Classification of foods by volatile molecules

Doctoral (PhD) thesis

Zsigmond Papp

Clinical Medical Sciences Doctoral School
Molecular Epidemiology of Tumours Doctoral Program

Doctoral School Leader: Prof. Dr. Lajos Bogár
Program Leader: Prof. Dr. István Kiss
Tutor: Dr. Zoltán Gyöngyi

University of Pécs
Clinical Medical Sciences Doctoral School
Department of Public Health Medicine

Pécs, 2025

1. Introduction

1.1. Volatile organic compounds

Volatile organic compounds (VOCs) are naturally occurring agents with different chemical structures, low molecular mass, and evaporating under ambient conditions (Rowan, 2011). Volatilomics is a metabolomics field that aims to detect, describe and quantify the VOCs in biological systems (Lytou et al., 2019; Kaldeli et al., 2024). The main instruments of volatilomics are gas chromatography (GC) and mass spectrometry, MS), but it can also apply multisensor systems like the electronic nose (Lytou et al., 2019) and the gas chromatography–ionmobility spectrometry (GC-IMS) (Wang et al., 2020).

An enormous amount of data is gathered during the volatilomic measurements, which can be analysed using multivariate data analysis methods. The applied statistical methods can be principal component analysis (PCA), linear discriminant analysis (LDA), and other methods (Lytou et al., 2019).

VOCs play a significant role in shaping the aromatic characteristics of different food materials (Rowan, 2011; Lytou et al., 2019; Kaldeli et al., 2024). To determine an aroma, a volatile profile can be established, which holds valuable information about the quality of certain foods (Lytou et al., 2019; Kaldeli et al., 2024). One goal of volatilomics is to detect and analyse these profiles, and it is gaining popularity because of its fast, cost-effective and noninvasive methods (Kaldeli et al., 2024).

1.2. Propolis

Propolis is a resinous material consisting of pollen, beeswax, plant and buds secreted by bees, and enzymes secreted by the bees' salivary glands (Sforcin, 2016; Anjum et al., 2019). A general description of propolis is difficult since its appearance and chemical structure may vary depending on its place of origin and date of collection (Sforcin, 2016; Anjum et al., 2019; Kasote et al., 2022).

More and more phytogenic compounds are identified in propolis. Currently, 800 of them are documented, most of them being polyphenols (Kasote et al., 2022). The primary polyphenols found in propolis are phenolic acids and their esthers, and flavonoids (Sforcin, 2016; Kasote et al., 2022). Other ingredients found in propolis are terpenes, chalcones, amino acids, sugars, alcohols, ketones, micro- and macroelements

and vitamins (Sforcin, 2016; Anjum et al., 2019; Przybyłek és Karpiński, 2019; Kasote et al., 2022).

Propolis's chemical composition results in a very complex mechanism of action. The most studied effect of propolis is the antimicrobial effect, but its anticancer, immunomodulatory, antiinflammatory, and wound-healing effects are also described (Sforcin, 2016). It is suspected that propolis' effectiveness against *Candida albicans* (*C. albicans*) results from damaging the fungal cell membrane (Corrêa et al., 2020).

Since the standardisation of propolis is difficult, it cannot gain wider access in conventional medicine (Sforcin, 2016). Even if there have been factors determined regarding the production and quality of propolis, the lack of quality standards and legal background hinders the widespread use of propolis-based products in the world market (Peixoto et al., 2022). Examination, classification and chemical assessment of different types of propolis are necessary to determine their mechanism of action and to produce new propolis-based medicines (Sforcin, 2016). There is some evidence of using the electronic nose to analyse different propolis samples, yet it remains largely unexplored.

1.3. Plant-based drinks

Plant-based drinks are liquid products intended for human consumption made from plants or plant parts with aqueous extraction (Pérez-Rodríguez et al., 2023). Plant-based drinks are often incorrectly referred to as „plant milks," which is not legally accepted, with a few exceptions (Angelino et al., 2020; Pérez-Rodríguez et al., 2023).

Plant-based drinks are gaining popularity. The estimated value of the diary-alternative market was 27.3 billion American dollars in 2022 (Pérez-Rodríguez et al., 2023). Manufacturers are producing newer and newer products to support the increasing demand (Angelino et al., 2020). Because of this, more and more types of plant-based drinks are becoming available, differing in taste, consistency, ingredients, and plant source (Xie et al., 2023).

The composition of plant-based drinks may vary depending on the type, in terms of trace elements, micronutrients and macronutrients (Walther et al., 2022). Furthermore, during the formulation of plant-based drinks, the manufacturer may add additives (colourants, flavourings, preservatives, stabilisers, thickeners), resulting in compositional differences (McClements et al., 2019).

Plant-based drinks are colloidal systems, and can be destroyed by physical effects, chemical reactions and microbiological processes (McClements et al., 2019). Also, the quality and stability of individual products can vary greatly (Patra et al., 2021). Inadequate quality of raw materials and insufficient sterilisation processes can cause the colonisation and proliferation of microorganisms in plant-based drinks (Xie et al., 2023). Furthermore, plant-based drinks can contain substances that cause severe allergic reactions, either through food fraud or cross-contamination (Ning et al., 2024). According to these possible scenarios, there is a legitimate need for strict quality control of plant-based drinks before, during and after production.

The instruments used for quality control of plant-based drinks are high-performance liquid chromatography (HPLC), GC and MS (Basile et al., 2023). These methods can be supplemented with lower-cost, complementary analytical methods (Basile et al., 2023). According to current data, the use of volatilomics and electronic noses in the quality control of plant-based drinks is an unexplored area.

2. Objectives

2.1. Classification of propolis samples based on their antifungal activity

Our study aimed to determine the groupability of propolis samples from 4 different Hungarian settlements based on their previously determined antifungal activity using an electronic nose and GC-MS.

2.2. Classification of plant-based drinks based on composition, type and manufacturer

In our study, we aimed to analyse 111 plant-based drinks using GC-IMS and an electronic nose. Our aim was to investigate how accurately the plant-based beverages can be separated from each other based on composition, type and manufacturer.

3. Materials and methods

3.1. Classification of propolis samples based on their antifungal activity

3.1.1. Collection of raw propolis samples and preparation of ethanol extracts

For our study, raw poplar-type propolis samples were collected in 2015 from four different areas in Hungary: Csikóstőttős (CS), Héhalom (HE), Somogybabod (SO), and Szolnok (SZ). The samples were extracted after grinding: 100 g of the propolis sample was extracted with 450 ml of 80% ethanol solution for 30 minutes in a water bath at 70 °C. The ethanol extracts were sterilised on a 0.22 μm pore size filter (Millipore, Burlington, MA, United States of America) to produce stock solutions with a 222.2 mg/ml concentration. All ethanolic propolis extracts (EPEs) were stored in the dark at 4 °C (Alencar et al., 2007).

3.1.2. Antifungal Test

The susceptibility of C. albicans ATCC 44829 to the four different EPKs was determined using the broth microdilution method of the CLSI M27-A2 standard (Torres-Rodríguez & Alvarado-Ramírez, 2007), maintaining the stock solution at 80% (v/v) ethanol concentration. 80% (v/v) ethanol was used as a control. The stock solution constituted 1% of the medium, resulting in a final ethanol concentration of 0.8% (v/v) in both the treated and control solutions. The cell culture was grown on yeast extract-peptone-glucose agar plates (0.5 % (w/v) yeast extract, 2 % (w/v) glucose, 1 % (w/v) bacteriological peptone, 2 % (w/v) agar, supplemented with 25 mg/l adenine; pH 5.6). In total, two-fold serial dilutions of the EPKs (6.25 - 400 g/ml) were mixed 1:1 with fungal suspensions in RPMI-1640 medium buffered with 0.165 mol/l 3-(N-morpholino)-propanesulfonic acid solution (pH 7.0). The cell number in the mixture was adjusted to a final concentration of 2.5×10^3 cells/ml in 96-well cell culture plates (REF3595; Costar®, Kennebunk, ME, USA) and incubated at 35 °C for 48 h. The solvent concentration in each well was kept constant at 1%. The absorbance of the suspensions proportional to growth was measured at 595 nm using a Multiskan EX plate reader (ThermoFisher Scientific, Waltham, MA, USA). Each experiment was repeated three times. The minimum inhibitory concentration (MIC) was defined as the lowest concentration that caused 80% growth inhibition (Boisard et al., 2015).

3.1.3. GC-MS analysis of propolis samples

The trimethylsilyl ether (TMS-ether) derivatives of ethanolic propolis extracts were subjected to GC-MS analysis. In summary, approximately 2.2 mg of freeze-dried EPK was mixed with 50 µl of dry pyridine (Merck, Budapest, Hungary) and 75 µl of N-methyl-N-(trimethylsilyl)-trifluoroacetamide (MSTFA) solution (Merck, Hungary), and the mixture was heated at 80 °C for 20 min. GC-MS analysis of the samples was performed using a QP-2020 GC-MS system (Shimadzu, Duisburg, Germany) equipped with a 30 m long, 0.25 mm inner diameter, one µm film thickness DB-5ms (Agilent, Santa Clara, CA, United States) capillary column. The temperature was programmed from 100 °C to 320 °C with a rate of 5 °C/min. Helium was used as the carrier gas with a flow rate of 40 cm/s. The split ratio was 1:20, the injector temperature was 280 °C, and the interface temperature was 320 °C. For MS, the electron ionisation (EI) ion source temperature was 230 °C, the ionisation voltage was 70 eV, and the delay time was 4.0 min. Data were recorded in scan mode with an event time of 0.3 s, between 4.5 and 60 min, in the m/z range of 45-600. To calculate the retention index values, a separation of a hexane solution of C7-C33 n-alkanes (Restek, Bellefonte, PA, USA) was performed under the conditions described above. For the identification of individual compounds by GC-MS, the obtained mass spectra were individually compared with the libraries of the NIST 17 and Smart Metabolites Database® databases, using the internal library search algorithm of the Shimadzu GC-MS Solutions V.4.45 program (Shimadzu, Duisburg, Germany).

3.1.4. Classification of propolis samples using an electronic nose

Propolis samples were separated based on their region of origin. A stock solution was prepared from each propolis sample: 15 µl EPK, 34 µl 96% (v/v) ethanol, and 4150 µl distilled water were mixed and homogenised to prepare stock solutions with a concentration of 800 µg/ml. 8 × 500 µl of each stock solution was measured into 20 ml glass vials. A control stock solution of 83 µl 96% (v/v) ethanol and 7920 µl distilled water was used as a control. 15 × 500 µl of this control stock solution was measured into 20 ml glass vials for HS measurements.

The NeOse Pro electronic nose system (Aryballe Technologies, Grenoble, France) was used for the measurements, which is an opto-electronic sensor array system using a sensor composed of 63 non-specific proteins printed on a gold layer (Brenet et

al., 2018). Dynamic measurements were performed with the following parameters: pump flow rate 40 ml/min; number of frames per second 20; ambient temperature 29 °C; core temperature 44 °C; humidity 25%. First, the ethanol control samples were measured, and then the propolis samples were measured with the same settings.

3.1.5. Statistical analysis

In case of the antifungal activity of propolis samples, each experiment was performed in triplicates, and the data were presented as mean ± standard deviation (SD). The data were first analysed by analysis of variance (ANOVA) for multiple comparisons between groups, then subjected to Dunn's post hoc test. The analysis results and graphs were prepared using OriginPro 2016 and Past 3.1 software. Differences between samples were considered significant at $p < 0.05$ (* $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$). To determine the significant differences between the antifungal activities of propolis samples, the different survival percentages were compared with the values of the non-antifungal sample (SO). When comparing with the SO sample, the lowest significance level was indicated for the other three samples (SZ, HE and CS) at all concentrations.

To compare propolis samples by GC-MS, the distribution ratio of the compounds identified in the non-antifungal sample (SO) was calculated and then compared with the average distribution ratio of the corresponding compound found in all antifungal samples (SZ, HE and CS).

During the measurements of the propolis samples with the opto-electronic nose, the total of the signals registered by the 63 individual sensors of the instrument were subjected to LDA analysis using the IBM SPSS Statistics for Windows software version 26 (IBM, Armonk, NY, United States of America), where Fisher's coefficient and Mahalanobis distance were used for stepwise analysis.

3.2. Classification of plant-based drinks based on composition, type and manufacturer

3.2.1. Collection of plant-based drink samples

A total of 111 plant-based drinks were purchased from local shopping centres (ALDI, DM, EcoFamily, Penny Market, SPAR, TESCO; Pécs, Hungary; Auchan; Szekszárd, Hungary). Until the instrumental analysis, the samples were frozen and stored at -80 °C to preserve their chemical nature. The tested samples were products from 13 different brands: Adez, Alnatura, Alpro, DMbio, Happy, Isola, Joya, Koko, My Bio, Natur Aktiv, Plant Pro, Riso Scotti, and The Bridge. The samples were divided into seven groups based on their plant source: cashew, coconut, almond, rice, soy, spelt, and oat.

The samples were divided into conventional and organic groups based on their preparation method. Organic products are produced using methods consistent with organic farming, prohibiting artificial and synthetic pesticides, fertilisers, and genetically modified ingredients (Giampieri et al., 2022). If the manufacturer stated on the packaging of the plant-based drink that the product was organic, it was classified as organic. Otherwise, the product was classified as conventional. We also created a specific subgroup (‘Barista’) for plant-based drinks intended for barista use, which was also indicated on the packaging. Products subjected to a special production method (roasting) were also placed in a separate group.

3.2.2. GC-IMS analysis of plant-based drinks

All plant samples were stored in a freezer at -80 °C until analysis. Before the analysis, the samples were thawed at room temperature and carefully shaken until the solid phases disappeared. The sample containers were opened only briefly, and the air gap volume was kept as low as possible. 1 ml sample from each container was measured into 20 ml glass vials with plastic caps. Based on the results of a previous study (Manousi & Zachariadis, 2019), 500 mg of sodium chloride was added to each sample to facilitate the movement of the analytes into the headspace.

Based on the results obtained during the measurements, the following temperatures were set for further measurements: 40 °C for GC, 70 °C for IMS, and 95 °C for incubation. The samples were incubated for 20 minutes at 95 °C. In the first 10 minutes, the vial caps were slightly loosened to avoid the risk of explosion, and in the

next 10 minutes, they were tightened entirely on the vials. No control samples were used during the GC-IMS measurements.

The BreathSpec GC-IMS device (Gesellschaft für Analytische Sensorsysteme, G.A.S., GmbH, Dortmund, Germany) was used for the measurements. Its central element (G.A.S, Dortmund, Germany) is equipped with a wide-bore GC column (MXT-WAX 30 m × 0.53 mm, RESTEK, United States of America).

From the headspace of each sample, 1 ml was measured using a heated (95 °C), 5 ml plastic syringe with a 51 mm needle, of which 200 µl was added to the heated (95 °C) splitless injector. After injection, the first separation of the analytes took place on the GC column, and from there the eluate was sent to the second separation by the IMS, which was equipped with a tritium ionising radioactive source (5000 eV) and a 9.8 cm long drift tube. The drift gas flow rate was 150 ml/min, and the pressure was 0.712 kPa.

Software analysis of the measured volatile components was performed using the VOCal program (Gesellschaft für Analytische Sensorsysteme GmbH, G.A.S.; Dortmund, Germany). Using the VOCal software, 58 different areas were selected from the signals obtained on the chromatograms to represent the volatile components examined. Area 66 served as an internal reference signal recorded by the GC-IMS, which was used to normalise the signal strength of the 58 examined areas. The ratios obtained by normalization were used for further analyses.

3.2.3. Analysis of plant-based drinks with the NeOse Pro electronic nose

The “Happy” almond-based sample, which had a neutral taste and smell and contained only 1% almond, was used as a control solution for all measurements with the electronic nose. All samples, including the controls, were stored at -80 °C until analysis. Before analysis, the samples were thawed at 4 °C. After thawing, 1 ml of each sample was measured into seven labelled vials. The sealed vials were incubated for 40 min at 95 °C and then allowed to cool to room temperature (25 °C) for 20 min.

All samples, including the controls, underwent a comprehensive analysis using the NeOse Pro electronic nose device (Aryballe Technologies, Grenoble, France). NeOse Pro, as previously described, is an opto-electronic sensor array system using a gold-printed sensor consisting of 63 non-specific proteins (Brenet et al., 2018).

Dynamic measurements were performed with the following parameters: pump flow rate 50 ml/min, frame rate 25 per second, and ambient temperature 32 °C.

The analysis of one sample took approximately two minutes, ensuring the analysis's accuracy and high quality. Seven samples of each material were measured. Control samples were measured at the beginning and at the end of each measurement phase, and at two intermediate times. The measurement results of the control samples served as a reference point during the calibration of the results. The first two measurements of each sample, including the controls, were discarded during the evaluation of the results, as these measurements were necessary to saturate the polytetrafluoroethylene membrane (32 mm diameter, 0.45 µm RephiQuick Syringe Filter; RephiLe Bioscience Ltd., Zhejiang, China).

3.2.4. Statistical analysis

The average values of the control sample results were subtracted from the measurement results of each tested sample, thereby correcting the data shift caused by the electronic nose drift phenomenon. The absolute value of the lowest negative value was added to each value of the corrected results to eliminate negative values in the dataset.

The classification of the samples measured with GC-IMS and electronic nose was performed using PCA and LDA procedures. The PCA score plots were prepared using ClustVis, a web-based tool for clustering multivariate data. The ellipses are defined by the 95% confidence level (Metsalu & Vilo, 2015). During LDA, Fisher's coefficient and Mahalanobis distance were used to perform stepwise analysis, using IBM SPSS Statistics for Windows, Version 28 software (IBM, Armonk, NY, United States of America).

The samples (n = 111) were compared by brand (Alpro: almond, cashew, rice, sugarfree soy; DMbio: almond, coconut, rice, soy, spelt), type (barista: almond, coconut, oat) and plant source (almond: roasted, barista, traditional, organic; coconut: Adez, Joya, Koko, Naturaktiv, Happy; traditional rice: Alpro traditional, Alpro sweetened, PlantPro, Happy; organic rice: Auchan, Isola, Riso Scotti, DMbio, TheBridge, MyBio).

The data and figures used in the PCA and LDA analyses, including the gallery plot panels, underwent no standardisation or normalisation within samples.

4. Results

4.1. Classification of propolis samples based on their antifungal activity

The cytotoxicity of the EPKs was characterized by determining the antifungal susceptibility of *C. albicans* 48 h after performing the microdilution method. All extracts showed concentration-dependent susceptibility (**Figure 1**).

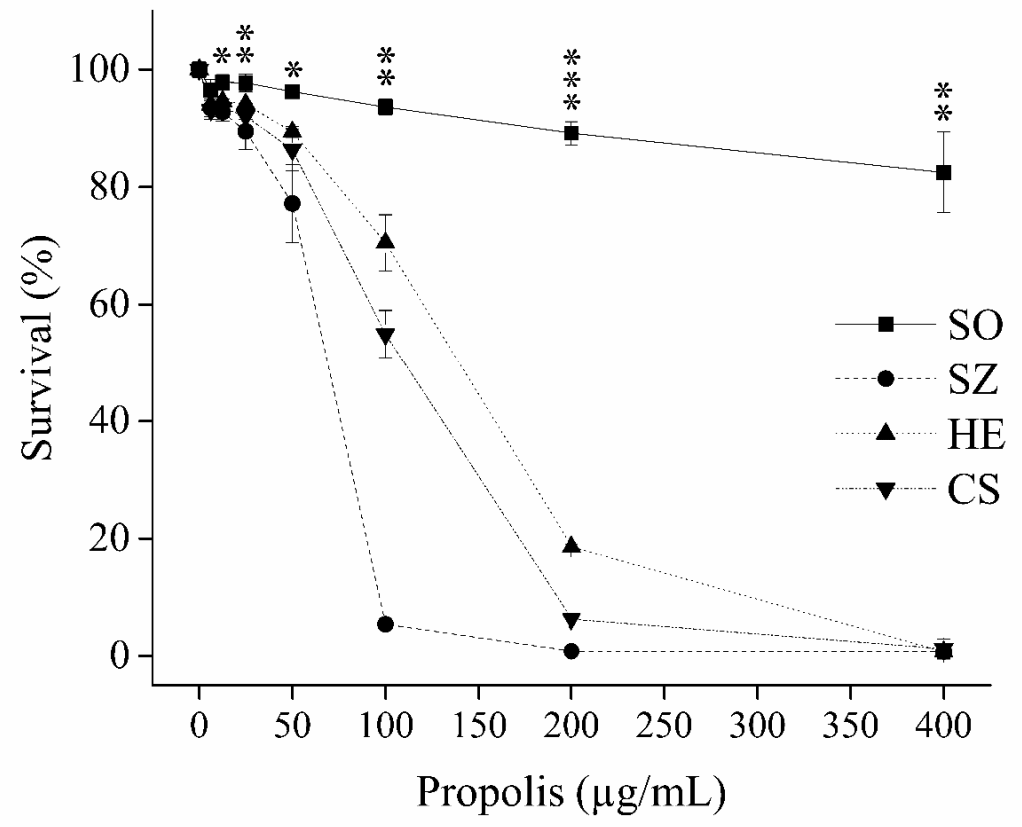

Figure 1. Inhibition curve. Survival of *C. albicans* ATCC 44829 after 48 h incubation at 35 °C (y axis) in solutions of different ethanolic propolis extracts (EPEs) at concentrations of 0, 6.25, 12.5, 25, 50, 100, 200 and 400 µg/ml (x axis).

The SZ, HE, and CS samples had vigorous antifungal activity, with MIC_{80} and IC_{50} values ranging from 100–200 µg/ml and 72–134 µg/ml, respectively (**Table I**). The SO sample showed significantly weaker antifungal activity compared to the other EPKs, with cell growth inhibition not exceeding 18% even at a concentration of 400 µg/ml.

Table I. Inhibition data. The minimum inhibitory concentration (MIC) was defined as the lowest concentration that resulted in 80% inhibition of growth. The 50% inhibitory concentration (IC_{50}) was also determined using four different ethanolic propolis extracts on *C. albicans* strains. The values are given in g/ml.

EPK (µg/ml)	SO	SZ	HE	CS
IC_{50}	No	72	134	108
MIC_{80}	No	100	200	200

During GC analysis, a total of 148 components were successfully identified in the EPKs, of which 134 were in the HE sample, 115 in the CS sample, 127 in the SZ sample, and 94 in the SO sample. Based on their chemical structures, the identified compounds were alcohols, aliphatic and aromatic aldehydes, alkanes, amino acids, aliphatic and aromatic carboxylic acids, essential oils, esters, fatty acids, fatty alcohols, flavonoids, ketones, polyphenols, sugars (monosaccharides, disaccharides), sugar acids, sugar alcohols, phenols, phenolic acids, terpenes, terpene alcohols, vitamin B_6, and compounds with other structures such as heptalene and urea.

In the samples with more potent antifungal activity, the five most abundant compounds were chrysin (25.65%; polyphenolic flavone/flavonoid), genistein (21.69%; isoflavone), ethyl gallate (6.62%; carboxylic acid), caffeic acid (5.69%; cinnamic acid derivative), and caffeic acid ethyl ester (4.22%; hydroxycinnamic acid derivative/polyphenol). In the sample with weaker antifungal activity, the five most abundant compounds were chrysin (19.01%; polyphenolic flavone/flavonoid); D-fructofuranose pentakis-(trimethylsilyl)-ether (isomer, psicose) (16.46%; monosaccharide), genistein (13.83%; isoflavone), sucrose (10.54%; disaccharide), and α-D-glucopyranose (6.44%; monosaccharide).

The proportion of compounds found in samples with stronger and weaker antifungal activity was also determined. **Table II.** contains the compounds whose concentration is at least five times higher in samples with more potent antifungal activity (HE, CS, SZ) than those with weaker antifungal activity (SO).

Table II. Relative dominance of compounds in samples with higher antifungal activity. The ratios of chemical concentrations (high-antifungal-activity/low-antifungal-activity sample) show more than 5 times higher concentration in samples with higher antifungal activity. The original data are from the results of the GC measurement and were calculated from the area values.

Compound	>5-Times Higher Concentration in Antifungal Samples [Times]
11,14-Eicosadienoic acid	16,84
Ferulic acid	14,87
Phenylpropionic acid	13,13
Farnesol	12,99
Cinnamic acid	12,97
Urea	12,16
Benzoic acid	11,66
17-octadecynoic acid	10,96
α/β-eudesmol	10,90
Vanillin	9,97
Ricinolenic acid	8,88
4-methoxycinnamic acid	8,64
Cis/trans-p-coumaric acid	8,60
Benzyl alcohol	8,21
Cis/trans-p-coumaric acid	8,10
Hexadecyl-p-coumarate	7,30
1,3,5-trihydroxybenzene	7,03
Coniferyl aldehyde	7,02
Isoferulic acid	6,89
Pyridoxine	6,79
Methyl ferulate	6,55
Propionic acid	6,19
α/β-eudesmol	5,67
Methyl 2-amino-3-hydroxybenzoate	5,46
Caffeic acid	5,07
Caffeic acid ethyl ester	5,06

The results of the electronic nose measurements were analysed using IBM SPSS Statistics software. Three groups were created, one for ethanol controls, one for samples with higher antifungal activity, and one for samples with lower antifungal activity. Discriminant analysis was used to examine the separation of the groups from each other. **Figure 2** illustrates the marked separation of the different samples: the control samples, the samples with higher antifungal activity and those with lower antifungal activity are clearly separated and form separate group centroids. Based on the result of the stepwise method, 98.4% of the grouped cases were correctly classified; however, using cross-validation, 95.2% of the grouped cases were correctly classified (**Table III**).

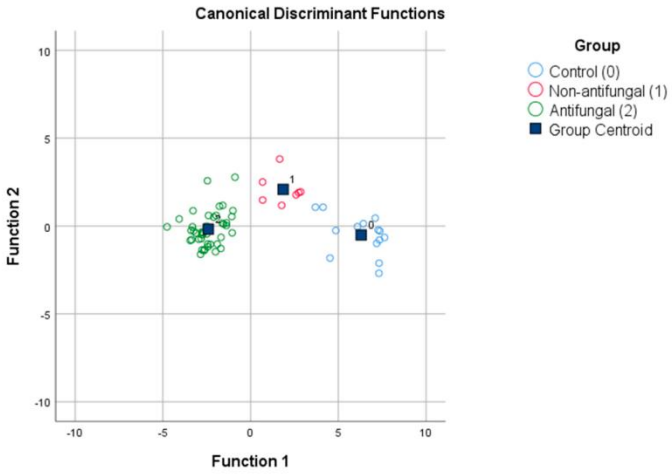

Figure 2. Linear discriminant analysis (LDA). Separation of EPK samples based on their low or high antifungal activity, and from the control ethanol solvent, using LDA. The original data comes from the NeOse Pro opto-electronic nasal sensor array measurements, containing 63 different peptide sequences.

Table III. Classification results. The table shows the original and cross-validated values of propolis samples with low antifungal activity ("Non-antifungal") and high antifungal activity ("Antifungal"), as well as the solvent control ("Control"). The samples were classified using linear discriminant analysis (LDA).

Classification Results[a,c]						
			Predicted Group Membership			
		Group	Control	Non-Antifungal	Antifungal	Total
Original	Count	Control	13	1	0	14
		Non-Antifungal	0	7	0	7
		Antifungal	0	0	42	42
	%	Control	92,9	7,1	0,0	100,0
		Non-Antifungal	0	100,0	0,0	100,0
		Antifungal	0	0,0	100,0	100,0
Cross-validated[b]	Count	Control	12	2	0	14
		Non-Antifungal	0	7	0	7
		Antifungal	0	1	41	42
	%	Control	85,7	14,3	0	100,0
		Non-Antifungal	0	100,0	0	100,0
		Antifungal	0	2,4	97,6	100,0

a. 98.4% of original grouped cases correctly classified.
b. Cross-validation was undertaken only for those cases in the analysis. In cross-validation, each case is classified by the functions derived from all cases other than that case
c. 95.2% of cross-validated grouped cases correctly classified.

4.2. Classification of plant-based drinks based on composition, type and manufacturer

Figures 3-4. show the PCA results of the GC-IMS and electronic nose measurements. The PCA results show more overlap and less separation for the GC-IMS data than for the electronic nose data.

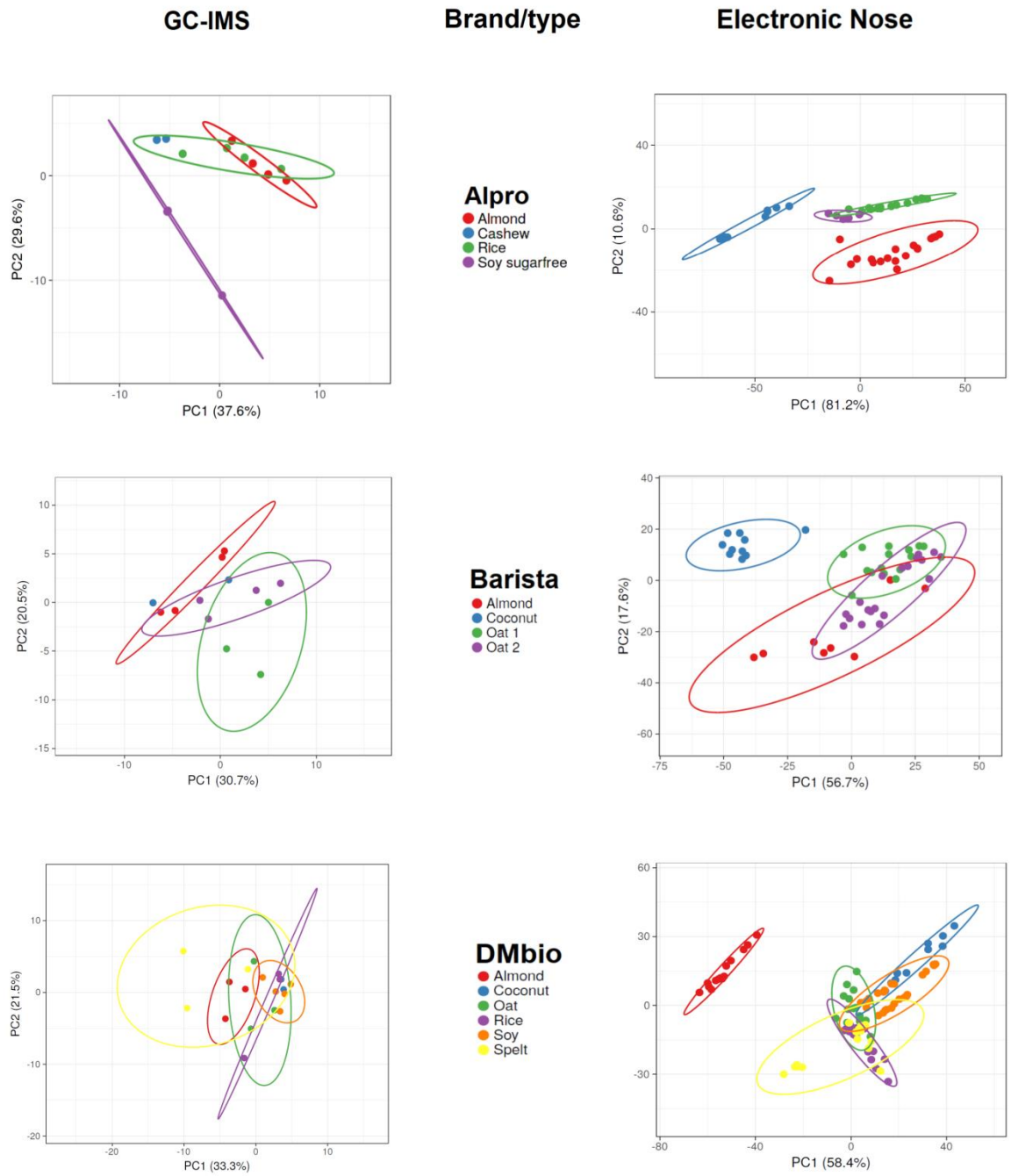

Figure 3. Principal component analysis results from gas chromatography-coupled ion mobility spectroscopy (GC-IMS) and electronic nose measurements for plant-based drinks by brand and type.

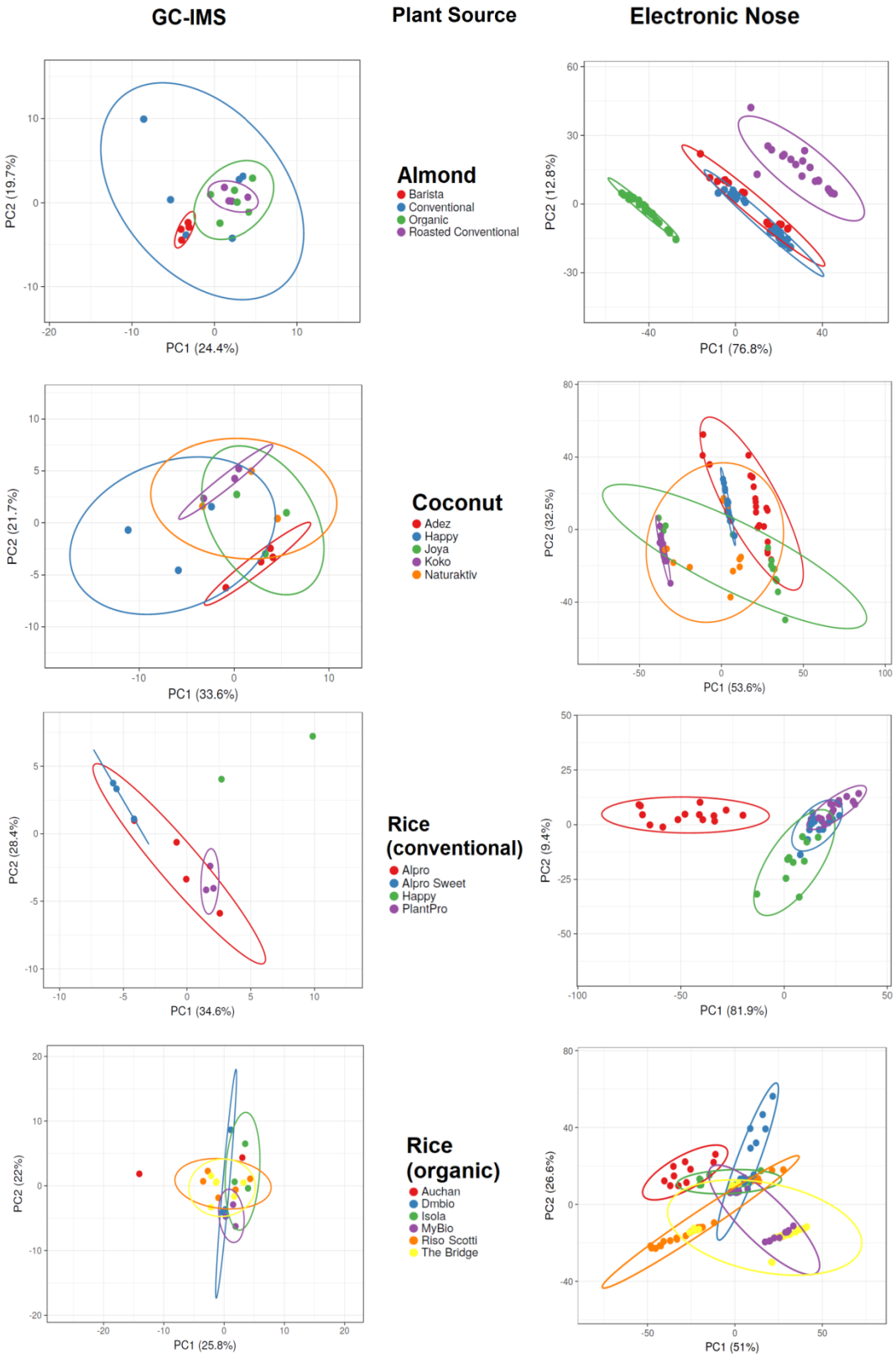

Figure 4. Principal component analysis results from gas chromatography-coupled ion mobility spectroscopy (GC-IMS) and electronic nose measurements for plant-based drinks by plant source.

In the case of the Alpro samples, GC-IMS allowed the complete separation of sugar-free soy-based samples from other Alpro samples and also separated cashew-based samples from almond-based samples. However, there was an overlap between almond-based and rice-based samples and between cashew-based and rice-based samples. Results with the electronic nose showed complete separation of almond-based and cashew-based samples, but there was overlap between rice-based and sugar-free soy-based samples.

In the barista samples, GC-IMS allowed the separation of almond-based samples from one of the two groups of oat-based samples. However, neither group of samples was completely separated from all others. In contrast, using the electronic nose, the coconut-based samples were completely separated from the other barista samples. However, overlaps were seen between the almond and oat samples from both groups.

When examining the DMbio samples by GC-IMS, none of the groups were completely separated from all the other samples. However, separation of the almond-based from rice-based samples was observed. However, overlaps were observed between all the samples when using GC-IMS. The electronic nose was able to completely separate the almond-based samples from the other DMbio samples, as well as the rice-based and coconut-based samples. However, overlaps were observed for all samples except the almond-based samples.

For the almond-based samples, the GC-IMS method separated the barista and roasted (conventional) samples. However, no complete separation was observed when all the samples were examined together. The electronic nose method completely separated the organic and roasted (conventional) samples, but the barista and conventional samples overlapped.

GC-IMS analysis could not completely separate the coconut-based samples, but the Adez and Koko samples were separated. The electronic nose also failed to achieve complete separation for any of the samples, but the Koko and Happy samples, as well as the Koko and Adez samples, were separated from each other.

During the GC-IMS analysis of the traditional rice-based samples, we observed complete separation of the Happy samples, and the sweetened Alpro and PlantPro samples were separated. The electronic nose method achieved complete separation of the Alpro samples, but the Alpro, Happy, and PlantPro samples overlapped with each other.

In the case of the organic, rice-based samples, the GC-IMS method was able to separate the Auchan and MyBio samples from each other, but when all the samples were analysed together, complete separation was not achieved for any of the groups. The electronic nose was able to separate the Auchan samples from the Riso Scotti and DMbio samples. However, when all the samples were analysed together, none of the groups was completely separated from the others.

The summarised LDA results are included in **Table IV**, which compares the original and cross-validated classification results of the groups analysed by GC-IMS and the electronic nose. After cross-validation, the LDA classification results obtained by GC-IMS proved less accurate than the results of the electronic nose. Although the lowest classification accuracy was 15.4% for the Alpro samples, the other accuracy values were 89.5% or higher. However, the electronic nose showed higher accuracy values in all aspects, with the lowest being 96.2%.

Table IV. Percentages of correct classifications by LDA before and after cross-validation. LDA: linear discriminant analysis; GC-IMS: gas chromatography-coupled ion immobilization spectrometry; conv.: conventional; org.: organic

Examined Group	**GC-IMS**		**Electronic Nose**	
	Original Grouped Cases Correctly Classified	*Cross-Validated Grouped Cases Correctly Classified*	*Original Grouped Cases Correctly Classified*	*Cross-Validated Grouped Cases Correctly Classified*
Alpro	100,0 %	15,4 %	100,0 %	100,0 %
Barista	100,0 %	92,3 %	100,0 %	100,0 %
DMbio	100,0 %	89,5 %	100,0 %	100,0 %
Almond	100,0 %	95,0 %	100,0%	100,0 %
Coconut	100,0 %	100,0 %	100,0 %	96,2 %
Rice (conv.)	100,0 %	91,7 %	100,0 %	100,0 %
Rice (org.)	100,0 %	90,9 %	100,0 %	100,0 %

Figures **5-6.** illustrate a visual comparison of LDA results based on GC-IMS and electronic nose measurements.

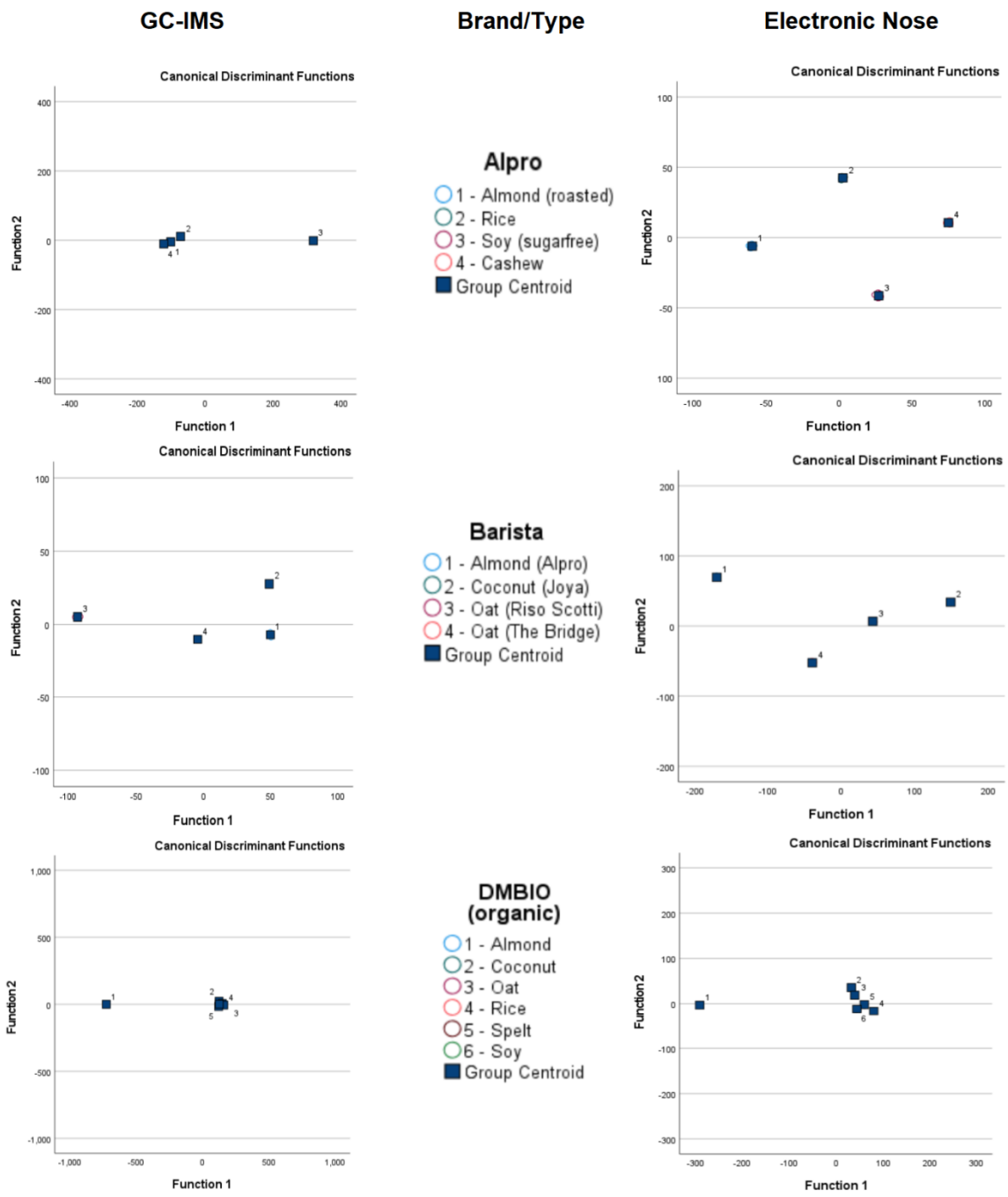

Figure 5. Linear discriminant analysis classification results from gas chromatography ion mobility spectroscopy (GC-IMS) and electronic nose measurements for plant-based drinks by brand and type.

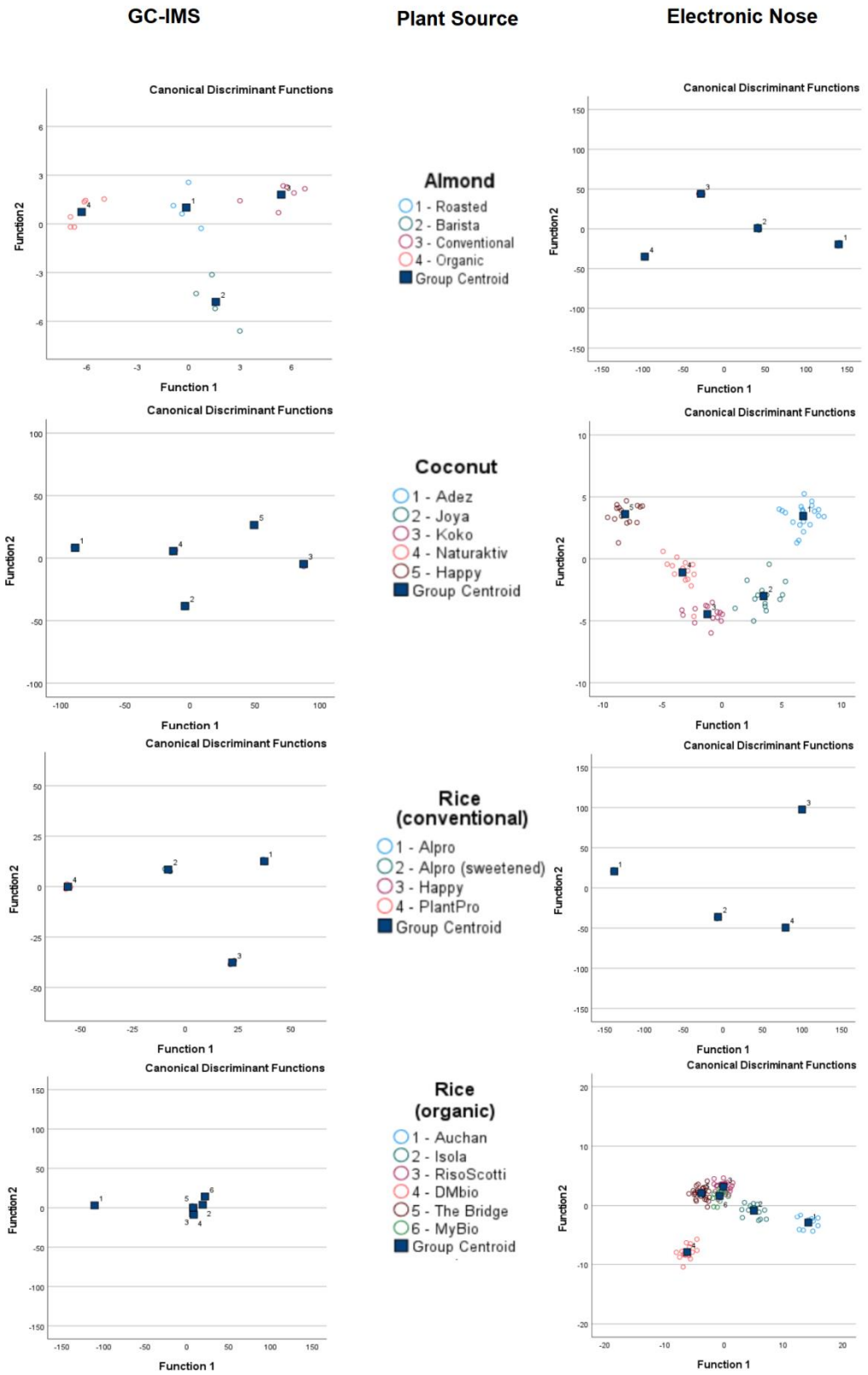

Figure 6. Linear discriminant analysis classification results from gas chromatography ion mobility spectroscopy (GC-IMS) and electronic nose measurements for plant-based drinks based on plant source.

5. Discussion

Our first study aimed to characterize and compare ethanolic solutions of poplar-type propolis from four different, non-overlapping regions of Hungary. One sample showed significantly lower antifungal activity against *C. albicans* compared to the other samples.

During the GC-MS analysis, we found significant compositional differences between antifungal and less antifungal samples. The active growth inhibitory effect of one of the main components of all samples, chrysin, has already been documented for *Candida* species (Agüero et al., 2014). The cinnamic acid derivatives, which have been described to inhibit fungal growth (Sova, 2012), were found in much higher amounts in the samples with higher antifungal activity. In the case of ethyl gallate, an IC_{50} value comparable to fluconazole was measured for Candida species, and a synergistic effect with 3-β-D-glucopyranosyloxysitosterol gallate was described (Dias Silva et al., 2019).

Chrysin, as an antifungal component, was also found in the lower antifungal samples, but at a lower concentration. The genistein content of the sample with low antifungal activity was also lower. However, the effect of genistein on fungal growth is questionable. Samples with low antifungal activity also contained significantly less cinnamic acid derivatives and ethyl gallate, but contained higher amounts of saccharides, which have no antifungal effect reported in the literature.

In our second approach, we compared the proportions of the components of the antifungal and the less antifungal samples. With this method, we identified 26 compounds; however, a significant part of them does not have antifungal activity according to the literature, however, in the case of farnesol (Ramage et al., 2002), vanillin (Boonchird & Flegel, 1982), p-coumaric acid (Shreaz et al., 2013), and methyl ferulate (Perez-Castillo et al., 2020), effective inhibition of biofilm formation of microorganisms, including *C. albicans*, has been reported.

In addition to suggesting new potential target molecules for research on antifungal activity, our primary goal was to identify propolis samples with high antifungal activity. The samples were classified using the NeOse Pro opto-electronic nose and the LDA machine learning algorithm. The opto-electronic nose we tested was able to classify propolis samples based on their antifungal activity with an accuracy of 95% using a protein sensor array and the selected data processing algorithm.

In our second study, we analysed 111 plant-based beverage samples using GC-IMS and an electronic nose, and our goal was to compare the accuracy of the two instruments. In order to reduce the number of dimensions, we compared methods with and without controls. We chose the two most commonly used methods for statistical analyses, PCA and LDA (Jena et al., 2023). Our study evaluated GC-IMS and the electronic nose as promising, fast, efficient and inexpensive methods and compared their ability to discriminate and classify plant-based beverages produced by different manufacturers and containing different ingredients.

Our study is among the first to compare GC-IMS and an electronic nose in separating different plant-based beverage samples. Our results show that most almond samples were correctly classified, with 95-100% accuracy. The low almond content of the samples, ranging from 1% to 7%, may be responsible for the overlaps. Overlaps may also result from mixing product ingredients, such as coconut and rice, or coconut and soybean. Although manufacturers may use different raw materials and protocols during production, different products may contain the same additives, such as sunflower oil, gellan gum, and vitamin B_2. Our results show that GC-IMS could not clearly distinguish between different plant-based beverage samples without fingerprint references to GC-IMS. In contrast, the electronic nose method was much more efficient in differentiating samples. Based on our results, the electronic nose method proved to be a more accurate and faster method than GC-IMS, allowing sample analysis within 2 minutes compared to the 25-minute duration of GC-IMS.

One limitation of the grouping of plant-based beverages is that the GC-IMS and electronic nose measurements were not performed with the same number of samples. The difference in the number of samples per group may bias the results. However, it should be noted that in the case of GC-IMS, the measurement of a single sample takes 25 minutes, while in the case of the electronic nose, it takes only 2 minutes. When implementing the study, we also had to consider how long it takes to perform the analysis. As a result, we could measure a sample fewer times with GC-IMS than with the electronic nose. The other limiting factor is that when evaluating the GC-IMS results, in some cases, we worked with fewer than three samples, which does not allow the determination of a 95% confidence interval.

6. New scientific results

Our studies tested the applicability of the electronic nose and GC-IMS in the quality control of propolis and plant-based drinks samples.

In our first study, four propolis samples from different regions of Hungary were analysed with GC-IMS and an opto-electronic nose. The GC-IMS and the electronic nose results correlated with the previously determined antifungal activity of the different propolis samples. 26 chemical agents were identified with GC-MS, which could be related to the increased antifungal activity. Possible connections were established between the antifungal activity, chemical composition and volatile profiles using the results. Our study suggested using opto-electronic noses in the medical field, expanding the current instrumental options.

Our second study analysed 111 plant-based drink samples with GC-IMS and an electronic nose. Separating the mentioned drinks by plant source, brand and type was investigated. Our research concluded that perfect separation of every plant-based drink was rarely obtainable. However, some particular groups separated from each other. The results with the electronic nose proved to be more accurate. However, this might be due to the different affinity of the two methods for polar molecules, which are abundant in plant-based drinks compared to apolar molecules. According to our results, both methods need to be refined. However, the rapid and accurate results of the electronic nose are remarkable. In the current state, none of these methods will replace GC-MS or HPLC-MS, but they may be useful in complementing their measurements.

7. References

Agüero, M. B., Svetaz, L., Baroni, V., Lima, B., Luna, L., Zacchino, S., Saavedra, P., Wunderlin, D., Feresin, G. E., & Tapia, A. (2014). Urban propolis from San Juan province (Argentina): Ethnopharmacological uses and antifungal activity against *Candida* and dermatophytes. *Industrial Crops and Products*, *57*, 166-173.

Angelino, D., Rosi, A., Vici, G., Dello Russo, M., Pellegrini, N., Martini, D., & SINU Young Working Group. (2020). Nutritional quality of plant-based drinks sold in Italy: the Food Labelling of Italian Products (FLIP) study. *Foods*, *9*(5), 682.

Anjum, S. I., Ullah, A., Khan, K. A., Attaullah, M., Khan, H., Ali, H., Bashir, M. A., Tahir, M., Ansari, M. J., Ghramh H. A., Adgaba, N., & Dash C. K. (2019). Composition and functional properties of propolis (bee glue): A review. *Saudi Journal of Biological Sciences*, *26*(7), 1695-1703.

Basile, T., Mallardi, D., & Cardone, M. F. (2023). Spectroscopy, a tool for the non-destructive sensory analysis of plant-based foods and beverages: A comprehensive review. *Chemosensors*, *11*(12), 579.

Boisard, S., Le Ray, A. M., Landreau, A., Kempf, M., Cassisa, V., Flurin, C., & Richomme, P. (2015). Antifungal and antibacterial metabolites from a French poplar type propolis. *Evidence-Based Complementary and Alternative Medicine*, *2015*(1), 319240.

Boonchird, C., & Flegel, T. W. (1982). In vitro antifungal activity of eugenol and vanillin against *Candida albicans* and *Cryptococcus neoformans*. *Canadian Journal of Microbiology*, *28*(11), 1235-1241.

Brenet, S., John-Herpin, A., Gallat, F. X., Musnier, B., Buhot, A., Herrier, C., Rousselle, T., Livache, T., & Hou, Y. (2018). Highly-selective optoelectronic nose based on surface plasmon resonance imaging for sensing volatile organic compounds. *Analytical Chemistry*, *90*(16), 9879-9887.

Corrêa, J. L., Veiga, F. F., Jarros, I. C., Costa, M. I., Castilho, P. F, de Oliveira, K. M. P., Rosseto, H. C, Bruschi, M. L., Svidzinski, T. I. E., & Negri, M. (2020). Propolis extract has bioactivity on the wall and cell membrane of *Candida albicans*. *Journal of Ethnopharmacology*, *256*, 112791.

Dias Silva, M. J., Simonet, A. M., Silva, N. C., Dias, A. L. T., Vilegas, W., & Macías, F. A. (2019). Bioassay-guided isolation of fungistatic compounds from *Mimosa caesalpiniifolia* leaves. *Journal of Natural Products*, *82*(6), 1496-1502.

Giampieri, F., Luca Mazzoni, L., Danila Cianciosi, D., Alvarez-Suarez, J. M., Regolo, L., Sánchez-González, C., Capocasa, F., Xiao, J., Mezzetti, B., & Battino, M. (2022). Organic vs conventional plant-based foods: A review. *Food Chemistry*, *383*, 132352.

Jena, S. P., Paikaray, B. K., Pramanik, J., Thapa, R., & Samal, A. K. (2023). Classifications on wine informatics using PCA, LDA, and supervised machine learning techniques. *International Journal of Work Innovation*, *4*(1), 58-73.

Kaldeli, A., Zakidou, P., & Paraskevopoulou, A. (2024). Volatilomics as a tool to ascertain food adulteration, authenticity, and origin. *Comprehensive Reviews in Food Science and Food Safety*, *23*(4), e13387.

Kasote, D., Bankova, V., & Viljoen, A. M. (2022). Propolis: chemical diversity and challenges in quality control. *Phytochemistry Reviews*, *21*(6), 1887-1911.

Lytou, A. E., Panagou, E. Z., & Nychas, G. E. (2019). Volatilomics for food quality and authentication. *Current Opinion in Food Science*, *28*, 88–95.

Manousi, N., & Zachariadis, A.G. (2019). Determination of volatile compounds in nut-based milk alternative beverages by HS-SPME prior to GC-MS analysis. *Molecules*, *24*(17), 3091.

McClements, D. J., Newman, E., & McClements, I. F. (2019). Plant-based milks: A review of the science underpinning their design, fabrication, and performance. *Comprehensive Reviews in Food Science and Food Safety*, *18*(6), 2047-2067.

Metsalu, T., & Vilo, J. (2015). ClustVis: a web tool for visualizing the clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Research*, *43*(W1), W566-W570.

Ning, Y., Liu, Z., Yang, Z., Ma, J., Zhang, Y., & Li, Q. (2024). Adulteration detection of plant protein beverages by UPLC-MS/MS based on signature peptide of allergen. *Food Science and Human Wellness*, *13*(6), 3371-3380.

Peixoto, M., Freitas, A. S., Cunha, A., Oliveira, R., & Almeida-Aguiar, C. (2022). Mixing propolis from different apiaries and harvesting years: towards propolis standardization? *Antibiotics*, *11*(9), 1181.

Perez-Castillo, Y., Lima, T. C., Ferreira, A. R., Silva, C. R., Campos, R. S., Neto, J. B. A., Magalhães, H. I. F., Cavalcanti, B. C., Júnior, H. V. N., & de Sousa, D. P. (2020). Bioactivity and molecular docking studies of derivatives from cinnamic and benzoic acids. *BioMed Research International*, *2020*(1), 6345429.

Pérez-Rodríguez, M. L., Serrano-Carretero, A., García-Herrera, P., Cámara-Hurtado, M., Sánchez-Mata, M.C. (2023). Plant-based beverages as milk alternatives? Nutritional and functional approach through food labelling. *Food Research International*, *173*(1), 113244.

Przybyłek, I., & Karpiński, T. M. (2019). Antibacterial properties of propolis. *Molecules*, *24*(11), 2047.

Ramage, G., Saville, S. P., Wickes, B. L., & López-Ribot, J. L. (2002). Inhibition of *Candida albicans* biofilm formation by farnesol, a quorum-sensing molecule. *Applied and Environmental Microbiology*, *68*(11), 5459-5463.

Rowan, D. D. (2011). Volatile metabolites. *Metabolites*, *1*(1), 41-63.

Sforcin, J. M. (2016). Biological properties and therapeutic applications of propolis. *Phytotherapy Research*, *30*(6), 894-905.

Shreaz, S., Bhatia, R., Khan, N., Muralidhar, S., Manzoor, N., & Khan, L. A. (2013). Influences of cinnamic aldehydes on H^+ extrusion activity and ultrastructure of *Candida*. *Journal of Medical Microbiology*, *62*(2), 232-240.

Sova, M. (2012). Antioxidant and antimicrobial activities of cinnamic acid derivatives. *Mini-Reviews in Medicinal Chemistry*, *12*(8), 749-767.

Torres-Rodríguez, J. M., & Alvarado-Ramírez, E. (2007). In vitro susceptibilities to yeasts using the ATB® FUNGUS 2 method, compared with Sensititre Yeast One® and standard CLSI (NCCLS) M27-A2 methods. *Journal of Antimicrobial Chemotherapy*, *60*(3), 658–661.

Walther, B., Guggisberg, D., Badertscher, R., Egger, L., Portmann, R., Dubois, S., Haldimann, M., Kopf-Bolanz, K., Rhyn, P., Zoller, O., Veraguth, R., & Rezzi, S. (2022). Comparison of nutritional composition between plant-based drinks and cow's milk. *Frontiers in Nutrition*, *9*, 988707.

Wang, S., Chen, H., & Sun, B. (2020). Recent progress in food flavor analysis using gas chromatography-ion mobility spectrometry (GC-IMS). *Food Chemistry*, *315*, 126158.

Xie, A., Dong, Y., Liu, Z., Li, Z., Shao, J., Li, M., & Yue, X. (2023). A review of plant-based drinks addressing nutrients, flavor, and processing technologies. *Foods*, *12*(21), 3952.

8. List of publications

Publications related to this PhD thesis:

Papp, Z., Bouchelaghem, S., Szekeres, A., Meszéna, R., Gyöngyi, Z., & Papp, G. (2021). The scent of antifungal propolis. *Sensors*, *21*(7), 2334. DOI: 10.3390/s21072334. PMID: 33801571.

Ranking: Q1
IF: 3,847

Papp, Z., Nemeth L. G,, Nzetchouang Siyapndjeu, S., Bufa, A., Marosvölgyi, T., & Gyöngyi Z. (2024). Classification of plant-based drinks based on volatile compounds. *Foods*, *13*(24), 4086. DOI: 10.3390/foods13244086.

Ranking: Q1
IF: 4,7