmer G, Ankenbrand M, von der Ohe K, von der Ohe W, et al. Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. Plant Biol. 2015;17:558–66.
37. Muthreich F, Zimmermann B, Birks HJB, Vila-Viçosa CM, Seddon AWR. Chemical variations in Quercus pollen as a tool for taxonomic identification: implications for long-term ecological and biogeographical research. J Biogeogr. 2020;47:1298–309.
38. Roulston TH, Cane JH. Pollen nutritional content and digestibility for animals. Plant Syst Evol. 2000;222:187–209.
39. Piana L. *Monografie: Analisi Pollinica e Mieli Uniflorali*. [Online]. Piana Consulenza e Ricerca (2016). [cited 2024 May 14]. 2016 https://www.pianaricerca.it/
40. Goulson D, Nicholls E, Botías C, Rotheray EL. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. Science. 2015;347:1255957.

41. Naug D. Nutritional stress due to habitat loss may explain recent honeybee colony collapses. Biol Conserv. 2009;142:2369–72.
42. Giupponi L, Leoni V, Gianoncelli C, Tamburini A, Giorgi A. Endemic plants can be resources for mountain agro-ecosystems: the case of *Sanguisorba dodecandra* Moretti. Sustainability. 2022;14:6825.
43. Lau P, Grebenok RJ, Rangel J, Behmer ST. Assessing pollen nutrient content: a unifying approach for the study of bee nutritional ecology. Philos Trans R Soc B. 2022;377(1853):20210510.
44. Danner N, Molitor AM, Schiele S, Härtel S, Steffan-Dewenter I. Season and landscape composition affect pollen foraging distances and habitat use of honeybees. Ecol Appl. 2016;26:1920–9.
45. Vaudo AD, Patch HM, Mortensen DA, Tooker JF, Grozinger CM. Macronutrient ratios in pollen shape bumble bee (*Bombus impatiens*) foraging strategies and floral preferences. Proc Natl Acad Sci. 2016;113(28):E4035–42.
46. Aronne G, De micco V. Traditional melissopalynology integrated by multivariate analysis and sampling methods to improve botanical and geographical characterisation of honeys. Plant Biosyst. 2010;144:833–40.
47. Leoni V, Panseri S, Giupponi L, Pavlovic R, Gianoncelli C, Sala S, et al. Formal analyses are fundamental for the definition of honey, a product representing specific territories and their changes: the case of north Tyrrhenian dunes (Italy). Sci Rep. 2023;13:17542.
48. Leoni V, Panseri S, Giupponi L, Pavlovic R, Gianoncelli C, Coatti G, et al. Phytochemical profiling of red raspberry (*Rubus idaeus* L.) honey and investigation of compounds related to its pollen occurrence. J Sci Food Agric. 2024;104:5391–406. https://doi.org/10.1002/jsfa.13375
49. Leoni V, Giupponi L, Pavlovic R, Gianoncelli C, Cecati F, Ranzato E, et al. Multidisciplinary analysis of Italian alpine wildflower honey reveals criticalities, diversity and value. Sci Rep. 2021;11:19316.
50. Kast C, Dübecke A, Kilchenmann V, Bieri K, Böhlen M, Zoller O, et al. Analysis of swiss honeys for pyrrolizidine alkaloids. J Apic Res. 2014;53:75–83.
51. Bogdanov S, Haldimann M, Luginbühl W, Gallmann P. Minerals in honey: environmental, geographical and botanical aspects. J Apic Res. 2007;46:269–75.
52. Kaškoniene V, Venskutonis PR. Floral markers in honey of various botanical and geographic origins: a review. Compr Rev Food Sci Food Saf. 2010;9:620–34.
53. Borgonovo F, Marzorati S, Piana L, Rizzi R, Panseri S, Verotta L, et al. Electronic nose for the classification of honeys of different floral origins. Chem Eng Trans. 2021;85:205–10.
54. Jardine PE, Gosling WD, Lomax BH, Julier ACM, Fraser WT. Chemotaxonomy of domesticated grasses: a pathway to understanding the origins of agriculture. J Micropalaeontol. 2019;38:83–95.
55. Liang M, Zhang P, Shu X, Liu C, Shu J. Characterization of pollen by MALDI-TOF lipid profiling. Int J Mass Spectrom. 2013;334:13–8.
56. Weidner S, Schultze RD, Enthaler B. Matrix-assisted laser desorption/ionization imaging mass spectrometry of pollen grains and their mixtures. Rapid Commun Mass Spectrom. 2013;27:896–903.
57. Krause B, Seifert S, Panne U, Kneipp J, Weidner SM. Matrix-assisted laser desorption/ionization mass spectrometric investigation of pollen and their classification by multivariate statistics. Rapid Commun Mass Spectrom. 2012;26:1032–8.
58. Lorenc Z, Paśko S, Pakuła A, Teper D, Sałbut L. An attempt to classify the botanical origin of honey using visible spectroscopy. J Sci Food Agric. 2021;101:5272–7.
59. Oteros J, Weber A, Kutzora S, Rojo J, Heinze S, Herr C, et al. An operational robotic pollen monitoring network based on automatic image recognition. Environ Res. 2020;191:110031.
60. Oconnor DJ, Healy DA, Hellebust S, Buters JTM, Sodeau JR. Using the WIBS-4 (waveband integrated bioaerosol sensor) technique for the on-line detection of pollen grains. Aerosol Sci Tech. 2014;48:341–9.
61. Polling M, Li C, Cao L, Verbeek F, de Weger LA, Belmonte J, et al. Neural networks for increased accuracy of allergenic pollen monitoring. Sci Rep. 2021;11:11357.
62. Dunker S, Motivans E, Rakosy D, Boho D, Mäder P, Hornick T, et al. Pollen analysis using multispectral imaging flow cytometry and deep learning. New Phytol. 2021;229:593–606.
63. Baksay S, Pornon A, Burrus M, Mariette J, Andalo C, Escaravage N. Experimental quantification of pollen with DNA metabarcoding using ITS1 and trnL. Sci Rep. 2020;10:4202.
64. Piñol J, Senar MA, Symondson WOC. The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. Mol Ecol. 2019;28:407–19.
65. Bell KL, Burgess KS, Okamoto KC, Aranda R, Brosi BJ. Review and future prospects for DNA barcoding methods in forensic palynology. Forensic Sci Int Genet. 2016;21:110–6.
66. Pornon A, Escaravage N, Burrus M, Holota H, Khimoun A, Mariette J, et al. Using metabarcoding to reveal and quantify plant-pollinator interactions. Sci Rep. 2016;6:27282.
67. Romera-Torres A, Romero-González R, Martínez Vidal JL, Garrido Frenich A. Comprehensive tropane alkaloids analysis and retrospective screening of contaminants in honey samples using liquid chromatography-high resolution mass spectrometry (orbitrap). Food Res Int. 2020;133:109130.
68. McCormick S. Control of male gametophyte development. Plant Cell. 2004;16:S142–53.
69. Denisow B, Denisow-Pietrzyk M. Biological and therapeutic properties of bee pollen: a review. J Sci Food Agric. 2016;96:4303–9.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Leoni V, Kast C, Bütikofer U, Giupponi L, Portmann R. Identification of pollen types of beekeeping interest by non-targeted mass spectrometry. JSFA Reports. 2025;5(2):62–73. https://doi.org/10.1002/jsf2.228