Food Research International 194 (2024) 114872

Contents lists available at ScienceDirect

Food Research International

journal homepage: www.elsevier.com/locate/foodres

High-resolution proteomics and machine-learning identify protein classifiers of honey made by Sicilian black honeybees (*Apis mellifera* ssp. *sicula*)

Giulia Biundo [a,1], Matteo Calligaris [a,b,1], Margot Lo Pinto [a], Danilo D'apolito [c], Salvatore Pasqua [c], Giulio Vitale [d], Giuseppe Gallo [e], Antonio Palumbo Piccionello [f], Simone D. Scilabra [a,*]

[a] Proteomics Group of Ri.MED Foundation, Research Department IRCCS ISMETT (Istituto Mediterraneo per i Trapianti e Terapie ad Alta Specializzazione), Via E. Tricomi 5, 90127 Palermo, Italy
[b] Department of Medicine (DMED), University of Udine, via Colugna 50, 33100, Udine, Italy
[c] Unità Prodotti Cellulari (GMP), Ri.MED Foundation, IRCCS-ISMETT, Via E. Tricomi 5, 90127 Palermo, Italy
[d] Associazione Apistica Spazio Miele, Via Dell'Acquedotto 10, 91026 Mazara del Vallo, TP, Italy
[e] Dipartimento di Scienze e Tecnologie Biologiche, Chimiche e Farmaceutiche-STEBICEF, Università degli Studi di Palermo, V.le delle Scienze Ed.16, 90128 Palermo, Italy
[f] Dipartimento di Scienze e Tecnologie Biologiche, Chimiche e Farmaceutiche-STEBICEF, Università degli Studi di Palermo, V.le delle Scienze Ed.17, 90128 Palermo, Italy

ARTICLE INFO

Keywords:
Sicilian black honeybee
Honey
Proteomics
Machine learning
Protein signature

ABSTRACT

Apis mellifera ssp. *sicula*, also known as the Sicilian black honeybee, is a Slow Food Presidium that produces honey with outstanding nutraceutical properties, including high antioxidant capacity. In this study, we used high-resolution proteomics to profile the honey produced by *sicula* and identify protein classifiers that distinguish it from that made by the more common Italian honeybee (*Apis mellifera* ssp. *ligustica)*.

We profiled the honey proteome of genetically pure *sicula* and *ligustica* honeybees bred in the same geographical area, so that chemical differences in their honey only reflected the genetic background of the two subspecies, rather than botanical environment. Differentially abundant proteins were validated in *sicula* and *ligustica* honeys of different origin, by using the so-called "rectangular strategy", a proteomic approach commonly used for biomarker discovery in clinical proteomics. Then, machine learning was employed to identify which proteins were the most effective in distinguishing *sicula* and *ligustica* honeys. This strategy enabled the identification of two proteins, laccase-5 and venome serine protease 34 isoform X2, that were fully effective in predicting whether honey was made by *sicula* or *ligustica* honeybees.

In conclusion, we profiled the proteome of sicula honey, identified two protein classifiers of *sicula* honey in respect to *ligustica*, and proved that the rectangular strategy can be applied to uncover biomarkers to ascertain food authenticity.

1. Introduction

Honey is a sweet and viscous natural substance produced by bees to nourish the colony, and which has been consumed by humans since ancient times, as a nutritional product and also for its medical properties (Ranneh et al., 2021; Mandal and Mandal, 2011). Nowadays, the nutraceutical properties of honey are widely appreciated, with antioxidant, anti-inflammatory, antibacterial and antidiabetic properties described (Ranneh et al., 2021; Israili, 2014; Erejuwa et al., 2012; Ahmed et al., 2018). In this regard, over the last few years there has been growing interest in the nutraceutical properties of honey produced by *Apis mellifera* ssp. *sicula*, also known as the Sicilian black honeybee. This subspecies differs from *Apis mellifera* ssp. *ligustica*, which is the most widely distributed subspecies in Italy (also known as the Italian honeybee)

Abbreviations: MS, Mass spectrometry; FASP, Filtered assisted sample preparation; STAGE, Stop-and-go extraction; LFQ, label-free quantification; DDA, Data-dependent acquisition; DIA, Data-independent acquisition.

* Corresponding author.
E-mail address: sdscilabra@fondazionerimed.com (S.D. Scilabra).

[1] These authors equally contributed.

https://doi.org/10.1016/j.foodres.2024.114872
Received 2 May 2024; Received in revised form 2 August 2024; Accepted 5 August 2024
Available online 8 August 2024

(Sinacori et al., 1998). The *Apis mellifera* ssp. *sicula* originated from an African lineage; which eventually became widespread in the Sicilian area (Franck et al., 2000). Compared to *ligustica,* the Sicilian black honeybee has a dark abdomen and small wings (Sinacori et al., 1998; Ruttner, 1988), it is more resistant to extreme temperatures (Attanzio et al., 2016); has better skills in pollination, and shows higher physical and immunological resistance to pathogens such as varroa mites and viruses that commonly affect commercial honey bees (Franck et al., 2000). The Sicilian black honeybee risked extinction when *ligustica* was imported from Northern Italy for honey production, but inclusion of this species and its honey as a the Slow Food Presidium project in 2012 has contributed to its genetic preservation and a growing appreciation of its exceptional nutraceutical properties. Analysis of the chemical composition of *sicula* honey revealed that it had a higher phenolic content and antioxidant capacity than honey from other subspecies (Mannina et al., 2015; Bambina et al., 2023; Lo Dico et al., 2019), with the antioxidant capacity mainly due to its ability to reduce free radicals (Attanzio et al., 2016). Moreover, *sicula* honey showed high antimicrobial activity against different microorganisms, including *Escherichia coli* and *Salmonella serovar Infantis* (Coniglio et al., 2013). The unique nutraceutical properties of *sicula* honey, together with its outstanding organoleptic characteristics, position it on the market as an asset whose value will potentially increase in the next years, reaching a value similar to that of most expensive honeys worldwide.

Honey is one of the most adultered food (Lang, 2013), which usually occurs by addition of cheap and low-quality sweeteners or by blending an expensive honey with a low-cost honey (Se et al., 2019; Calle et al., 2023). Only few Sicilian producers breed genetically pure *sicula*, while most of them still use *ligustica* for honey production. In the case price of *sicula* honey will rise, as it is currently forecasted, it will become an attractive target to counterfeit and it is easy to speculate that cheaper *ligustica* honey may be subjected to fraudulent mislabeling as *sicula*. Indeed, mislabeling of the animal species used in a food product is another common example of food fraud (Moore et al., 2012). In addition to the negative effects that such fraud can cause to both farmers and the consumers, it may have detrimental effects on the safeguard of the Sicilian black honeybee, and therefore methods to distinguish *sicula* from *ligustica* honey are needed. In the last years, mass spectrometry has provided some of the most powerful tools for food authentication, as it allows the detection and identification of chemical markers of specific foods, such as honey (Lo Dico et al., 2019; Dou et al., 2023; Bocian et al., 2019). Proteomics, which is the mass spectrometry-based identification and quantification of proteins in complex biological samples, is extensively used in food research nowadays (Afzaal et al., 2022; Ortea et al., 2016). Proteins act as an indicator of origin, properties, and processes conducted on food, and also as classifiers of different but related animal species (Barik et al., 2013).

In this report, we characterized *sicula* honey by using high-resolution quantitative proteomics and set up a workflow that identified protein classifiers able to distinguish *sicula* honey from *ligustica* honey, even when *sicula* and *ligustica* honeys were of a different botanical origin. These findings may be developed into tools to ascertain the authenticity of *sicula* honey.

2. Materials and methods

2.1. Collection of honey samples

Six *Apis mellifera* ssp. *ligustica* queen bees were imported from Northern Italy (Apicoltura Ortolani, Reda di Faenza – RA) and six queen bees from genetically pure *Apis mellifera* ssp. *sicula* were imported from the islands of Lampedusa and Linosa (Associazione Apistica Spazio Miele). Morphometric characteristic analysis was performed by the Council for Agricultural Research and Economics (CREA) Bologna (Italy) to ensure that the honeybees used in this study truly belonged to *ligustica* and *sicula* subspecies (Supplementary material). Colonies of *Apis mellifera* ssp. *sicula* and *Apis mellifera* ssp. *ligustica* were grown in the area of Mazara del Vallo (Sicily, Italy), in close proximity to each other (about 60 m apart) to ensure that honey produced by both types of bees had the same botanical source, thereby minimizing their diversity due to environmental differences. Honey was extracted and filtered through a strainer into jars, which were kept at room temperature (RT) until sample preparation for proteomic analysis.

For the analysis of *sicula* and *ligustica* honey produced in different geographical areas, *sicula* honey samples were from Apicoltura Amodeo Carlo (C.da Madonna Diana, 90,018 Termini Imerese, PA) Nettare di Sicilia s.a.s. (C.da Cammarella, 90,022 Caltavuturo, PA) and La Mantia Giuseppe (fondo micciulla 26, 90135, Palermo.). Samples of *ligustica* honey were from Apicoltura Ortolani (Via Cangia 15 Reda di Faenza, 48018), Apicoltura Gardi-Petrarchini (str. Perugia Ponte Valleceppi 112, 06,135 Perugia) and Apicoltura Cantoni e Ottani (via Meucci 11, 40,017 San Giovanni in Persiceto, BO).

2.2. Protein extraction from honey

Acetone precipitation was used to extract proteins from honey, as previously described (Bocian et al., 2019). Each sample was diluted with ddH_2O in 1:10 ratio (w/v), incubated for 5 min at 35 °C and vortexed to homogenize the mixture. 2 ml of each sample were diluted with 5 volumes of cold acetone overnight at −20 °C. Then, honey samples were centrifuged for 30 min at 10,000 x g, supernatants were discarded, and protein pellets dried to let residual acetone evaporate. Proteins were collected with 300 μl of STET buffer (50 mM Tris, pH 7,5, 150 mM NaCl, 2 mM EDTA, 1 % Triton) containing EDTA-free protease inhibitor cocktail (1:25, Thermo Fisher Scientific, Waltham, Massachusetts, US). Protein concentration was measured by using a Pierce Bicinchoninic Acid (BCA) assay kit (Thermo Fischer Scientific, Waltham, Massachusetts, US) and a plate reader from Tecan (Männedorf, CH).

2.3. SDS-PAGE analysis and Coomassie blue staining

30 μg of protein extracted from the different honey samples was loaded onto acrylamide gels and analysed by SDS-PAGE electrophoresis. Gels were washed three times in ddH_2O and then stained with Coomassie Blue (QC Colloidal Coomassie Stain, BioRad, Hercules, California, US) for 15 min in the dark while shaking. Stained gels were destained in ddH_2O and were imaged using ChemiDoc Imaging System (BioRad).

2.4. Sample preparation and LC-MS/MS analysis

For each sample, 10 μg of protein was subjected to filter-aided sample preparation (FASP – with 10 kDa Vivacon 500 spin filter columns from Sartorius, Göttingen, Germany) (Wisniewski et al., 2009). Briefly, samples were loaded onto a 10 kDa Vivacon spin column and concentrated by centrifugation (14,000 x g; 10 min). Samples were then reduced and alkylated with dithiothreitol (DTT, Sigma-Aldrich, part of Merck Group, St. Louis, Missouri, US) in 200 μl of UA buffer (8 M Urea in 0.1 M Tris/HCl, pH 8.5) and alkylated with 50 mM iodoacetamide (IAA) (Sigma-Aldrich). After reduction and alkylation, samples were washed three times and sequentially digested with LysC (1:50 enzyme to protein ratio, from Promega, Madison, Wisconsin, US) and trypsin (1:100 enzyme to protein ratio, Promega). Then, peptides were eluted from filter columns by centrifugation (14,000 x g; 60 min) and acidified with 20 μl of 8 % formic acid. Peptides were desalted by stop-and-go extraction (STAGE) on reverse phase tips packed with Empore C18 disks in-house (Sigma-Aldrich), as previously described (Rappsilber et al., 2003). Peptides were eluted; vacuum-dried and resuspended in 20 μl 0.1 % formic acid. After measuring their concentration by Nanodrop 2000 (Thermo Scientific), 1 μg of peptides were separated using a nanoLC system (Vanquish Neo UHPLC, Thermo Scientific) equipped with an Acclaim PEPMap C18 column (25 cm x 75 μm ID, Thermo Scientific) in a 130 min binary gradient of water and acetonitrile

containing 0.1 % formic acid. Separated peptides were ionized using a nano electrospray ion source and analysed into an Exploris 480 mass spectrometer (Thermo Fischer Scientific) for tandem mass spectrometry analysis.

2.5. Proteomic data analysis

Honey proteins were identified and quantified with label-free quantification (LFQ) using two different data acquisition approaches: (i) data-dependent acquisition (DDA) and (ii) data-independent acquisition (DIA). DDA was performed using an MS1 full scan (300–1450 *m/z*) at a resolution of 120,000, an automatic gain control (AGC) of 1 x 10^6 ions and a maximum injection time of 50 ms, followed by sequential fragmentation of the most abundant precursors for a cycle time of 1 s. Precursors were filtered according to 1 x 10^4 intensity threshold, 2–6 charge state and a dynamic exclusion of 60 s. Fragmentation was performed at 30 % higher-energy collisional dissociation (HCD) and fragments were identified at a resolution of 15,000 and an AGC of 1 x 10^5. The data were analysed using the software MaxQuant (version 2.0.1) and an *Apis mellifera* RefSeq protein reference database (https://ftp.ncbi.nlm.nih.gov/genomes/refseq/invertebrate/Apis_mellifera/latest_assembly_versions/GCF_003254395.2_Amel_HAv3.1/GCF_003254395.2_Amel_HAv3.1_protein.faa.gz) with standard settings. (ii) DIA was performed using an MS1 full scan followed by 60 sequential DIA windows with an overlap of 1 *m/z* and window placement optimization option enabled. Full scans were acquired with a resolution of 120000, AGC of 3 x 10^6, and maximum injection time of 50 ms. Afterwards, 60 isolation windows were scanned with a resolution of 30,000, an AGC of 8 x 10^5 and maximum injection time was set as auto to achieve the optimal cycle time. Collision-induced dissociation fragmentation was induced with 30 % of the normalized HCD. The data were analysed using the software DIA-NN (version 1.8.1) and a predicted library generated from in silico digestion of the Apis mellifera proteome (RefSeq protein reference database –https://ftp.ncbi.nlm.nih.gov/genomes/refseq/invertebrate/Apis_mellifera/latest_assembly_versions/GCF_003254395.2_Amel_HAv3.1/GCF_003254395.2_Amel_HAv3.1_protein.faa.gz), with cleavages at K* and R*, two missed cleavages allowed and a minimal peptide length set at 6 residues. The reference proteome of *Apis mellifera* consisted of 9919 proteins, with 4,246,929 precursors generated by the in silico digestion. For both DDA and DIA, LFQ values were log2 transformed, a two-sided Student's *t*-test and false discovery rate (FDR) for multiple testing correction were used to evaluate proteins differentially abundant in the honey produced by *sicula* and *ligustica*.

2.6. Machine learning for the identification of a sicula honey proteomic signature

LFQ protein intensities were Z-score normalized within groups and missing intensities were replaced with 0 using Perseus software version 1.6.1.3 (Tyanova and Cox, 2018). After Z-scoring; changes in protein abundance were evaluated by a Student *t*-test, with false discovery rate (FDR) correction for multiple hypothesis testing. Unsupervised clustering was applied to differentially abundant proteins ($q < 0.05$) using pheatmap (1.0.12) and factoextra (1.0.7) packages in R studio (4.2.2) with standard parameters. A clustered heatmap and principal component analysis (PCA) were plotted using integrated tools.

The Boruta algorithm was employed to select relevant features for honey classification by machine learning (Kursa and Rudnicki, 2010). The 19 important features identified by the Boruta algorithm were iteratively added to a Random Forest ensemble method according to their importance. The dataset was shuffled and 66 % of our dataset samples were employed to train the Random Forest model, and the remaining 33 % of samples used to test the prediction model. The prediction ability of the trained algorithm was evaluated by a confusion matrix. In addition, the algorithm was used to seek how many and which features were necessary to discriminate between *sicula* and *ligustica* honeys with full accuracy by using this model. The resulting dataset, comprising two protein classifiers, was used for PCA and hierarchical clustering analysis.

3. Results

3.1. Quantitative proteomic analysis identifies differences in honey produced by Apis mellifera ssp. Sicula and ligustica of the same botanical origin

3.1.1. Protein extraction from honey

Acetone precipitation was used to extract proteins from honey prior to tryptic digestion and mass spectrometry analysis. This yielded an average of 1727.5 μg of protein per g of honey (Fig. 1A). This method was more effective in extracting proteins from honey than others tested, including dialysis through filter columns (Supplementary Figure S1). Protein yield did not differ significantly between *sicula* and *ligustica* honey (Fig. 1A). Honey proteins were analysed by sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) and Coomassie blue staining which allowed separation and visualization of individual proteins in the complex honey mixture. This analysis confirmed that acetone precipitation enabled effective extraction of proteins from honey, and that *sicula* and *ligustica* honeys displayed a similar pattern of protein bands, with two major bands running slightly below 50 kDa and at around 70 kDa, with other less intense bands between 10 kDa and 20 kDa, and at 30 kDa, 75 kDa and 150 kDa (Fig. 1B).

3.1.2. Assessment of different data acquisition strategies for analysis of honey proteins

Honey proteins from *sicula* and *ligustica* were applied to a high-resolution mass spectrometry-based workflow, which comprised tryptic digestion on filter columns (FASP – (Wisniewski et al., 2009) and protein analysis via LC-MS/MS followed by label-free quantification. Protein detection and quantification were performed by using either data-dependent acquisition (DDA) or data independent acquisition (DIA). DDA performs a first MS scan to determine mass-to-charge ratio (*m/z*) and abundance of peptide ions entering the mass spectrometer at each time, then selects the most intense peptide ions for subsequent fragmentation and sequence determination. This method, which is very accurate for quantification of the most abundant proteins in a complex protein sample, is quite limited for identification of low abundance proteins. In comparison, DIA is not limited to sequencing of the most abundant peptides, but fragments and sequences all peptides within a defined *m/z* window regardless of their intensity. This makes DIA particularly advantageous for detection of low-abundance proteins. DIA identified and quantified 168 different proteins present in both *sicula* and *ligustica*, which was 3-fold more proteins than were identified by DDA (Fig. 2A, Fig. 2B and Supplemental Table). 30 of these 168 proteins were also identified by DDA (Fig. 2B). Intriguingly, DDA identified 25 proteins that were not detected by DIA. This indicated that, although DIA was clearly more effective than DDA for characterizing the honey proteome, the two analyses were complementary, and together identified and quantified almost 200 proteins present in both species (Fig. 2B).

3.1.3. High-resolution mass spectrometry-based analysis of honey proteins from sicula and ligustica

A similar number of proteins were detected in *sicula* and *ligustica* honey, irrespective of whether samples analysed by DIA or DDA (Fig. 2C and 2D). Then, we evaluated differences in the relative abundance of proteins extracted from *sicula* and *ligustica* honey. Out of the 168 proteins detected by DIA in both *sicula* and *ligustica* honey, 26 proteins were differentially abundant (Fig. 2E, Supplemental Table). Levels of 24 proteins were higher in *sicula* honey (Fig. 2E, Table 1), including phospholipase A1 (A0A7M7MR18), laccase-5 (A0A7M7RC42), bee-milk protein (A0A7M7SQQ4) and venom serine protease isoform X2 (A0A7M7L3N3). 2 proteins were more abundant in *ligustica* honey:

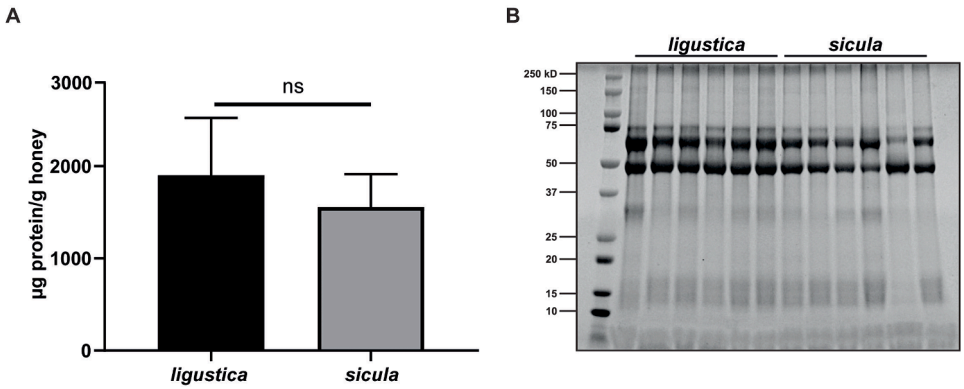

Fig. 1. Protein extraction from *sicula* and *ligustica* honey. A. Proteins were extracted from honey produced by *Apis mellifera* ssp. *sicula* or *ligustica* through acetone precipitation and quantified by BCA assay. Histograms show the yield of extracted proteins. A two-sided Student *t*-test was used to evaluate differences in protein content between honey produced by each subspecies (ns, $p > 0.05$). **B**. SDS-PAGE analysis and Coomassie blue staining of proteins extracted from *sicula* or *ligustica* honey.

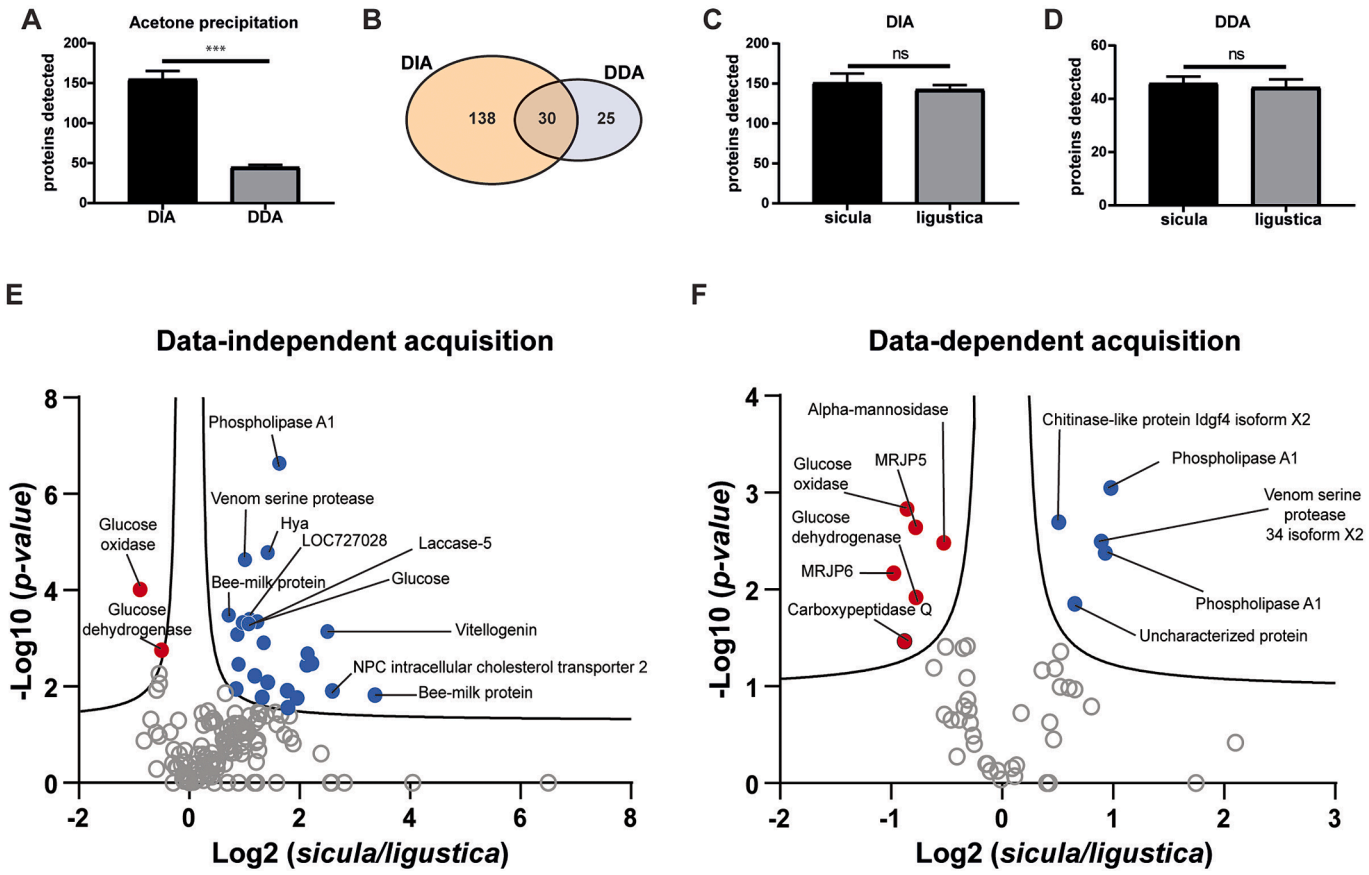

Fig. 2. High-resolution proteomic characterization of *sicula* and *ligustica* honey. A. After extraction by acetone precipitation, proteins from *sicula* honey were digested by FASP and applied to LC-MS/MS. Proteins were identified and quantified using two different approaches: data-dependent acquisition (DDA) and data-independent acquisition (DIA). Histograms show the average number of proteins identified and quantified in *sicula* honey by DIA or DDA (151 ± 11 and 45 ± 4, respectively; *** $p < 0.005$; $n = 6$). **B**. Venn diagrams showing the number of proteins found by DIA and/or DDA data acquisition methods in *sicula* honey. **C-D**. Proteins from *sicula* and *ligustica* honey were analysed by LC-MS/MS followed by DIA or DDA. Histograms show the number of proteins detected in *sicula* honey and *ligustica* honey by DIA (151 ± 11 and 143 ± 7, respectively) and DDA (45 ± 4 and 44 ± 4, respectively). A two-sided Student *t*-test was used to evaluate statistical differences in the number of proteins detected and quantified in *sicula* or *ligustica* honey (ns $p > 0.05$). **E.** Volcano plot showing Log_2 of relative abundance ratio (*sicula/ligustica,* acquired by DIA) versus $-Log_{10}$ p-value of change for 168 proteins detected in *sicula* and *ligustica* honey. The black hyperbolas indicate the false discovery rate (FDR). Proteins displayed above the FDR curves were considered significantly different between honey of *sicula* and *ligustica*. Proteins significantly more abundant in *sicula* honey are displayed as filled blue dots,and proteins more abundant in *ligustica* honey as filled red dots. Proteins below the FDR curve did not significantly differ in the honey of *sicula* and *ligustica*, and they are represented by open grey circles. **F.** Volcano plot showing Log_2 of relative abundance ratio (*sicula/ligustica*, acquired by DDA) versus $-Log_{10}$ p-value of change for 46 proteins detected in *sicula* and *ligustica* honey. Proteins significantly more abundant in *sicula* honey are displayed as filled blue dots, and proteins more abundant in *ligustica* honey as filled red dots, while proteins that did not significantly differ are represented by open grey circles.

Table 1
Proteins significantly more abundant in the honey of *Apis mellifera* ssp. *sicula* analysed by LC-MS/MS followed by DIA.

Protein name	Uniprot-ID	Difference	p-value
Phospholipase A1	A0A7M7MR18	1.63	2.30E-07
Hyaluronidase (Hya)	Q08169	1.42	1.64E-05
Venom serine protease 34 isoform X2	A0A7M7L3N3	1.01	2.32E-05
Bee-milk protein	Q4ZJX1	0.71	3.27E-04
Uncharacterized protein LOC727028	A0A7M7FZE0	1.09	3.99E-04
Glucose dehydrogenase [FAD, quinone]	A0A7M7R506	1.23	4.45E-04
Phospholipase A1 isoform	A0A7M7IQ52	0.97	4.67E-04
Laccase-5	A0A7M7RC42	1.07	4.99E-04
Vitellogenin	Q868N5	2.5	7.13E-04
Glucose dehydrogenase [FAD, quinone]	A0A7M7GC96	0.87	8.29E-04
Calmodulin	A0A7M7LRU3	1.35	1.24E-03
Chymotrypsin inhibitor	A0A7M7GQH1	2.14	2.09E-03
Apolipophorins	A0A7M7SQ18	2.22	3.27E-03
Venom acid phosphatase Acph-1	Q5BLY5	1.18	5.95E-03
Prophenoloxidase (phenoloxidase subunit A3)	Q86MV4	1.42	8.18E-03
Apolipophorin-III-like protein	B0LUE8	0.85	1.11E-02
UDP-glucose 6-dehydrogenase	A0A7M7GZT2	1.78	1.21E-02
NPC intracellular cholesterol transporter 2	A0A7M7G1N6	2.59	1.23E-02
Bee-milk protein	A0A7M7SQQ4	3.37	1.50E-02
40S ribosomal protein S28	A0A7M7GS09	1.32	1.68E-02
Odorant binding protein 14 precursor	A0A8U0WQC3	1.95	1.74E-02
DNA topoisomerase	A0A7M7RC27	1.78	2.75E-02

Uniprot-ID: accession number of the protein. Difference: log2 transformed mean ratio of label-free quantification intensities between *Apis mellifera* ssp. *sicula and ligustica* (n = 6). p-value: p-value of intensities change between *Apis mellifera* ssp. *sicula and ligustica* (n = 6).

glucose oxidase (Q9U8X6) and glucose dehydrogenase isoform X1 (A0A7M7RBJ1) (Table 2). When data were acquired by DDA, 56 proteins were detected and quantified in the honey of *sicula* and *ligustica*, with 11 proteins differentially abundant (Fig. 2F, Supplemental Table). 5 proteins were higher in *sicula* honey (Table 3), including phospholipase A1 (A0A7M7MR18), phospholipase A1 isoform (A0A7M7IQ52) and venom serine protease 34 isoform X2 (A0A7M7L3N3), three

Table 2
Proteins significantly more abundant in the honey of *Apis mellifera* ssp. *ligustica* analysed by LC-MS/MS followed by DIA.

Protein name	Uniprot-ID	Difference	p-value
Glucose oxidase	Q9U8X6	−0.89	9.62E-05
Glucose dehydrogenase [FAD, quinone] isoform X1	A0A7M7RBJ1	−0.5	1.76E-03

Uniprot-ID: accession number of the protein. Difference: log2 transformed mean ratio of label-free quantification intensities between *Apis mellifera* ssp. *sicula and ligustica* (n = 6). p-value: p-value of intensities change between *Apis mellifera* ssp. *sicula and ligustica* (n = 6).

Table 3
Proteins significantly more abundant in the honey of *Apis mellifera* ssp. *sicula* analysed by LC-MS/MS followed by DDA.

Protein name	Uniprot-ID	Difference	p-value
Phospholipase A1	A0A7M7MR18	0.98	8.94E-04
Venom serine protease 34 isoform X2	A0A7M7L3N3	0.89	3.19E-03
Phospholipase A1 isoform	A0A7M7IQ52	0.93	4.20E-03
Chitinase-like protein Idgf4 isoform X2	A0A7M7M4R3	0.51	2.02E-03
Uncharacterized protein	A0A7M7LN70	0.66	1.41E-02

Uniprot-ID: accession number of the protein. Difference: log2 transformed mean ratio of label-free quantification intensities between *Apis mellifera* ssp. *sicula and ligustica* (n = 6). p-value: p-value of intensities change between *Apis mellifera* ssp. *sicula and ligustica* (n = 6).

proteins that were also found to be more abundant by DIA. On the other hand, DDA found levels of 6 proteins higher in *ligustica* honey (Table 4), including glucose oxidase (Q9U8X6).

3.2. Proteomic analysis of sicula and ligustica honey from different geographical areas

When bees were kept in the same area and honeys were of same botanical origin, proteomic analysis led to identification of a number of proteins that were more abundant in *sicula* compared to *ligustica*. Thus, we investigated whether these proteins could be distinctive identifiers of *sicula* honey and also allow its identification over *ligustica* when honeybees were kept in different geographical areas. In order to do so, we analysed honeys produced by *sicula* in different areas of Sicily (e.g. in the Aeolian islands Vulcano and Vulcanello, and in Palermo area) and *ligustica* honeys produced in Northern Italy (Tuscany, Umbria and Emilia Romagna). DIA proteomics was applied to acetone precipitated honey proteins, as for the initial analysis. 192 proteins were detected and quantified in both *sicula* and *ligustica* honey (Fig. 3, Supplemental Table). Glucose dehydrogenase was the only protein more abundant in *ligustica* honey, in agreement with the previous analysis (Fig. 3). 8 proteins were higher in *sicula* (Table 5), 4 of which [i.e. hyaluronidase (Hya) (A0A7M7MTB6), venom serine protease 34 isoform X2 (A0A7M7L3N3), bee-milk protein (A0A7M7SQQ4) and laccase-5 (A0A7M7RC42)], were also more abundant in the previous analysis when *sicula* and *ligustica* were kept in the same geographical area and their honeys of the same botanical origin (compare Fig. 2E and 3). Phospholipase A1 and Phospholipase A1 isoform, which were higher in

Table 4
Proteins significantly more abundant in the honey of *Apis mellifera* ssp. *ligustica* analysed by LC-MS/MS followed by DDA.

Protein name	Uniprot-ID	Difference	p-value
Alpha-mannosidase	A0A7M7MGK9	−0.53	3.31E-03
Glucose dehydrogenase [FAD, quinone] isoform X2	A0A7M7IL97	−0.78	1.21E-02
Carboxypeptidase Q	A0A7M7LQ78	−0.88	3.42E-02
Major royal jelly protein MRJP6	A0A8U0WQ84	−0.98	6.80E-03
Major royal jelly protein 5	A0A8U0WQ67	−0.78	2.28E-03
Glucose oxidase	O97432	−0.86	1.47E-03

Uniprot-ID: accession number of the protein. Difference: log2 transformed mean ratio of label-free quantification intensities between *Apis mellifera* ssp. *sicula and ligustica* (n = 6). p-value: p-value of intensities change between *Apis mellifera* ssp. *sicula and ligustica* (n = 6).

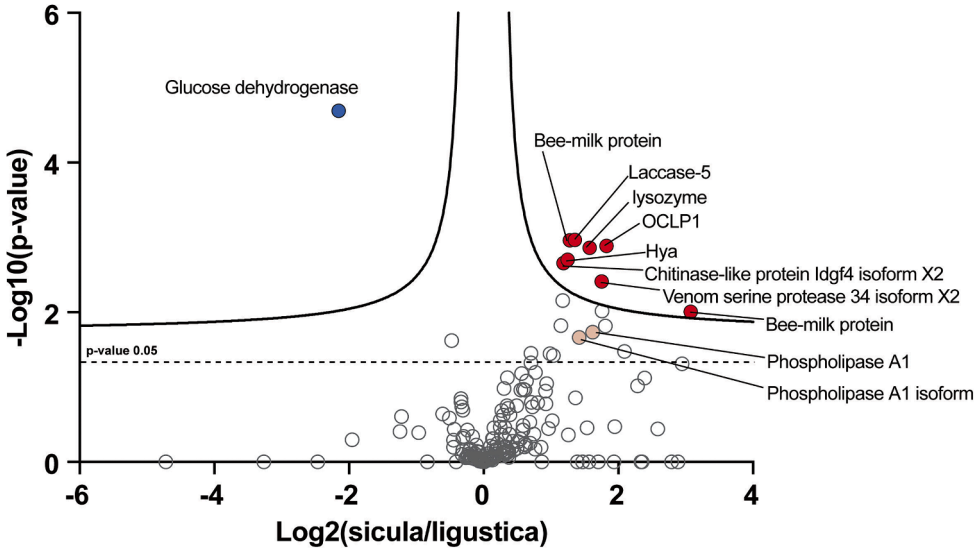

Fig. 3. Proteomic analysis of honey produced by *sicula* and *ligustica* in different botanical areas. Volcano plot showing Log_2 of relative abundance ratio (*sicula/lingustica*) versus $-Log_{10}$ p-value of change for 192 proteins detected in *sicula* and *ligustica* honey. The black hyperbolas indicate the FDR. Levels of proteins displayed above the FDR curves were considered significantly different in honey of *sicula* and *ligustica*. Proteins significantly more abundant in *sicula* honey are displayed as filled blue dots and proteins more abundant in *ligustica* honey as filled red dots. Proteins below the FDR curve did not significantly differ in the honey of *sicula* and *ligustica*, and are represented by open grey circles.

Table 5
Proteins significantly more abundant in the honey of *sicula* compared to *ligustica* of different botanical origin, analysed by LC-MS/MS followed by DIA.

Protein name	Uniprot-ID	Difference	p-value
Laccase-5	A0A7M7RC42	1.31	1.08E-03
Lysozyme (EC 3.2.1.17)	A0A7M7R3V2	1.57	1.38E-03
Omega-conotoxin-like protein 1	H9KQJ7	1.82	1.29E-03
Chitinase-like protein ldgf4 isoform X2	A0A7M7M4R3	1.25	1.99E-03
Venom serine proase 34 isoform X2	A0A7M7L3N3	1.75	3.85E-03
Bee-milk protein	A0A7M7SQQ4	3.07	9.83E-03
Hyaluronidase (Hya) (EC 3.2.1.35)	A0A7M7MTB6	1.19	2.21E-03
Bee-milk protein	A0A7M7MQN8	1.28	1.09E-03

Uniprot-ID: accession number of the protein. Difference: log2 transformed mean ratio of label-free quantification intensities between *Apis mellifera ssp*. sicula and ligustica (n = 6). p-value: p-value of intensities change between *Apis mellifera ssp*. sicula and ligustica (n = 6).

sicula in all previous datasets, were still more abundant in *sicula* honey, although this did not reach significance when the more stringent false-discovery rate (FDR) correction for multiple hypothesis testing was applied (Fig. 3). In conclusion, proteomics revealed a number of proteins that were consistently higher in *sicula* honey compared to *ligustica* regardless of their botanical origin and the geographical area in which the honeys were produced.

3.3. Identification of a proteomic signature of sicula honey

3.3.1. Selection of sicula honey protein classifiers by using a "rectangular" proteome profiling strategy

Proteomics has been largely used for unbiased identification of protein biomarkers associated with specific conditions or diseases (Bader et al., 2020; Geyer et al., 2017). A common workflow for such hypothesis-free biomarker discovery comprises proteome profiling of biological samples (e.g. plasma, biopsies, etc.) collected from a study cohort that shares a specific condition and from a control cohort. This analysis enables selection of candidate proteins associated with the condition, which are eventually validated in a different cohort by orthogonal means, including immunoassays (Geyer et al., 2017). This workflow, which goes from the identification of a large number of proteins through proteomics to the validation of a small number of them by immuoassays, is referred to as the "triangular" strategy of biomarker discovery. In the last decade, advances in mass spectrometers resolution, development of more powerful methods for proteomic data acquisition and for in silico model predictions have allowed a "rectangular" strategy for biomarker discovery. When using this approach, a large amount of proteomic data obtained in an initial proteome profiling is subsequently validated by shotgun proteomics in a different study cohort. This strategy has the advantage of enabling discovering and validation of several protein biomarkers and patterns characteristic of a particular condition, rather than the single biomarker candidates that arise from the "triangular" strategy.

We applied a "rectangular" strategy to identify proteins able to discriminate *sicula* from *ligustica* honey (Fig. 4A). First, we assessed whether *sicula* and *ligustica* honeys clustered together based on 30 proteins that were differently abundant in the two datasets (i.e. honey proteomes from *sicula* and *ligustica* bees of same or different origin), after protein intensities were Z-scored and a Student *t*-test with FDR correction applied. 26 of these proteins were more abundant in *sicula* honey (e.g. hyaluronidase, bee milk protein, etc.) and 4 proteins in *ligustica* (e.g. glucose dehydrogenase) (Fig. 4B). Unsupervised clustering and principal component analysis clearly separated honeys produced by the two honeybee subspecies, both when they were of the same botanical origin and when they were produced in different geographical areas (Fig. 4B and 4C).

3.3.2. Classification of sicula honey by machine learning

Next, we used machine learning to assess which of this collection of 30 proteins could be used to classify *sicula* honey compared to *ligustica*. First, to determine whether these proteins could be an important feature for honey classification, we employed the Boruta algorithm (Kursa and Rudnicki, 2010). The main idea of this approach was to compare the importance of these proteins as real predictor variables with the so-

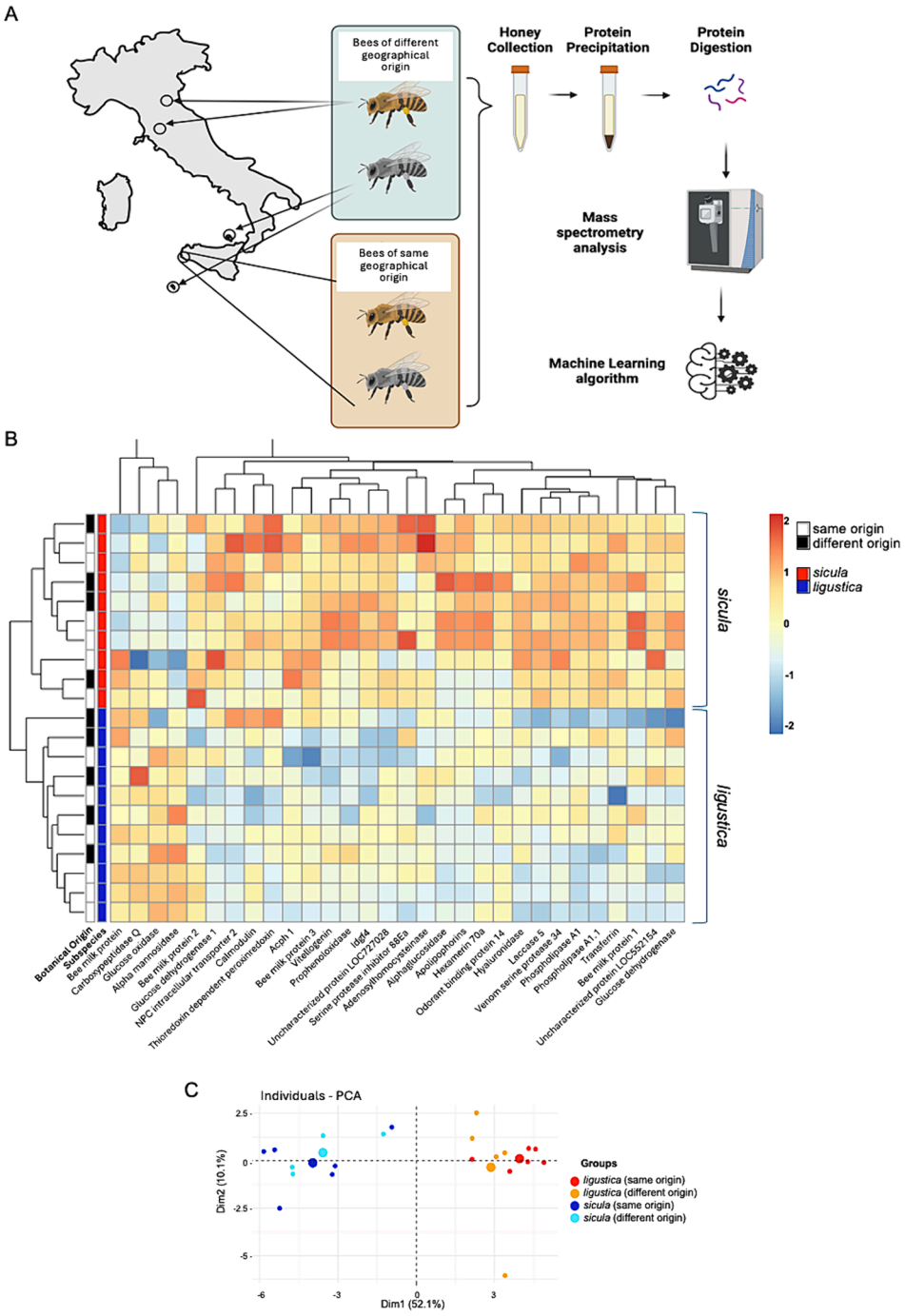

Fig. 4. Identification of protein classifiers to distinguish *sicula* honey from *ligustica* honey A. Schematic representation of the "rectangular strategy" used to identify protein classifiers of *sicula* honey. **B.** Protein intensities were Z-scored and proteins that differed significantly in abundance between *sicula* honey and *ligustica* honey ($q < 0.05$) were used for hierarchical clustering. As shown by the heat map, *sicula* and *ligustica* honeys each clustered together and were clearly separated. **C.** Principal component analysis of 22 honeys (*sicula* or *ligustica* of same or different origin) showed good separation of *sicula* and *ligustica* honeys (large points represent the median value of each group).

called random shadow variables. The Boruta algorithm confirmed the importance of 19 proteins as classifiers of *sicula* honey (Fig. 5A). Next, to further reduce the number of candidate protein classifiers to those with a highest predicted capability to distinguish between *sicula* and *ligustica* honey, we iteratively added important proteins, according to the Boruta algorithm, to a Random Forest ensemble method. By this, we found that laccase-5 and venom serine protease 34 isoform X2 were sufficient to create a model that was fully accurate in discriminating between *sicula* and *ligustica* honey (Fig. 5B). In line with this, unsupervised clustering and PCA analysis further confirmed an optimal separation of *sicula* and *ligustica* honeys, regardless of their origin, based on these two protein classifiers (Fig. 5C and D).

4. Discussion

The *Apis mellifera* ssp. *sicula*, also known as the Sicilian black honeybee, is a Sicilian native subspecies which was appointed a Slow Food Presidium in 2012, thereby contributing to its genetic preservation. Various beneficial pharmacological properties of *sicula* honey have been demonstrated and, as a consequence, interest in this honey has increased

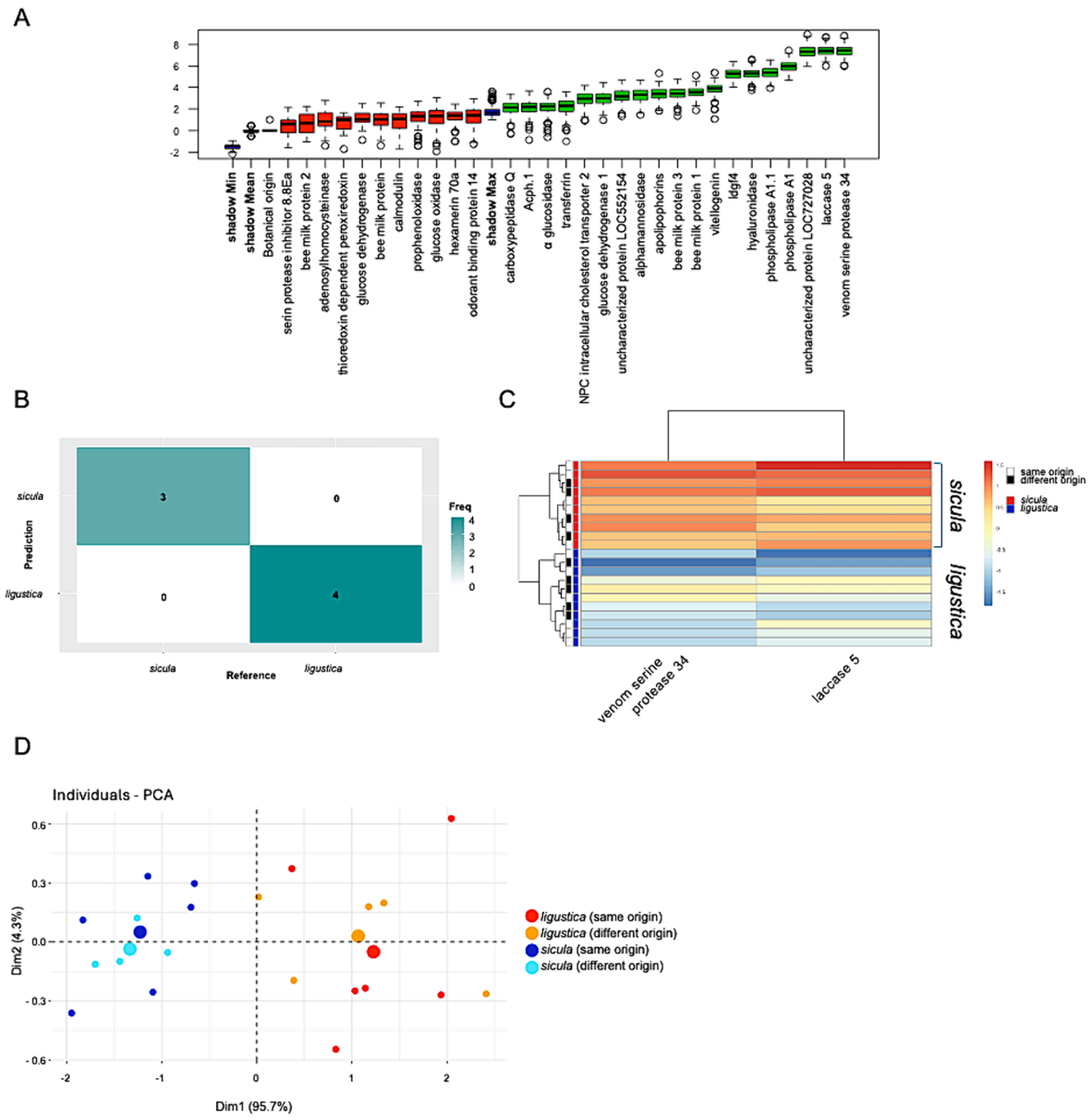

Fig. 5. Machine learning identifies laccase-5 and venom serine protease 34 isoform X2 as sufficient to fully discriminate between *sicula* and *ligustica* honey. A. Feature selection based on the Boruta algorithm. The horizontal axis is the name of each variable (proteins that differed in abundance between *sicula* and *ligustica* honey) and the vertical axis is the Z-score of each variable. The green boxplots represent the important variables and the red represent unimportant variables. The blue boxplots correspond to minimal, average, and maximum Z-score of a shadow attribute, which are automatically generated by the algorithm. **B.** Confusion matrix that evaluates the performance of a supervised machine learning model based on abundance of laccase 5 and venom serine protease 34 isoform X2 to predict *sicula* over *ligustica* honey. **C-D.** Heatmap showing hierarchical clustering and principal component analysis of *sicula* and *ligustica* honeys based on abundance of laccase 5 and venom serine protease 34 isoform X2.

considerably, and its commercial value is expected to grow in the near future (Attanzio et al., 2016; Mannina et al., 2015; Bambina et al., 2023; Tenore et al., 2012). As its price will rise, we expect that *sicula* honey will be subjected to food fraud. Given that only few producers breed the Sicilian black honeybee and keep it genetically pure on small islands around Sicily, while most Sicilian producers still use *ligustica* for honey production, the risk that *ligustica* honey will be sold as *sicula* is very high. Clearly, methods to discriminate honey made by the two subspecies are needed to protect producers and consumers from food fraud, and also to safeguard the *Apis mellifera* ssp. *sicula* from extinction. In this study, we characterized the proteome of honey produced by the Sicilian black honeybee and identified proteins that can distinguish *sicula* from *ligustica*, and may be utilized for ascertaining its authenticity. This proteomic signature may also shed light on the peculiar behaviors of the Sicilian black honeybee, such as its high immunological resistance to Varroa. Our approach detected and quantified almost 200 bee proteins contained in the honey which is an unprecedented high number. By matching peptides detected by mass spectrometry with the honeybee proteome, our method selectively identified honeybee proteins rather than botanical proteins contained in the honey. We used a two-phase design to carry out this study. First, we applied proteomics to honeys harvested from colonies of *sicula* and *ligustica* that were grown in close proximity to each other. We expect that, in such a case, proteomic differences between *sicula* and *ligustica* honey would only reflect the different genetic background of these two subspecies, rather than external agents (e.g. botanical origin). Then, to confirm that some proteins can be distinctive of *sicula* honey regardless of its origin, we repeated the analysis on *sicula* and *ligustica* honeys produced in different

geographical areas. The first analysis revealed that levels of 24 proteins were higher in *sicula* honey. Among them, there were several lyases, hydrolases and oxidases that may exert antioxidant and anti-inflammatory effects (Naqvi et al., 2023). In addition, this analysis found that vitellogenin was more abundant in *sicula* honey than *ligustica*. Vitellogenin plays a crucial role in the bee immune system. Similar to other insects, bees can recognize specific pathogens and transfer immune elicitors such as pathogen fragments into developing oocytes, thereby priming offspring immunity. This process, known as *trans*-generational immune priming, is mediated by vitellogenin that transports immune elicitors and favors transmission of immune resistance to the offspring (Dickel et al., 2022; , 2023; Salmela et al., 2015). Vitellogenin is currently used to "vaccinate" honeybees, and its higher levels in *sicula* honey suggest a higher production of vitellogenin by the Sicilian black honeybees, which, in turn, may explain its increased resistance to varroasis and virosis (Franck et al., 2000). Finally, to identify protein classifiers that could distinguish *sicula* from *ligustica* honey, we utilized a workflow commonly used for discovery of biomarkers associated with human health. Based on this workflow, known as the "rectangular" strategy for biomarker discovery, candidate protein biomarkers that emerged from the proteomic analysis of *sicula* and *ligustica* honeys of same geographical origin were validated by a sequential proteomic analysis of *sicula* and *ligustica* honeys of different geographical origin. This analysis, which was complemented by machine learning to assess the importance of candidate proteins to serve as predictive classifiers, identified a proteomic signature of *sicula* honey comprising 19 proteins, 2 of which, laccase-5 and venom serine protease isoform X2, were able to fully predict *sicula* honey in our model. This represents a first step towards the development of tools to prove its authenticity over *ligustica* honey, in line with the aims of the "Slow Food" project to safeguard food biodiversity and the related growing interest in identifying and using biomarkers for more accurate and objective measurements of food's provenance and quality. Our work could be developed by establishing methods for determining absolute concentration of laccase-5 and venom serine protease isoform X2 in *sicula* honey, as well as a concentration threshold for these two proteins above which *sicula* honey could be considered "authentic". Furthermore, evaluation of the peptide library generated by our mass spectrometry analysis could lead to the identification of unique peptides of *sicula* honey that may be used as additional markers, in a similar manner as unique peptides were selected as authentication markers for mānuka honey (Bong et al., 2021). These mass spectrometry-based approaches could be coupled with other methods for honey authentication, including identification of DNA signatures and sugar profiling. In the future, such an integrated strategy could be used for conclusive identification of *sicula* honey.

5. Conclusions

After near extinction, *Apis mellifera* ssp. *sicula* became a Slow Food "presidium" in 2012, and its honey has been reported to have beneficial nutraceutical properties. For this reason, the commercial value of *sicula* honey is predicted to raise in the coming years. To date, only a limited number of producers can access genetically pure *sicula*, as they breed it on small islands surrounding Sicily, thereby preventing its hybridization with the more common *ligustica*. It is speculated that once the value of *sicula* honey increases, cheaper *ligustica* honey will be mislabeled as *sicula* honey, and tools to discriminate the two will be needed to prevent fraud. For this reason we sought protein biomarkers that could distinguish *sicula* honey from *ligustica*. To do so, we used the innovative "rectangular strategy" proteomic workflow, commonly used for clinical biomarker discovery, and coupled this with machine learning. Our analysis identified two honey proteins, laccase-5 and venom serine protease isoform X2, that were able to distinguish *sicula* honey from *ligustica* regardless of their botanical origin or the geographical area in which honeys were produced. Mass spectrometry-based or antibody-based assays for an easy detection and quantification of these two proteins could be developed and used, either alone or in combination with other techniques for honey authentication, to ascertain the authenticity of *sicula* honey.

Funding

This work was funded by "National Biodiversity Future Center—NBFC" (CUP: B73C2100130006) CN_00,000,033 granted by Next Generation EU Programme—PNRR—M4C2—CN5 Spoke 6 and by "OCM Apicoltura, bando finalizzato al miglioramento della produzione e commercializzazione del miele, campagna 2021/2022 (UE) N.1308/2013".

CRediT authorship contribution statement

Giulia Biundo: Writing – original draft, Project administration, Formal analysis. **Matteo Calligaris:** Writing – original draft, Supervision, Formal analysis. **Margot Lo Pinto:** Formal analysis. **Danilo D'apolito:** Formal analysis. **Salvatore Pasqua:** Formal analysis. **Giulio Vitale:** Resources. **Giuseppe Gallo:** Writing – review & editing. **Antonio Palumbo Piccionello:** Writing – original draft, Supervision, Formal analysis. **Simone D. Scilabra:** Writing – original draft, Supervision, Project administration, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: [Giulio Vitale is a member of "Associazione allevatori Apis mellifera siciliana" and president of "Associazione apistica spazio miele"].

Data availability

Data will be made available on request.

Acknowledgements

We would like to thank Apicoltura Amodeo Carlo, Nettare di Sicilia s.a.s., La Mantia Giuseppe, Apicoltura Ortolani, Apicoltura Gardi-Petrarchini and Leandro Cilia (Apicoltura Cantoni e Ottani) for providing *sicula* and *ligustica* honey used in this study. In addition, we would like to thank Sergio Sapienza who helped with beekeeping. Finally, we would like to thank Dr Linda Troeberg (University of East Anglia Norwich, UK) for her help in editing the manuscript, and Dr Stephan A. Müller for his insightful comments.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodres.2024.114872.

References

Ranneh, Y., Akim, A. M., Hamid, H. A., Khazaai, H., Fadel, A., Zakaria, Z. A., Albujja, M., & Bakar, M. F. A. (2021). Honey and its nutritional and anti-inflammatory value. *BMC Complement Med Ther, 21*, 30.

Mandal, M. D., & Mandal, S. (2011). Honey: Its medicinal property and antibacterial activity. *Asian Pacific Journal of Tropical Biomedicine, 1*, 154–160.

Israili, Z. H. (2014). Antimicrobial properties of honey. *American Journal of Therapeutics, 21*, 304–323.

Erejuwa, O. O., Sulaiman, S. A., & Wahab, M. S. (2012). Honey–a novel antidiabetic agent. *International Journal of Biological Sciences, 8*, 913–934.

Ahmed, S., Sulaiman, S. A., Baig, A. A., Ibrahim, M., Liaqat, S., Fatima, S., Jabeen, S., Shamim, N., & Othman, N. H. (2018). Honey as a Potential Natural Antioxidant Medicine: An Insight into Its Molecular Mechanisms of Action. *Oxidative Medicine and Cellular Longevity, 2018*, 8367846.

Sinacori, A., Rinderer, T. E., Lancaster, V., & Sheppard, W. S. (1998). A morphological and mitochondrial assessment of Apis mellifera from Palermo, Italy. *Apidologie, 29*, 481–490.

Franck, P., Garnery, L., Celebrano, G., Solignac, M., & Cornuet, J. M. (2000). Hybrid origins of honeybees from italy (Apis mellifera ligustica) and sicily (A. m. sicula). *Molecular Ecology, 9*, 907–921.

Ruttner, F. (1988). Honeybees of the Western Mediterranean. In *Biogeography and Taxonomy of Honeybees* (pp. 228–239). Berlin Heidelberg, Berlin, Heidelberg: Springer.

Attanzio, A., Tesoriere, L., Allegra, M., & Livrea, M. A. (2016). Monofloral honeys by Sicilian black honeybee (Apis mellifera ssp. sicula) have high reducing power and antioxidant capacity. *Heliyon, 2*, e00193.

Mannina, L., Sobolev, A. P., Di Lorenzo, A., Vista, S., Tenore, G. C., & Daglia, M. (2015). Chemical composition of different botanical origin honeys produced by sicilian black honeybees (Apis mellifera ssp. sicula). *Journal of Agricultural and Food Chemistry, 63*, 5864–5874.

Bambina, P., Malvano, F., Cinquanta, C., Albanese, D., Cirrito, A., Mazza, F., & Corona, O. (2023). Qualitative characteristics of four Sicilian monofloral honeys from Apis mellifera ssp. sicula. *Italian Journal of Food Science*, 35.

Lo Dico, G. M., Ulrici, A., Pulvirenti, A., Cammilleri, G., Macaluso, A., Vella, A., Giaccone, V., Lo Cascio, G., Graci, S., Scuto, M., Trovato Salinaro, A., Calabrese, V., Lo Dico, R., & Ferrantelli, V. (2019). Multivariate statistical analysis of the polyphenols content for the discrimination of honey produced in Sicily (Southern Italy). *Journal of Food Composition and Analysis, 82*, Article 103225.

Coniglio, M. A., Faro, G., Giammanco, G., Pignato, S., & Marranzano, M. (2013). Antimicrobial potential of Sicilian honeys against commensal Escherichia coli and pathogenic Salmonella serovar infantis. *Journal of Preventive Medicine and Hygiene, 54*, 223–226.

Lang, R.E.D. Report on the Food Crisis, Fraud in the Food Chain and the Control Thereof. Tech. Report No. 2013/2091(INI) (Committee on the Environment, Public Health and Food Safety, European Parliament, 2013).

Se, K. W., Wahab, R. A., Yaacob, S. N. S., & Ghoshal, S. K. (2019). Detection techniques for adulterants in honey: Challenges and recent trends. *Journal of Food Composition and Analysis, 80*, 16–32.

Calle, J. L. P., Punta-Sánchez, I., González-de-Peredo, A. V., Ruiz-Rodríguez, A., Ferreiro-González, M., & Palma, M. (2023). Rapid and automated method for detecting and quantifying adulterations in high-quality honey using Vis-NIRs in combination with machine learning. *Foods, 12*.

Moore, J. C., Spink, J., & Lipp, M. (2012). Development and application of a database of food ingredient fraud and economically motivated adulteration from 1980 to 2010. *Journal of Food Science, 77*, R118–R126.

Dou, X., Zhang, L., Yang, R., Wang, X., Yu, L., Yue, X., Ma, F., Mao, J., Wang, X., Zhang, W., & Li, P. (2023). Mass spectrometry in food authentication and origin traceability. *Mass Spectrometry Reviews, 42*, 1772–1807.

Bocian, A., Buczkowicz, J., Jaromin, M., Hus, K. K., & Legáth, J. (2019). An effective method of isolating honey proteins. *Molecules, 24*.

Afzaal, M., Saeed, F., Hussain, M., Shahid, F., Siddeeg, A., & Al-Farga, A. (2022). Proteomics as a promising biomarker in food authentication, quality and safety: a review. *Food Science & Nutrition, 10*, 2333–2346.

Ortea, I., O'Connor, G., & Maquet, A. (2016). Review on proteomics for food authentication. *Journal of Proteomics, 147*, 212–225.

Barik, S. K., Banerjee, S., Bhattacharjee, S., Das Gupta, S. K., Mohanty, S., & Mohanty, B. P. (2013). Proteomic analysis of sarcoplasmic peptides of two related fish species for food authentication. *Applied Biochemistry and Biotechnology, 171*, 1011–1021.

Wisniewski, J. R., Zougman, A., Nagaraj, N., & Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nature Methods, 6*, 359–362.

Rappsilber, J., Ishihama, Y., & Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Analytical Chemistry, 75*, 663–670.

Tyanova, S., & Cox, J. (2018). Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. *Methods in Molecular Biology, 1711*, 133–148.

Kursa, M. B., & Rudnicki, W. R. (2010). Feature Selection with the Boruta Package. *Journal of Statistical Software, 36*, 1–13.

Bader, J. M., Geyer, P. E., Müller, J. B., Strauss, M. T., Koch, M., Leypoldt, F., Koertvelyessy, P., Bittner, D., Schipke, C. G., Incesoy, E. I., Peters, O., Deigendesch, N., Simons, M., Jensen, M. K., Zetterberg, H., & Mann, M. (2020). Proteome profiling in cerebrospinal fluid reveals novel biomarkers of Alzheimer's disease. *Molecular Systems Biology, 16*, e9356.

Geyer, P. E., Holdt, L. M., Teupser, D., & Mann, M. (2017). Revisiting biomarker discovery by plasma proteomics. *Molecular Systems Biology, 13*, 942.

Tenore, G. C., Ritieni, A., Campiglia, P., & Novellino, E. (2012). Nutraceutical potential of monofloral honeys produced by the Sicilian black honeybees (Apis mellifera ssp. sicula). *Food and Chemical Toxicology, 50*, 1955–1961.

Naqvi, F., Dastagir, N., & Jabeen, A. (2023). Honey proteins regulate oxidative stress, inflammation and ameliorates hyperglycemia in streptozotocin induced diabetic rats. *BMC Complementary Medicine and Therapies, 23*, 14.

Dickel, F., Bos, N. M. P., Hughes, H., Martín-Hernández, R., Higes, M., Kleiser, A., & Freitak, D. (2022). The oral vaccination with Paenibacillus larvae bacterin can decrease susceptibility to American Foulbrood infection in honey bees-A safety and efficacy study. *Frontier Veterinary Science, 9*, Article 946237.

First bee vaccine. *Nature* Biotechnology 41 (2023), 163-163.

Salmela, H., Amdam, G. V., & Freitak, D. (2015). Transfer of Immunity from Mother to Offspring Is Mediated via Egg-Yolk Protein Vitellogenin. *PLoS Pathogens, 11*, e1005015.

Bong, J., Middleditch, M., Loomes, K. M., & Stephens, J. M. (2021). Proteomic analysis of honey. Identification of unique peptide markers for authentication of NZ mānuka (Leptospermum scoparium) honey. *Food Chemistry, 350*, Article 128442.