Article

Not peer-reviewed version

Metabolomics on Apple (*Malus domestica*) Cuticle—Search for Authenticity Markers

Kamila Bechynska , Jiri Sedlak , Leos Uttl , Vit Kosek , Petra Vackova , Vladimir Kocourek , Jana Hajslova *

Posted Date: 8 April 2024

doi: 10.20944/preprints202404.0476.v1

Keywords: *Malus domestica;* apple cuticle; UHPLC-HRMS/MS; metabolomic fingerprints; classification models; markers; wax esters

 doi:10.20944/preprints202404.0476.v1

Article

Metabolomics on Apple (*Malus domestica*) Cuticle—Search for Authenticity Markers

Kamila Bechynska [1], Jiri Sedlak [2], Leos Uttl [1], Vit Kosek [1], Petra Vackova [1], Vladimir Kocourek [1] and Jana Hajslova [1,*]

[1] University of Chemistry and Technology, Department of Food Analysis and Nutrition, Technicka 3, Prague 6, 16628, Czech Republic; bechynsk@vscht.cz (K.B.); uttll@vscht.cz (L.U.); kocourev@vscht.cz (V.K.); kosekv@vscht.cz (V.K.); vackovap@vscht.cz (P.V.)

[2] Reserach and Breeding Institute of Pomology Holovousy Ltd., Holovousy 129, Holovousy, 50801, Czech Republic; jiri.sedlak@vsuo.cz

* Correspondence: jana.hajslova@vscht.cz; Tel.: +420-602833424

Abstract: The profile of secondary metabolites occurring in apple cuticular layer is not only characteristic for particular apple cultivar, but also dynamically reflects various external factors in growing environment. In this study, the possibility to authenticate apple samples based on analysis of their cuticular layer extracts was investigated. Ultra-high-performance liquid chromatography coupled to high resolution tandem mass spectrometry (UHPLC-HRMS/MS) was employed for obtaining metabolomic fingerprints. In total, 274 authentic apple samples, four cultivars harvested within the seasons 2020 to 2022 in the Czech Republic and Poland were analysed. The generated complex data processed by univariate and multivariate statistical methods enabled to build classification models for discrimination of apple cultivars as well as their geographical origin. The models showed very good performance in differentiation of Czech and Polish samples for three of four cultivars: 'Gala', 'Golden Delicious' and 'Idared', moreover, the validity of the models was verified over harvest seasons. The diagnostic markers, besides metabolites of triterpene biosynthetic pathway, mainly represented wax esters. 'Jonagold', which is known to be susceptible to mutations, was the only cultivar for which the unambiguous classification of geographical origin was not possible.

Keywords: *Malus domestica;* apple cuticle; UHPLC-HRMS/MS; metabolomic fingerprints; classification models; markers; wax esters

1. Introduction

Apple (*Malus* x *domestica* Borkh) belongs to the most widely grown fruits in the temperate climatic zone, their annual world harvest is about 96 million tons [1]. In the Czech Republic, in 2023, more than 100,000 tons of apples were harvested in production orchards (based on the data from Central Institute for Supervising and Testing in Agriculture, 2023). In spite of their higher cost compared to imported apples, regional fruit is preferred by consumers as they believe (thanks to widely used integral farming practices by local growers), that less pesticides were used in their production. Unfortunately, under such conditions, dishonest traders tend to falsely declare correct information about this popular fruit when placing it at the market. Besides of the intentional false declaration of geographical origin, the other fraudulent practice is misclassification of respective apple cultivar. To detect such economically motivated fraud, development of methods for its detection is a challenging task, aim of which is both taking preventive measures against fraudulent practices and consumers' protection [2,3].

To date, a wide range of instrumental techniques for authentication of apples geographic origin and/or cultivar have been published including near-infrared spectroscopy (NIR) [4], fluorescent spectroscopy [5], head-space solid phase microextraction coupled with gas chromatography and mass spectrometry (SPME-GC-MS) [6,7], isotope ratio mass spectrometry (IR-MS), either separately [2,8,9] or in combination with elemental analysis by inductively coupled plasma mass spectrometry (ICP-MS) [10], also electronic nose and electronic tongue were used [3]. More information about these

 doi:10.20944/preprints202404.0476.v1

authentication studies is summarized in **Table 1**. However, as shown here, some of them were conducted with only a limited number of samples or did not cover sufficiently factors of natural variability (different apple cultivars, growing localities, agricultural systems, different harvest years etc.), moreover, in some cases, the description of statistical methods for data handling was not sufficient. Under these conditions, the generic applicability of results might be rather difficult.

Table 1. The overview of studies dealing with authentication of apples origin and/or variety.

Analytical method	Description of apple samples	Classification factor	Number of samples	Number of classes to be distinguished within the sample set	Classification method	Performance of classification	Reference
NIR	Surface of whole apple fruits	Variety Geographical origin	300	3 (varieties Fuji, Red Star, Gala) 2 (grown in different Chinese provinces)	NN, SVM, ELM	calibration set 98 % (ELM) prediction set 97% (ELM)	[4]
fluorescent spectroscopy	Apple juice (squeezed with a juice extractor)	Variety	89	2 (grown in different Chinese provinces)	PLS	calibration set 100% prediction set 96 %	[5]
SPME-GC-MS	Apple juice (squeezed with a juicer)	Variety Geographical origin	50	6 (varieties Starkrimson, Qinguan, Gala, Jonagold, Golden Delicioius, Fuji) 5 (grown in different counties within Chinese province)	LDA, SLDA	predicition set 100 % (SLDA) predicition set 90 % (SLDA)	[6]
SPME-GC-MS	Apple juice (squeezed with hand press)	Variety Geographical origin	4 (3 kg of apples per sample)	4 (varieties Rijo, Verde, Ribeiro, Azedo) 2 (different civil parishes of Madeira)	PLS-DA, HCA	Vague description of model performance	[7]

 doi:10.20944/preprints202404.0476.v1

IR-MS + conventional methods	Pulp, juice	Variety	19	6 (varieties Topaz, Idared, Golden Delicious, Goldrush, Gala, Gloster)	LDA	Insufficient description of models	[2]
		Geographical origin		4 (different regions of Slovenia)			
		Agricultural practice		2 (way of farming organic, conventional)			
IR-MS	Whole apples, peel, pulp, seed	Variety	128	4 (varieties Cripps Pink, Gala, Golden Delicious, Granny Smith)	LDA	71 % correctly classified samples	[8]
		Geographical origin		4 (grown in different districts of northerm Italy)		99% (LOOCV)	
IR-MS	Peel, petiole, pulp, seed	Geographical origin	48	2 (grown in different districts of northern Italy)	LDA	Limited information on on classification models performance	[9]
IR-MS, ICP-MS	Apple juice (concentrated to sugar content 65.0°Brix)	Geographical origin	135	6 (grown in different Chinese provinces)	LDA, PLS-DA	Only description of sample clustering in PLS-DA model without information about model validation	[10]

 doi:10.20944/preprints202404.0476.v1

Electronic nose, electronic tongue	Apple juice (centrifugal juicer)	Variety	126	10 (varieties Fuji, Jonagold, Corolla, Gala, Red Delicous, Red Chief Delicious, Cattle Apple, Ralls Janet, Ourin, Tail, Golden Delicous)	LDA, PLS-DA, SVM	100% (prediction ability of PLS-DA) 100% (accuracy testing rate of SVM)	[3]
		Geographical origin		7 (grown in different Chinese provinces)			

ELM – extreme learning machine, HCA – Hierarchical Cluster Analysis, LDA – Linear Discriminant Analysis, LOOCV – Leave-one-out cross validation, NN – neural networks, PLS – Partial Least Square, PLS-DA – Partial Least Square Discriminant Analysis, SLDA – stepwise Linear Discriminant Analysis, SVM – support vector machine

Various parts and/or processed forms of apples were used within the studies listed in **Table 1**, nevertheless, only in one of them [10], the authentication was based on the data collected from apple surface. However, the NIR technique used in that particular study did not allow identification of characteristic marker compounds that might be applicable for authentication by targeted analysis. In this context, needed to say, that cuticular layer is an interesting matrix worth for authenticity markers search. Cuticular layer contains a number of secondary metabolites and profile of which are characteristic not only for respective cultivar (genotype), but it also can be impacted by various external factors in the particular growing environment, including local weather conditions, application of pesticides and growth regulators, diseases and pests. Cuticle, the outer protective layer of the fruit, consists of structural polymers, which are covered with a wax layer [11]. While intracuticular waxes are incorporated directly into cutin, epicuticular waxes cover a surface of the cutin polymers [12]. Cuticular layer is a complex mixture of secondary metabolites that includes long-chain hydrocarbons and their derivatives such as carboxylic acids, alcohols, aldehydes and ketones, esters, etc., the other group of typical metabolites are various triterpenoids [13–15].

The composition of cuticular waxes was investigated in detail in several studies utilizing various analytical methods. For wax extraction, various authors used cuticle membranes enzymatically isolated by pectinase and cellulase [14,16]; on the other hand, simple procedures based only on rinsing apple surface with a solvent (mostly chloroform, but also dichloromethane, petroleum ether, hexane etc.) were also reported [11,17–19]. The most common method for isolated compounds analysis was GC-MS [11,14,16,18], nevertheless, this method requires derivatization. As an alternative, reversed phase liquid chromatography coupled to mass spectrometry (RPLC-MS) has been increasingly used, both atmospheric pressure chemical ionization (APCI) and electrospray ionization (ESI) were used [20–25]. The advantage of LC-ESI-MS is the possibility of simultaneous detection of different lipid classes, including difficult to ionize neutral wax esters, when additives are used. The use of high-resolution (HR) MS allows identification based on molecular or adduct ion exact masses and the application of collision energy (tandem mass spectrometry, MS/MS) is suitable for obtaining information about molecular structure [20].

The aim of the current study was to investigate the potential of UHPL-HRMS based metabolomic fingerprinting of cuticular layer components followed by advanced statistics for classification of apple cultivars ´Gala´, ´Delicious', 'Idared', 'Jonagold', grown either in the Czech Republic or Poland. To our knowledge, up to now, no other published study employed such approach.

 doi:10.20944/preprints202404.0476.v1

2. Materials and Methods

2.1. Samples

Apple (Malus domestica) samples analyzed in this study were collected by the Research and Breeding Institute of Pomology Holovousy Ltd. (Holovousy, Czech Republic). In total, 274 authentic apple samples of known geographical origin and characterized cultivar were provided over a span of 3 harvest years (2020-2022). The cultivars available in this study were 'Gala', 'Golden Delicious', 'Idared' and 'Jonagold' originating either from the Czech Republic (CZE) or Poland (POL). 35 samples of the 'Gala' cultivar (16 CZE, 19 POL), 37 samples of the 'Golden Delicious' cultivar (19 CZE, 18 POL), 32 samples of the 'Idared' cultivar (17 CZE, 15 POL) and 33 samples of the 'Jonagold' cultivar (13 CZE, 19 POL) were available. After delivery to laboratory, samples were stored in maximum 3 days at 4°C before further processing.

2.2. Chemicals

Analytical grade methanol (MeOH), dichloromethane (DCM), ethyl acetate (EtAC), methyl-tert-butyl-ether (MTBE) and isopropanol (iPrOH) were purchased from Merck (Darmstadt, Germany). Deionized water (dH2O) was obtained from a Milli-Q Integral system (Millipore supplied by Merck (Darmstadt, Germany)). Mobile phase modifiers (ammonium formate, formic acid) were purchased from Sigma-Aldrich (Darmstadt, Germany).

2.3. Methods

2.3.1. Sample preparation

For the extraction of apple cuticular layer, an apple sample was placed into a beaker and submerged into 400 mL of solvent mixture (DCM:MeOH (1:1, v:v) as the optimized mixture used for the whole sample set and mixtures hexane:EtAc (1:1, v:v), DCM, DCM:MeOH (2:1, v:v) and MTBE:MeOH (10:3, v:v) for optimization experiment) and shaken for 10 minutes at 120 min^{-1}. The procedure was two times repeated, three representative apples were always washed. The combined extracts of individual samples were evaporated to dryness and then stored at -80°C. Before analysis, the residue was reconstituted in calculated amount of DCM:MeOH (1:1, v:v) to obtain a standardized concentration of extracted material 33.33 mg/mL. The solution was filtered using syringe filters (pore size 0.22 μm) and 1.5 mL aliquot of respective extracts were then transferred into glass vials for LC-MS analysis. Quality control (QC) sample was prepared as a pool of aliquots of all prepared standardized apple extracts.

2.3.2. UHPLC-HRMS/MS analysis

For UHPLC-HRMS/MS analysis, liquid chromatograph Dionex UltiMate 3000 RS (Thermo Fisher Scientific, Waltham, USA) coupled to quadrupole-time-of-flight TripleTOF™ 6600 mass spectrometer (Sciex, Concord, Canada) was used. For sample components separation, Acquity UPLC BEH C18 (2.1 x 100 mm, 1.7 μm) (Waters, Milford, USA) column was employed. The mobile phase consisted of (A) 5mM ammonium formate in a mixture of dH_2O:MeOH (95:5, v:v) with 0,1% formic acid and (B) 5mM ammonium formate in a mixture of iPrOH:MeOH:dH_2O (95:5, v:v) with 0,1% formic acid; following gradient was applied in case of both positive and negative ionization modes: 0 min (70% A), 2 min (50% A), 7 min (20% A), 13 min (0% A), 20 min (0% A), 20.1 min (70% A), 22 min (70% A) with a constant flow rate of 0,4 mL/min. The column temperature was held at 60°C, the autosampler temperature at 5°C and the sample injection volume was 1 μL.

Mass spectrometer was operated in both positive (ESI+) and negative (ESI-) mode with following ion source settings: nebulizing gas pressure 55 psi, drying gas pressure 55 psi, capillary voltage +4500 V (in ESI+)/-4000 V (in ESI-), ion temperature 500°C. Both MS and MS/MS data were acquired using Full scan and Information Dependent Acquisition (IDA) methods. Mass range in MS mode was set to 100-1200 *m/z* and in MS/MS mode to 50-1200 *m/z*. The collision energy was 35V with the spread of

 doi:10.20944/preprints202404.0476.v1

± 15 V. Mass spectrometer calibration was performed regularly after every 10 samples based on APCI calibration solution (Sciex, Concord, Canada).

Samples were injected in a randomized order, QC samples were injected throughout the entire analytical run (after 10 previous sample injections). Solvent blanks (extraction solvent mixture) were injected at the beginning of the sequence to capture the background features.

2.3.3. Data processing

Processing of acquired UHPLC-HRMS/MS data was performed by open-source software MS-Dial (version 4.8) [26]. In the first step, the data was converted to specific *.ibf format. The peak picking parameters were set as follows for both data acquired in ESI+ and ESI- modes: a minimum signal intensity threshold (peak height) of 10 000, mass accuracy of 0.01 Da for MS data, and 0.025 Da for MS/MS data. For data alignment, retention time tolerance of 0.05 min (ESI+) and 0.3 min (ESI-) along with a m/z tolerance of 0.015 Da were used. In ESI+ mode $[M+H]^+$, $[M+Na]^+$ and $[M+NH4]^+$ adducts and for ESI- mode $[M-H]^-$ and $[M+HCOO]^-$ adducts were considered.

The exported data matrices consisted of all detected features characterized by m/z and retention time were filtered according to the relative standard deviation (RSD) of signal intensity (peak area) in QC samples, with a maximum RSD threshold set to 20% both for ESI+ and ESI-. Furthermore, all features with signal to noise ratio (SNR) below 3 were filtered out to obtain the final data matrices.

2.3.4. Statistical analysis

Before performing statistical analysis, data underwent pre-processing to prevent any potential misinterpretation of data variability. In this study, the total area sum normalization and logarithmic transformation were performed prior any univariate statistics, followed by Pareto scaling in case of multivariate model building.

Within the chemometric processing the aim was to create model for classification of apple samples using both univariate and multivariate statistic tools. Principal Component Analysis (PCA) was used to overview the data. Diagnostics features were selected using t-test/ Analysis of Variance (ANOVA), fold change and Receiver Operating Characteristics (ROC) methods. The combination of used methods, when each of them assesses the feature significance based on a different algorithm, allows selection of relevant markers. Based on the selected feature subset, both Partial Least Square Discriminant Analysis (PLS-DA) and Orthogonal Partial Least Square Discriminant Analysis (OPLS-DA) were applied. Developed classification models were validated using 7-fold internal cross validation and characterized by described variance (R^2X and R^2Y), predicted variance (Q^2Y), the root mean square error of estimation (RMSEE) and permutation tests for R^2Y and Q^2Y.

All statistical analyses were performed using SIMCA® (Sartorius, Gottingen, Germany), Metaboanalyst (metaboanalyst.ca) and using custom built R scripts.

2.3.5. Marker identification

All significant features (significance described by univariate statistical analysis results or Variable Importance on the Projection (VIP) score from PLS-DA/OPLS-DA) used for classification models building, were subject of structure identification. For these features, *.mat files (containing both MS and MS/MS spectral information) were exported from MS-Dial and imported into the open-source software SIRIUS 4 [27,28], which integrates also CSI:FingerID [29] and CANOPUS [30,31]. Together, these three tools suggest possible molecular formulas, potential structures and compound classes for a given feature, which is compared with online spectral databases (BioCyc, HMDB, COCONUT) [32–34]. Tentatively identified markers were characterized by elemental formula, mass error and compound name. In addition, a confidence level of markers identification was classified according to the approaches employed in earlier studies [35–37] for identification of compounds based on LC-MS metabolomic data. Confidence levels range from Level 4 (unknown reproducible signal defined by m/z, retention time and MS spectrum), Level 3 (known compound class with many isomers possibilities), Level 2 (annotated compound based on matched MS/MS spectra and library),

 doi:10.20944/preprints202404.0476.v1

Level 1 (identified compound confirmed with analytical standard) to Level 0 (identified compound including full stereochemistry).

3. Results and discussion

As mentioned in the Introduction, apple cuticular layer contains a wide range of secondary metabolites, some of which might be specific for a particular fruit genotype and, moreover, thanks to sensitivity of surface metabolome to the impact of external factors, its profile may vary depending on conditions in the growing locality. This study, aimed at apple cultivar and its geographic origin authentication, was based on assumption that characteristic metabolites, authenticity markers, would be identified by statistical processing of HPLC-HRMS/MS metabolic fingerprints of cuticular layer extracts. In the paragraphs below, the steps taken to verify this working hypothesis are described.

3.1. Selection of extraction solvent/mixture

In the first step extraction solvent enabling reproducible extraction (not necessarily quantitative) of the widest possible range of substances from the apple cuticle was searched. The tested solvents / solvent mixtures differing in their selectivity involved hexane:EtAc (1:1, *v:v*), MTBE:MeOH (10:3, *v:v*), DCM, DCM:MeOH (2:1, *v:v*) and DCM:MeOH (1:1, *v:v*). The comparison of the tested solvents total ion chromatograms is illustrated in **Supplementary material** (**Figure S1)**. The suitability of extraction solvent was assessed against the distribution of chromatographic peaks in terms of their retention times and the overall number of detected features by reversed phase UHPLC-HRMS method commonly used by authors in metabolomic studies. As wax esters, known to occur in high quantities in apple cuticular layer [13], poorly ionize in ESI- mode, only ESI+ was used for experiments. Similar approach was employed in other studies concerned with analysis of similar matrix [20,22]. Based on the criteria mentioned above, the best solvent mixture was DCM:MeOH (1:1. *v:v*), altogether 11 581features corresponding to compounds representing wide range of polarities was detected. Only slightly lower features number was detected in DCM and DCM:MeOH (2:1, *v:v*) extracts, while in chromatograms obtained by the other solvent mixtures, hexane:EtAc (1:1, *v:v*), MTBE:MeOH (10:3, *v:v*), the more polar metabolites, eluted at lower retention times were not sufficiently represented.

3.2. UHPLC-HRMS/MS analysis

Number of analytical strategies have been applied for apples authenticity, either geographical or varietal. As shown in **Table 1**, in most cases, non target screening performed by various instrumental techniques such as NIR, GC-MS or IR-MS, was employed for analyses of various apple parts /forms. Contrary to those approaches, UHPLC-HRMS/MS technique used this study enables not only obtaining the metabolomic fingerprints of respective sample, but also identification of diagnostic markers (without pre-analytical derivatization) which may be used for authentication based on simpler target screening applicable under routine conditions. The complexity of apple cuticular layer extract is documented by total ion chromatograms (TIC) in **Figure 1A** and **1B.** As expected, wide range of lipophilic compounds was isolated, their main groups are indicated in the figures.

 doi:10.20944/preprints202404.0476.v1

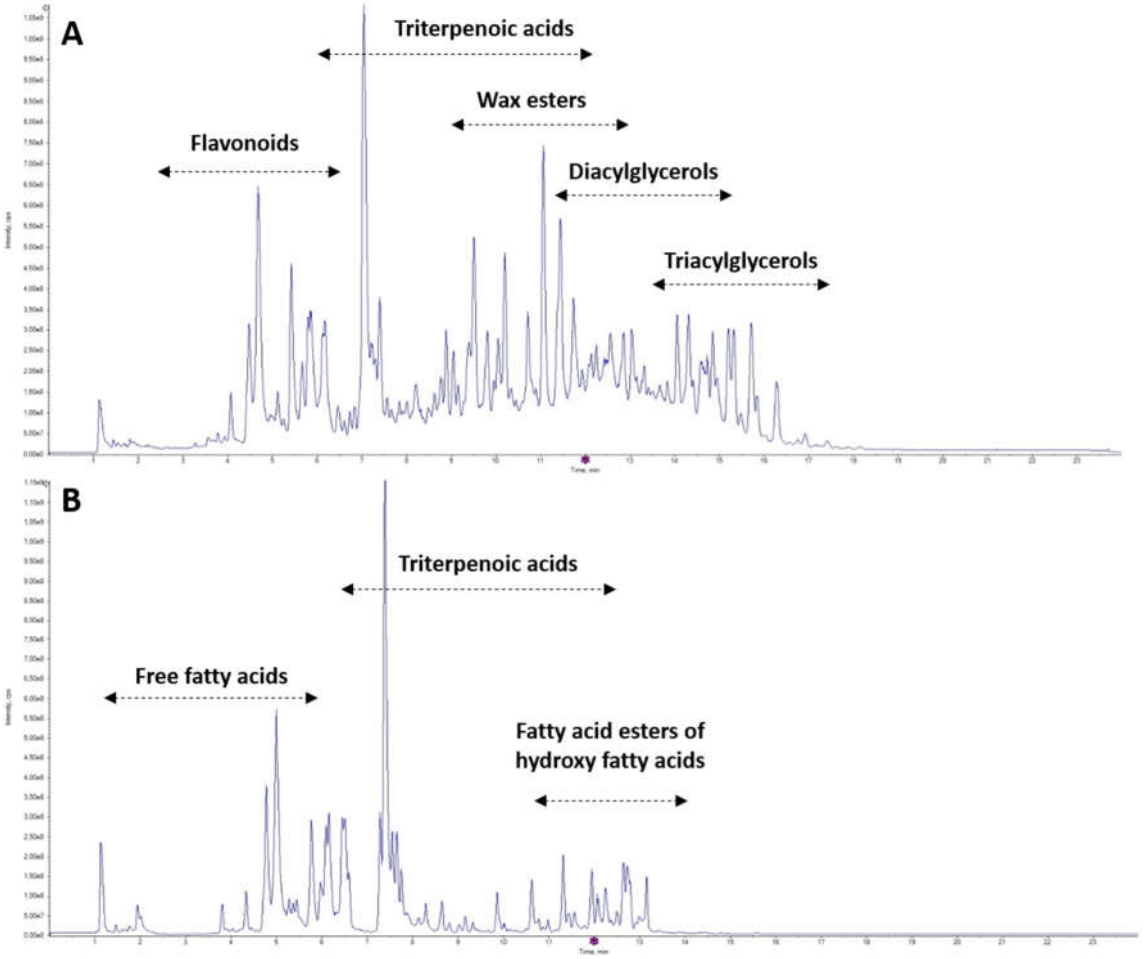

Figure 1. Total ion chromatogram of apple cuticle layer extract, QC sample; (A) ESI+ mode and (B) ESI- mode

3.3. Chemometric analysis

The processing of all raw UHPLC-HRMS/MS data using MS-Dial software resulted in detection of 96 072 features in positive ionization mode and 21 040 features in negative ionization mode. Both obtained aligned data matrices were subsequently filtered according to the criteria mentioned in Materials and methods (2.3.3. Data processing), based on RSD in QC sample (20 %) and SNR in blank sample (≥3). As a result of feature filtering, the final data matrices contained 16 044 features and 2 132 features, in ESI+ and ESI- mode, respectively. These data were used for further processing.

3.3.1. Data overview

Data visualization by PCA revealed clustering of samples based on cultivar (**Figure 2A**), nevertheless, in case of geographical origin the separation of apples from Czech Republic and Poland was not pronounced (**Figure 2B**). The data shown in these figures were obtained for ESI+ mode, similar trends are documented PCA Score plots for ESI- mode, see **Supplementary material** (**Figure S2**).

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 April 2024 doi:10.20944/preprints202404.0476.v1

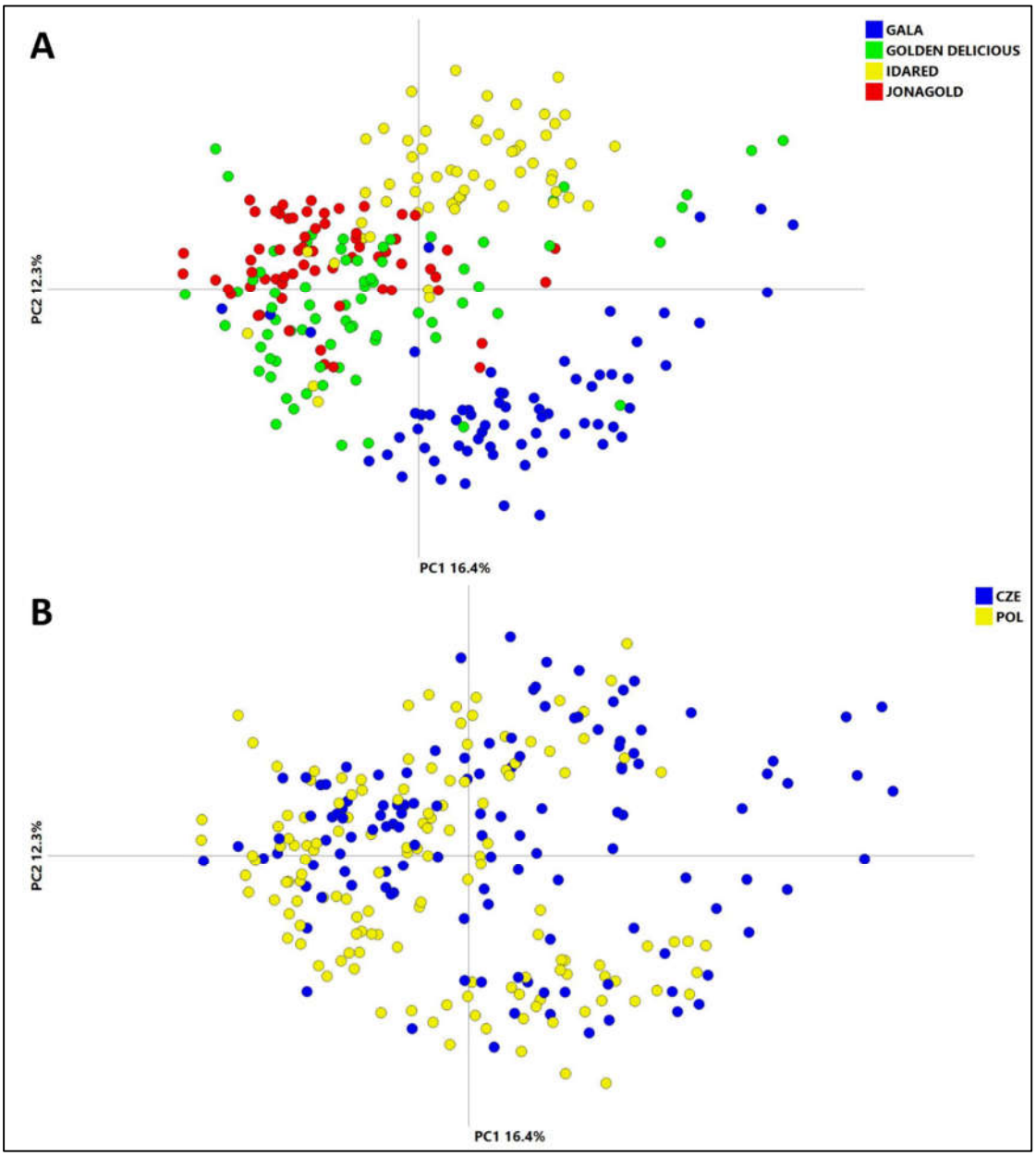

Figure 2. PCA Score plots of the complete dataset of ESI+ features. Samples are coloured according to (A) their cultivar and (B) geographical origin.

To get reliable models for classification of both apple cultivars and geographical origin, two alternative methods of supervised learning were combined. For cultivars classification, PLS-DA was employed, while, because of a less pronounced impact of growing locality on metabolites profile in cuticular layer, a systematic strategy was used for geographicity classification. Individual binary OPLS-DA models were created for each apple cultivar. The use of this particular chemometric method allowed to reduce the effect of collinearity and model overfitting [38].

3.3.2. Apple cultivars classification

Before building the PLS-DA classification models, ANOVA False Discovery Rate (FDR) p-value threshold of 0.05 was used to filter out features unimportant for cultivar differentiation. This process resulted in a subset of 14 551 and 1 678 significant features for ESI+ and ESI- mode, respectively, which were used for PLS-DA models building (all four groups of cultivars included). In the case of model based on ESI+ data, validation parameters were $R^2Y=0.790$ and $Q^2Y=0.756$, for the ESI- data, similar qualitative results were achieved with $R^2Y=0.770$ and $Q^2Y=0.756$. Permutation tests (n=100)

 doi:10.20944/preprints202404.0476.v1

were performed to validate developed models, for both R^2Y and Q^2Y p-value was below 0.01 what indicated valid models [39]. In general terms, VIP score > 1 is considered as a threshold value for significant features [40], nevertheless, in here presented study, this threshold was increased up to 1.5, thus omitting less significant metabolites. In this way, only 197 most significant markers (ESI+ and ESI- combined) were selected for further processing. In **Table 2**, there is the overview of 13 metabolites molecular structure of which was possible to identify, considering criteria specified in Materials and methods (2.3.5. Marker identification).

Table 2. Identified significant metabolites used for apple cultivar classification, (compounds sorted in descending order according to the PLS-DA VIP score).

Marker ion (m/z)	Retention time [min]	Adduct type	Elemental formula	Mass error [ppm]	Tentative identification	PLSDA VIP score	Confidence level
701.7138	14.03	[M+H]+	C48H92O2	-5.4	Wax ester (30:1/18:1)	3.1	2
317.064	2.06	[M+H]+	C16H12O7	-6.7	Isorhamnetine	2.9	3
673.6829	13.69	[M+H]+	C46H88O2	-5	Wax ester (28:1/18:1)	2.9	2
461.1111	2.14	[M-H]-	C22H22O11	5.9	Isorhamnetin rhamnoside	2.8	3
671.6652	13.43	[M+H]+	C46H86O2	-8	Wax esters (46:3)	2.8	3
699.691	14.64	[M+Na]+	C46H92O2	-12.1	Wax esters (46:0)	2.7	3
979.8971	14.81	[M+Na]+	C63H120O5	-6.4	TAG (60:2)	2.4	3
509.4234	12.24	[M+H]+	C31H56O5	5.6	DAG (28:2)	2.1	3
533.0917	1.33	[M-H]-	C24H22O14	2.6	Luteolin-O-malonylglucoside	2.1	3
663.3906	5.99	[M+HCOO]-	C39H54O6	1.3	Caffeoylbetulinic acid	2	3
535.4747	12.87	[M-H]-	C34H64O4	3.8	FAHFA (18:1/16:0)	1.9	2
549.3436	3.7	[M+HCOO]-	C30H48O6	1.6	Triterpenic acid	1.6	3
749.6105	13.21	[M-H]-	C49H82O5	2.8	DAG (46:7)	1.6	3

Flavonoids, wax esters and other lipids together with triterpenoids were the three classes of compounds identified as markers. The first one involved isorhamnetin (methylated metabolite of quercetin), isorhamnetin rhamnoside and luteolin malonylglucoside, secondary metabolites reported

 doi:10.20944/preprints202404.0476.v1

to occur in small quantities in apple skin [41,42]; all of them were present in higher concentrations in 'Idared' and 'Jonagold' apple cultivars. Second group of significant markers (according to the VIP score) was identified as wax esters, a lipid subclass defined by Lipid Maps Structure Database (LMSD) [43]. Although many studies have pointed at the presence of wax esters in apple cuticular layer [12,44–46], the (tentative) identification of individual representatives relied solely on gas chromatography coupled to flame ionization detector or mass-spectrometry with a simple mass analyzer, the sample preparation procedure typically involved hydrolysis of ester bond and derivatization step aimed at an volatility increase of released fatty acids and alcohols Under such conditions, part of information on wax structure is lost. Contrary to that approach, in another earlier comprehensive study [47], GC-MS analysis of whole molecules was performed; the database involving electron ionization mass spectra of 154 wax ester standards (various straight-chain and methyl-branched saturated and unsaturated species) was created. As regards LC-MS, several authors investigated in a greater detail mass spectra of wax esters obtained by this technique. In two older studies [48,49], atmospheric pressure ionization (APCI) and two types of mass analysers (ion trap and orbitrap) were used for this purpose. The dominating signals in mass spectra were protonated molecular ions $[M+H]^+$. Similarly in our study, in which instead of APCI, ESI was used, the most intensive ions in wax ester spectra were $[M+H]^+$, on the other hand, some studies exist, in which ESI was also used (and ammonium formiate was a component of a mobile phase), but the most intensive adducts of wax esters were $[M+NH_4]^+$ [50,51]. In our study employing Q-ToF mass analyser, fragmentation spectra of $[M+H]^+$ ion were further studied and compared with those reported by research team [20] focusing on a systematic investigation of Collision Induced Dissociation (CID) pattern of various wax ester standards by HPLC-ESI-Q-ToF-MS. Based on that, fragmentation spectra of some wax esters could be interpreted, an example of one of identified markers as wax ester (28:1/18:1) is shown in **Figure 3**.

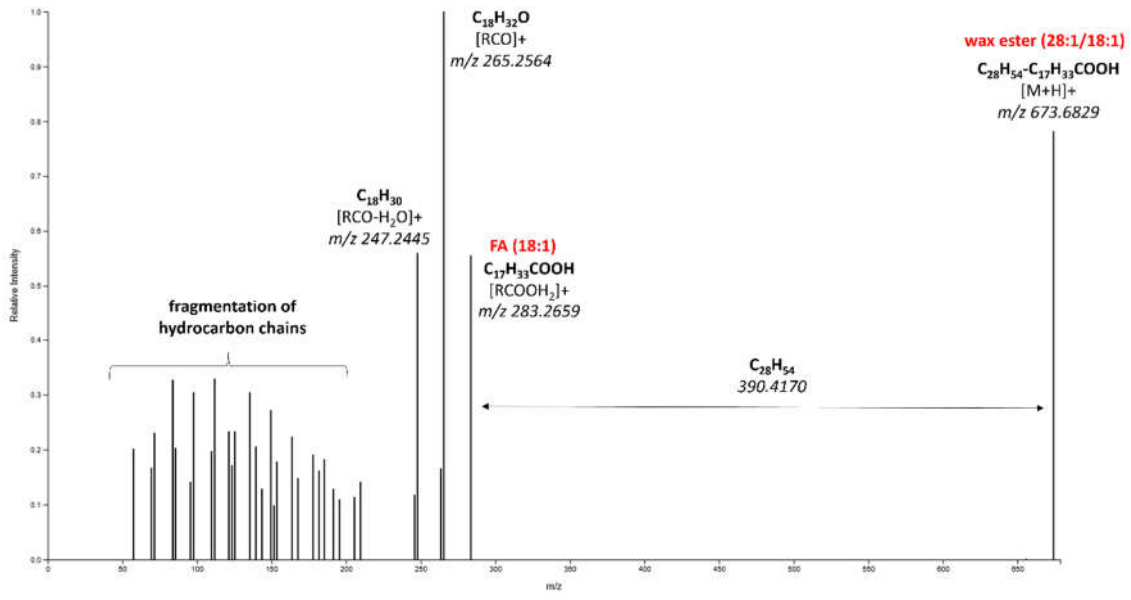

Figure 3. Fragmentation spectrum of marker identified as wax ester (28:1/18:1).

For precursor ion *m/z* 673.6829 a molecular formula $C_{46}H_{88}O_2$ based on its accurate mass and isotope envelope was calculated. In MS/MS spectrum the base ion *m/z* 283.2659 was identified as octadecenoic acid fragment $[RCOOH_2]^+$ originated by breaking ester linkage in wax molecule. Fragments ions *m/z* 256.2564 and *m/z* 247.2445 were residues of oleic acid as $[RCO]^+$ and $[RCO\text{-}H_2O]^+$, respectively, which is in line with spectra interpretation introduced by the mentioned authors [20]. The calculated difference between precursor $[M+H]^+$ ion (*m/z* 673.6829) and fatty acyl fragment (*m/z* 283.2659) 390.4170, corresponded to elemental composition $C_{28}H_{54}$, i.e. the loss of octacosenol alkyl

 doi:10.20944/preprints202404.0476.v1

moiety. Fragmentation spectra for other wax esters with identified fatty acid and fatty alcohol moiety are shown in **Appendix A**. The description of wax esters structure (number of carbon atoms in alcohol: number of double bonds/number of carbons in fatty acid: number of double bounds) is in line with nomenclature used in LMSD for this lipid subclass [52]. Another identified marker belonging to lipids, was fatty acid ester of hydroxy fatty acid (FAHFA, 18:1/16:1). Theses FAHFAs (earlier known as indicators of inflammation in human samples) have been quantified in some food commodities including apple skin relatively recently [53]. The other identified markers represented class of acylglycerols were DAG (46:7), DAG (28:2), TAG (60:2).

The last group of identified markers of apple cultivar were triterpene acids, group of compounds which was the only one studied by several authors in apple cuticle by UHPLC-HRMS/MS technique [22,23,25]. Caffeoylbetulinic acid one of tentatively identified triterpene acids, was, together with caffeoyloleanolic acid, recognized earlier as typical metabolite occurring in skins of russeted apple [25]. It should be noted that the susceptibility to russeting is rather cultivar specific, some apples such as Idared develop this defect very rarely.

Needed to emphasize, that in spite of a relatively low number of markers used for cultivars classification, the model performance characteristics were acceptable with R^2Y=0.691, Q^2Y=0.666 and recognition ability of 91% [39]. The boxplots illustrating markers function are shown in **Supplementary material** (**Figure S3**).

3.3.3. Classification of apple geographical origin

Comparison of PCA Score plots (**Figures 2** and also **S1**) documents a much more pronounced impact of apple variety on the fingerprint of cuticular layer metabolome than that of geographical origin. On this account, the generated data were investigated more in depth, and separately for each variety. The input file was a filtered data matrix with 16044 and 2132 features, obtained in ESI+ and ESI- modes, respectively. Both t-test and ROC were subsequently applied on the data and the fold change was calculated. An OPLS-DA model was created based on the subset of features that met the t-test FDR p-value < 0.05 and the area under curve (AUC) value was higher than 0.75 [54]. Except for model validation performed by using 7-round cross-validation [55], also the validity of the model over several harvest seasons was proved. Predictive ability of the model was calculated by inserting samples from harvest seasons 2020 and 2022 into model created from samples harvested in 2021. The performance characteristics of all created OPLS-DA models for all cultivars are summarized in **Table 3**.

Table 3. Parameters of OPLS-DA models for classification of apple samples according to the geographical origin (Poland vs. Czech Republic).

OPLS-DA model parameters	ESI+				ESI-			
	Gala	Golden Delicious	Idared	Jonagold	Gala	Golden Delicious	Idared	Jonagold
number of features	506	1048	156	11	13	44	24	9
R^2X	0.783	0.596	0.570	0.946	0.667	0.567	0.850	0.921
R^2Y	0.735	0.635	0.886	0.561	0.639	0.738	0.646	0.480
Q^2Y	0.624	0.554	0.809	0.501	0.543	0.686	0.574	0.436
RMSEE	0.265	0.309	0.175	0.335	0.307	0.261	0.308	0.362
p-value of permutation for R^2Y	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

 doi:10.20944/preprints202404.0476.v1

p-value of permutation for Q^2Y	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
validity of the model over time	82%	65%	85%	88%	78%	77%	78%	88%

As shown in **Table 3**, all OPLS-DA models, except for that one classifying geographical origin of 'Jonagold' apples (ESI- data), performed satisfactorily [39]. The highest number of features differentiating between apples from Poland and the Czech Republic was obtained for 'Golden Delicious' cultivar. On the other side, the poorest performance of developed classification models was found for 'Jonagold' cultivar known to be prone to mutations (according to the experts from Czech Research and Breeding Institute of Pomology Holovousy there are 23 known mutations of 'Jonagold' compared to 4 known mutations of 'Golden Delicious', 3 of 'Idared' and 8 of 'Gala'). High susceptibility to mutations was obviously associated with a rather high variability of apple metabolites pattern of this cultivar, even within respective country. On this account, it was difficult to identify reliable diagnostic markers.

All features selected for geographical origin classification (distinguishing between Czech and Polish apple samples), i.e. those with ROC AUC value > 0.75 and t-test FDR p-value < 0.05, were subjected to the identification process. Altogether, 40, 72 and 6 markers were identified for 'Gala' (**Table S1**), 'Golden Delicious' (**Table S2**) and 'Idared' 6 (**Table S3**) cultivars, respectively. With regards to the fact mentioned above, no marker was identified for 'Jonagold' cultivar.

The most frequent markers of geographical origin were wax esters, 23 of them for 'Gala' and 28 for 'Golden Delicious'. Except for fatty acids, also hydroxy fatty acids were bound in wax esters molecules. The example of fragmentation spectrum of such marker (*m*/*z* 631.5975, retention time 13.02 min) identified as hydroxy wax ester (24:0/18:3-O) is shown in **Figure 5**. The obtained spectra were similar to those of wax esters discussed in previous paragraph nevertheless, the calculated molecular formula ($C_{42}H_{78}O_3$ in particular case) contained three oxygen atoms instead of two what is common in plant waxes. In fragmentation spectra of these compounds, alike in that shown in **Figure 5**, there was a visible neutral loss of H_2O molecule from both precursor and fragment ions corresponding to hydroxy fatty acyl [56]. In total, 27 hydroxy wax esters were identified as markers of geographical origin in tested apples. For those, fatty acid composition of which was tentatively determined, corresponding fragmentation spectra are summarized in **Appendix A**.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 April 2024 doi:10.20944/preprints202404.0476.v1

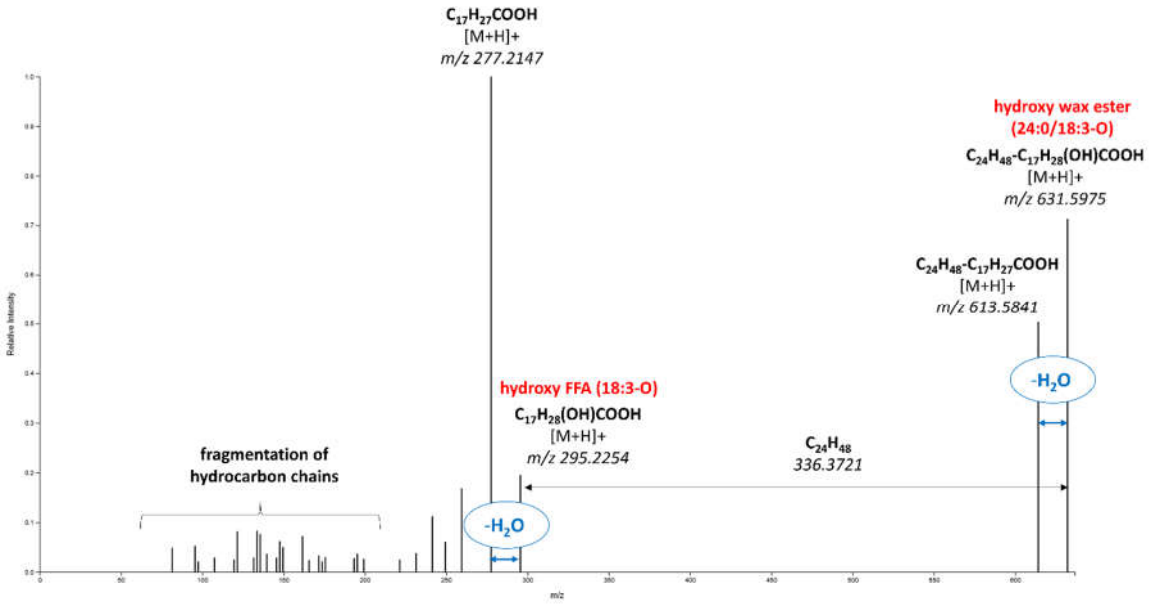

Figure 4. Fragmentation spectra of marker identified as hydroxy wax ester (24:0/18:3-O).

Generally, the most common fatty acids bound in waxes were hexadecanoic and hexadecenoic acid (16:0, 16:1) and unsaturated fatty acids containing 18 carbons (18:1, 18:2, 18:3, 18:4), bonded to C22-C28 aliphatic alcohol. The heatmaps in **Figure 5** illustrate that in both 'Gala' and 'Golden Delicious' cultivars, wax esters as well as hydroxy wax esters, were upregulated in apples harvested in the Czech Republic (exception were shorter chain wax esters, 30:3, 30:4, 30:5). For 'Gala' the differences were more distinct than for 'Golden Delicious', fold change of wax esters ranged from 1.21 to 2.37 and from 1.11 to 3,17, respectively. The higher intensities of wax esters signals (i.e. their higher amounts) observed in samples from the Czech Republic might be due to a higher average altitude (thus other climatic conditions) of orchards in which apples were collected (450 m in the Czech Republic versus 173 m Poland), which was previously reported as a parameter associated with thicker cuticles [57].

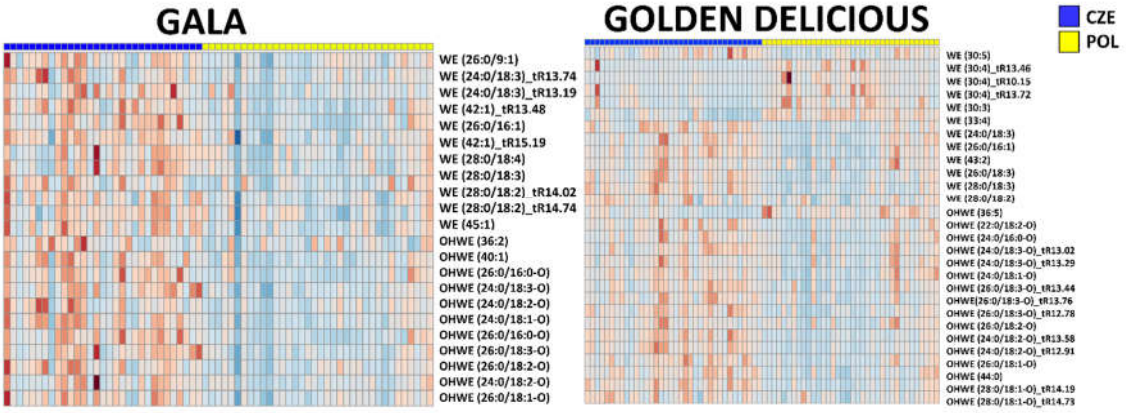

Figure 5. Heatmaps showing fold changes of wax esters (WE) and hydroxy wax esters (OHWE) intensities over median value of respected compound in apple samples (on horizontal axis) harvested either in the Czech Republic or Poland

The next group of identified geographic origin markers were triterpene acids, specifically ursane-type triterpene acids. As mentioned earlier, these compounds were identified in apple cuticle layer together with their derivatives (oxo, dihydroxy, oxohydroxy) by several authors [22,23].

 doi:10.20944/preprints202404.0476.v1

In case of our data, for 'Gala' cultivar, 8 derivatives of ursolic acid were identified among markers, interestingly, all ursenoic acids showed an increased intensity in Czech apple samples. The higher intensity was statistically significant based on t-test p-value < 0.05 (except for oxohydroxy ursenoic acid) with fold changes between 1.3 and 4. For 'Golden Delicious' in addition to 5 ursolic acid derivatives, also their precursors lupeone, uvaol, amyrin and hydroxybetulin were found. Among markers for cultivar 'Golden Delicious' there was no clear trend. As an example, the boxplots of all ursenoic acids selected as markers are shown in **Supplementary material, Figure S4 and S5**.

As relatively polar metabolites such as sorbitol, fucose, heptulose and mannose (trace amounts were contained in analysed extracts) alike some flavonoids were represented by phloretin, chlorogenic acid common apple flavonoids [58,59] were also on the list of geographic origin markers.

4. Conclusion

In this study, UHPLC-HRMS/MS based metabolomic fingerprinting of cuticular layer extracts followed by advanced data processing was shown to be an applicable strategy for authentication of apples cultivars and their geographic origin. The study was based on the assumption that not only respective cultivar (thanks to genetics), but also specific external factors in particular localities, may impact the metabolome of the apple cuticular layer. The outcome of the study, within which set of 274 apple samples involving four cultivars harvested either in the Czech Republic or in Poland within three seasons was investigated, can be summarized as follows:

- Apple cultivar classification of samples based on the developed PLS-DA model was achieved using a subset of significant features as well as 13 identified markers.
- The created OPLS-DA models enabled safe classification of geographical origins 'Gala', 'Golden Delicious' and 'Idared' cultivars.
- However, the geographical origin of the 'Jonagold' cultivar could not be classified due to its susceptibility to mutations thus variability of cuticle layer metabolome even within one geographic area.
- Wax esters, including those with bound hydroxy fatty acids (reported for the first time in apple cuticular wax) represented a significant group of identified markers, amount of which in 'Golden Delicious' and 'Gala' cultivars was higher (upregulated) in samples from Czech Republic compared those from Poland.

Overall, our findings underscore the potential of apple cuticular layer analysis as a robust tool for apple authentication, offering insights into both cultivar and geographical origin distinctions while revealing novel compounds that enhance our understanding of apple wax composition.

Supplementary Materials: The following supporting information can be downloaded at: Preprints.org. **Figure S1**: Total ion chromatograms (ESI+) showing intensities of compounds extracted by various solvent mixtures. **Figure S2**: PCA Score plots of the complete dataset of ESI- features. Samples are coloured according to their cultivar (A) and geographical origin (*B*). **Figure S3**: Boxplots of identified markers for classification of apple cultivar. **Figure S4**: Boxplots of markers for classification of geographical origin for cultivar 'Gala' identified as ursane-type triterpene acids and their derivatives and precursors. **Figure S5**: Boxplots of markers for classification of geographical origin for cultivar 'Golden Delicious' identified as ursane-type triterpene acids and their derivatives and precursors. **Table S1**: Identification of metabolites used for geographical origin classification of apple cultivar 'Gala'. Markers are in descending order according to the AUC ROC value. The log2 FC value indicates whether the marker is increased in Czech samples (log2 FC > 0) or in Polish samples (log2 FC < 0). **Table S2**: Identification of metabolites used for geographical origin classification of apple cultivar 'Golden Delicious'. Markers are in descending order according to the AUC ROC value. The log2 FC value indicates whether the marker is increased in Czech samples (log2 FC > 0) or in Polish samples (log2 FC < 0). **Table S3**: Identification of metabolites used for geographical origin classification of apple cultivar 'Idared'. Markers are in descending order according to the AUC ROC value. The log2 FC value indicates whether the marker is increased in Czech samples (log2 FC > 0) or in Polish samples (log2 FC < 0).

Author Contributions: Conceptualization, K.B., J.S., V.Koc. and J.H.; methodology, K.B., L.U. and P.V.; software, K.B.; formal analysis, K.B.; investigation, K.B., L.U., P.V.; resources, J.S.; data curation, K.B.; writing—original draft preparation, K.B., J.H.; writing—review and editing, J.H., V.Koc., V.Kos.; supervision, J.H., V.Kos., V. Koc., J.H.; project administration, J.S., V.Koc., J.H.. All authors have read and agreed to the published version of the manuscript.".

Funding: This project was funded by the Ministry of Agriculture of the Czech Republic, grant NAZV QK1910104 „Research of metabolomic methods for laboratory authentication of apples geographicity“. Regarding the instrumental research facilities, this study was supported by METROFOOD-CZ research infrastructure project (MEYS Grant no. LM2023064).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors acknowledge Michaela Rektorisova and Klara Navratilova for the support of this research.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Fragmentation spectra of all wax esters with identified fatty acid and fatty alcohol moiety.

References

1. Shanbandeh, M. Apple production worldwide in 2022. Available online: https://www.statista.com/statistics/237605/production-of-apples-worldwide-by-region-2007/ (accessed on 29.02.2024).
2. Bat, K.; Vidrih, R.; Nečemer, M.; Mozetič Vodopivec, B.; Mulič, I.; Kump, P.; Ogrinc, N. Characterization of Slovenian apples with respect to their botanical and geographical origin and agricultural production practice. *Food Technol. Biotechnol.* **2012**, *50*, 107–116.
3. Wu, H.; Yue, T.; Yuan, Y. Authenticity Tracing of Apples According to Variety and Geographical Origin Based on Electronic Nose and Electronic Tongue. *Food Anal. Methods* **2018**, *11*, 522–532. https://doi.org/10.1007/s12161-017-1023-y
4. Li, C.; Li, L.; Wu, Y.; Lu, M.; Yang, Y.; Li, L. Apple Variety Identification Using Near-Infrared Spectroscopy. *J. Spectrosc.* **2018**, e6935197. https://doi.org/10.1155/2018/6935197
5. Bian, H.; Sheng, L.; Yao, H.; Ji, R.; Yu, Y.; Chen, R.; Wei, D.; Han, Y. Application of fluorescence spectroscopy in classifying apple juice according to the variety. *Optik* **2021**, *231*, 166361. https://doi.org/10.1016/j.ijleo.2021.166361
6. Guo, J.; Yue, T.; Yuan, Y. Feature selection and recognition from nonspecific volatile profiles for discrimination of apple juices according to variety and geographical origin. *J. Food Sci.* **2012**, *77*, C1090–C1096. https://doi.org/10.1111/j.1750-3841.2012.02914.x
7. Medina, S.; Perestrelo, R.; Santos, R.; Pereira, R.; Câmara, J.S. Differential volatile organic compounds signatures of apple juices from Madeira Island according to variety and geographical origin. *Microchem. J.* **2019**, *150*, 104094. https://doi.org/10.1016/j.microc.2019.104094
8. Mimmo, T.; Camin, F.; Bontempo, L.; Capici, C.; Tagliavini, M.; Cesco, S.; Scampicchio, M. Traceability of different apple varieties by multivariate analysis of isotope ratio mass spectrometry data. *Rapid Commun. Mass Spectrom.* **2015**, *29*, 1984–1990. https://doi.org/10.1002/rcm.7306
9. Brombin, V.; Mistri, E.; Bianchini, G. Multi stable isotope ratio analysis for the traceability of northern Italian apples. *Food Chem.: X* **2022**, *16*, 100514. https://doi.org/10.1016/j.fochx.2022.100514
10. Liu, X.; Zhao, Y.; Mu, J.; Zhang, J.; Zhang, A. Determination of geographical origin of concentrated apple juice through analysis of stable isotopic and mineral elemental fingerprints: Preliminary results. *J. Sci. Food Agric.* **2021**, *101*, 3795–3803. https://doi.org/10.1002/jsfa.11012
11. Belding, R.D.; Blankenship, S.M.; Young, E.; Leidy, R.B. Composition and Variability of Epicuticular Waxes in Apple Cultivars. *J. Am. Soc. Hort. Sci.* **1998**, *123*, 348–356. https://doi.org/10.21273/JASHS.123.3.348
12. Yang, Y.; Zhou, B.; Zhang, J.; Wang, C.; Liu, C.; Liu, Y.; Zhu, X.; Ren, X. Relationships between cuticular waxes and skin greasiness of apples during storage. *Postharv. Biol. Technol.* **2017**, *131*, 55–67. https://doi.org/10.1016/j.postharvbio.2017.05.006
13. Wu, W.; Jiang, B.; Liu, R.; Han, Y.; Fang, X.; Mu, H.; Farag, M.; Simal-Gandara, J.; Prieto M.A.; Chen, H.; Xiao, J.; Gao, H. Structures and functions of cuticular wax in postharvest fruit and its regulation: A comprehensive review with future perspectives. *Engineering* **2023**, *23*, 118-129. https://doi.org/10.1016/j.eng.2022.12.006

14. Leide, J.; Xavier de Souza, A.; Papp, I.; Riederer, M. Specific characteristics of the apple fruit cuticle: Investigation of early and late season cultivars 'Prima' and 'Florina' (Malus domestica Borkh.). *Scientia Horticulturae* **2018**, *229*, 137–147. https://doi.org/10.1016/j.scienta.2017.10.042
15. Shepherd, T.; Wynne Griffiths, D. The effects of stress on plant cuticular waxes. *New Phytol.* **2006**, *171*, 469–499. https://doi.org/10.1111/j.1469-8137.2006.01826.x
16. Fernandez-Moreno, J.-P.; Malitsky, S.; Lashbrooke, J.; Biswal, A.K.; Racovita, R.C.; Mellerowicz, E.J.; Jetter, R.; Orzaez, D.; Aharoni, A.; Granell, A. An efficient method for medium throughput screening of cuticular wax composition in different plant species. *Metabolomics* **2016**, *12*, 73. https://doi.org/10.1007/s11306-016-0982-0
17. Lara, I.; Belge, B.; Goulao, L.F. A Focus on the Biosynthesis and Composition of Cuticle in Fruits. *J. Agric. Food Chem.* **2015**, *63*, 4005–4019. https://doi.org/10.1021/acs.jafc.5b00013
18. Li, F.; Min, D.; Song, B.; Shao, S.; Zhang, X. Ethylene effects on apple fruit cuticular wax composition and content during cold storage. *Postharv. Biol. Technol.* **2017**, *134*, 98–105. https://doi.org/10.1016/j.postharvbio.2017.08.011
19. Li, F.; Min, D.; Ren, C.; Dong, L.; Shu, P.; Cui, X.; Zhang, X. Ethylene altered fruit cuticular wax, the expression of cuticular wax synthesis-related genes and fruit quality during cold storage of apple (Malus domestica Borkh. C.v. Starkrimson) fruit. *Postharv. Biol. Technol.* **2019**, *149*, 58–65. https://doi.org/10.1016/j.postharvbio.2018.11.016
20. Chen, J.; Green, K.B.; Nichols, K.K. Characterization of Wax Esters by Electrospray Ionization Tandem Mass Spectrometry: Double Bond Effect and Unusual Product Ions. *Lipids* **2015**, *50*, 821–836. https://doi.org/10.1007/s11745-015-4044-6
21. Tada, A.; Jin, Z.-L.; Sugimoto, N.; Sato, K.; Yamazaki, T.; Tanamoto, K. Analysis of the constituents in jojoba wax used as a food additive by LC/MS/MS. Shokuhin Eiseigaku Zasshi. *J. Food Hyg. Soc. Japan* **2005**, *46*, 198–204. https://doi.org/10.3358/shokueishi.46.198
22. McGhie, T.K.; Hudault, S.; Lunken, R.C.M.; Christeller, J.T. Apple peels, from seven cultivars, have lipase-inhibitory activity and contain numerous ursenoic acids as identified by LC-ESI-QTOF-HRMS. *J. Agric. Food Chem.* **2012**, *60*, 482–491. https://doi.org/10.1021/jf203970j
23. Poirier, B.C.; Buchanan, D.A.; Rudell, D.R.; Mattheis, J.P. Differential Partitioning of Triterpenes and Triterpene Esters in Apple Peel. *J. Agric. Food Chem.* **2018**, *66*, 1800–1806. https://doi.org/10.1021/acs.jafc.7b04509
24. Andre, C.M.; Greenwood, J.M.; Walker, E.G.; Rassam, M.; Sullivan, M.; Evers, D.; Perry, N.B.; Laing, W.A. Anti-inflammatory procyanidins and triterpenes in 109 apple varieties. *J. Agric. Food Chem.* **2012**, *60*, 10546–10554. https://doi.org/10.1021/jf302809k
25. Andre, C.M.; Larsen, L.; Burgess, E.J.; Jensen, D.J.; Cooney, J.M.; Evers, D.; Zhang, J.; Perry, N.B.; Laing, W.A. Unusual immuno-modulatory triterpene-caffeates in the skins of russeted varieties of apples and pears. *J. Agric. Food Chem.* **2013**, *61*, 2773–2779. https://doi.org/10.1021/jf305190e
26. Tsugawa, H.; Kind, T.; Nakabayashi, R.; Yukihira, D.; Tanaka, W.; Cajka, T.; Saito, K.; Fiehn, O.; Arita, M. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. *Anal. Chem.* **2016**, *88*, 7946–7958. https://doi.org/10.1021/acs.analchem.6b00770
27. Dührkop, K.; Fleischauer, M.; Ludwig, M.; Aksenov, A.A.; Melnik, A.V.; Meusel, M.; Dorrestein, P.C.; Rousu, J.; Böcker, S. SIRIUS 4: A rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods* **2019**, *16*, 4. https://doi.org/10.1038/s41592-019-0344-8
28. Dührkop, K.; Nothias, L.-F.; Fleischauer, M.; Reher, R.; Ludwig, M.; Hoffmann, M.A.; Petras, D.; Gerwick, W.H.; Rousu, J.; Dorrestein, P.C.; Böcker, S. Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. *Nat. Biotechnol.* **2021**, *39*, 4. https://doi.org/10.1038/s41587-020-0740-8
29. Dührkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Böcker, S. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 12580–12585. https://doi.org/10.1073/pnas.1509788112
30. Kim, H.W.; Wang, M.; Leber, C.A.; Nothias, L.-F.; Reher, R.; Kang, K.B.; van der Hooft, J.J.J.; Dorrestein, P.C.; Gerwick, W.H.; Cottrell, G.W. NPClassifier: A Deep Neural Network-Based Structural Classification Tool for Natural Products. *J. Nat. Prod.* **2021**, *84*, 2795–2807. https://doi.org/10.1021/acs.jnatprod.1c00399
31. Djoumbou Feunang, Y.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.; Steinbeck, C.; Subramanian, S.; Bolton, E.; Greiner, R.; Wishart, D.S. ClassyFire: Automated chemical classification with

a comprehensive, computable taxonomy. *J. Cheminformatics* **2016**, *8*, 61. https://doi.org/10.1186/s13321-016-0174-y

32. Karp, P.D.; Billington, R.; Caspi, R.; Fulcher, C.A.; Latendresse, M.; Kothari, A.; Keseler, I.M.; Krummenacker, M.; Midford, P.E.; Ong, Q.; Ong, W.K.; Paley, S.M.; Subhraveti, P. The BioCyc collection of microbial genomes and metabolic pathways. *Brief. Bioinform.* **2019**, *20*, 1085–1093. https://doi.org/10.1093%2Fbib%2Fbbx085.
33. Sorokina, M.; Merseburger, P.; Rajan, K.; Yirik, M.A.; Steinbeck, Ch. COCONUT online: Collection of open natural products database. *J. Cheminformatics* **2021**, *13*, 2. https://doi.org/10.1186/s13321-020-00478-9
34. Wishart, D. S.; Guo, A. C.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B. L.; Berjanskii, M.; Mah, R.; Yamamoto, M.; Jovel, J.; Torres-Calzada, C.; Hiebert-Giesbrecht, M.; Lui, V. W.; Varshavi, D.; Varshavi, D.; Allen, D.; Arndt, D.; Khetarpal, N.; Sivakumaran, A.; Harford, K.; Sanford, S.; Yee, K.; Cao, X.; Budinski, Z.; Liigand, J.; Zhang, L.; Zheng, J.; Mandal, R.; Karu, N.; Dambrova, M.; Schiöth, H. B.; Greiner, R.; Gautam, V. HMDB 5.0: the Human Metabolome Database for 2022. *Nucleic Acids Res.* **2022**, *50(D1)*, D622-D631. https://doi.org/10.1093%2Fnar%2Fgkab1062
35. Reisdorph, N.A.; Walmsley, S.; Reisdorph, R. A Perspective and Framework for Developing Sample Type Specific Databases for LC/MS-Based Clinical Metabolomics. *Metabolites* **2019**, *10*, 8. https://doi.org/10.3390/metabo10010008
36. Blaženović, I.; Kind, T.; Ji, J.; Fiehn, O. Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics. *Metabolites* **2018**, *8*, 31. https://doi.org/10.3390/metabo8020031
37. Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W-M.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; Hankemeier, T.; Hardy, N.; Harnly, J.; Higashi, R.; Kopka, J.; Lane, A. N.; Lindon, J. C.; Marriott, P.; Nicholls, A. W.; Reily, M. D.; Thaden, J. J.; Viant, M. R. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **2007**, *3*, 211–221. https://doi.org/10.1007/s11306-007-0082-2
38. Bylesjö, M.; Rantalainen, M.; Cloarec, O.; Nicholson, J.K.; Holmes, E.; Trygg, J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J. Chemometrics* **2006**, *20*, 341–351. https://doi.org/10.1002/cem.1006
39. Triba, M. N.; Le Moyec, L.; Amathieu, R.; Goossens, C.; Bouchemal, N.; Nahon, P.; Rutledge, D. N.; Savarin, P. PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters. *Mol. BioSyst.* **2015**, *11*, 13–19. doi:10.1039/c4mb00414k
40. Cocchi, M.; Biancolillo, A.; Marini, F. Chapter Ten - Chemometric Methods for Classification and Feature Selection. In *Comprehensive Analytical Chemistry*; Editor 1,Jaumot, J., Editor 2,Bedia, C., Editor 3, Tauler, R.; Data Analysis for Omic Sciences: Methods and Applications; Elsevier, 2018; Vol. 82, pp. 265–299. https://doi.org/10.1016/bs.coac.2018.08.006
41. Pennington, J.A.T. Food Composition Databases for Bioactive Food Components. *J. Food Comp. Anal.* **2002**, *15*, 419–434. https://doi.org/10.1006/jfca.2002.1073
42. Ramírez-Ambrosi, M.; López-Márquez, D.M.; Abad-García, B.; Dapena, E.; Berrueta, L.Á.; Gallo, B. Comparative study of phenolic profile of fruit and juice samples of a progeny of 'Meana' × 'Florina' from an Asturian cider apple breeding program. Eur. *Food Res. Technol.* **2015**, *241*, 769–784. https://doi.org/10.1007/s00217-015-2502-2
43. Sud, M.; Fahy, E.; Cotter, D.; Brown, A.; Dennis, E. A.; Glass, C. K.; Merrill Jr, A. H.; Murphy, R. C.; Raetz, C. R. H.; Russell, D. W.; Subramaniam, S. LMSD: LIPID MAPS structure database. *Nucleic Acids Res.* **2007**, *35*, D527–D532. https://doi.org/10.1093%2Fnar%2Fgkl838
44. Veraverbeke, E.A.; Lammertyn, J.; Saevels, S.; Nicolaï, B.M. Changes in chemical wax composition of three different apple (Malus domestica Borkh.) cultivars during storage. Postharv. *Biol. Technol.* **2001**, *23*, 197–208. https://doi.org/10.1016/S0925-5214(01)00128-4
45. Klein, B.; Falk, R.B.; Thewes, F.R.; Anese, R. de O.; Santos, I.D. dos; Ribeiro, S.R.; Donadel, J.Z.; Brackmann, A.; Barin, J.S.; Cichoski, A.J.; Wagner, R. Dynamic controlled atmosphere: Effects on the chemical composition of cuticular wax of 'Cripps Pink' apples after long-term storage. *Postharv. Biol. Technol.* **2020**, *164*, 111170. https://doi.org/10.1016/j.postharvbio.2020.111170
46. Morice, I.M.; Shorland, F.B. Composition of the surface waxes of apple fruits and changes during storage. *J. Sci. Food Agric.* **1973**, *24*, 1331–1339. https://doi.org/10.1002/jsfa.2740241104
47. Urbanova, K.; Vrkoslav, V.; Valterova, I.; Haková, M.; Cvacka, J. Structural characterization of wax esters by electron ionization mass spectrometry. *J. Lipid Res.* **2011**, *53*, 204–213. doi:10.1194/jlr.d020834

48. Vrkoslav, V.; Urbanová, K.; Cvacka, J. Analysis of wax ester molecular species by high performance liquid chromatography/atmospheric pressure chemical ionisation mass spectrometry. *J. Chromatogr. A* **2010**, *1217*, 4184–4194. doi:10.1016/j.chroma.2009.12.048
49. Vrkoslav, V.; Urbanova, K.; Haková, M.; Cvacka, J. Analysis of wax esters by silver-ion high-performance liquid chromatography–tandem mass spectrometry. *J. Chromatogr. A* **2013**, *1302*, 105–110. doi:10.1016/j.chroma.2013.06.031
50. Sekosan, G. LC-MS Identification of Wax Esters in Cloudy Canola Oil. *LCGC Suppl.* **2014**, *12*, 16-21.
51. Iven, T.; Herrfurth, C.; Hornung, E.; Heilmann, M.; Hofvander, P.; Stymne, S.; Feussner, I. Wax ester profiling of seed oil by nano-electrospray ionization tandem mass spectrometry. *Plant Methods* **2013**, *9*, 24. doi:10.1186/1746-4811-9-24
52. Liebisch, G.; Fahy, E.; Aoki, J.; Dennis, E. A.; Durand, T.; Ejsing, C. S.; Fedorova, M.; Feussner, I.; Griffiths, W. J.; Köfeler, H.; Merrill Jr, A. H.; Murphy, R. C.; O'Donnell, V. B.; Oskolkova, O.; Subramaniam, S.; Wakelam, M. J. O.; Spener, F. Update on LIPID MAPS classification, nomenclature, and shorthand notation for MS-derived lipid structures. *J. Lipid Res.* **2020**, *61*, 1539–1555. doi:10.1194/jlr.s120001025
53. Liberati-Čizmek, A.-M.; Biluš, M.; Brkić, A.L.; Barić, I.C.; Bakula, M.; Hozić, A.; Cindrić, M. Analysis of Fatty Acid Esters of Hydroxyl Fatty Acid in Selected Plant Food. *Plant Foods Hum. Nutr.* **2019**, *74*, 235–240. https://doi.org/10.1007/s11130-019-00728-8
54. Mandrekar, J.N. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *J. Thorac. Oncol.* **2010**, *5*, 1315–1316. https://doi.org/10.1097/jto.0b013e3181ec173d
55. Westerhuis, J.A.; Hoefsloot, H.C.J.; Smit, S.; Vis, D.J.; Smilde, A.K.; van Velzen, E.J.J.; van Duijnhoven, J.P.M.; van Dorsten, F.A. Assessment of PLSDA cross validation. *Metabolomics* **2008**, *4*, 81–89. https://doi.org/10.1007/s11306-007-0099-6
56. Spickett, C.M.; Pitt, A.R. Oxidative Lipidomics Coming of Age: Advances in Analysis of Oxidized Phospholipids in Physiology and Pathology. *Antioxid. Redox Signal.* **2015**, *22*, 1646–1666. doi:10.1089/ars.2014.6098
57. Tessmer, M.A.; Antoniolli, L.R.; Appezzato-da-Glória, B. Cuticle of "Gala" and "Galaxy" apples cultivars under different environmental conditions. *Braz. Arch. Biol. Technol.* **2012**, *55*, 709–714. https://doi.org/10.1590/S1516-89132012000500010
58. Lommen, A.; Godejohann, M.; Venema, D.P.; Hollman, P.C.; Spraul, M. Application of directly coupled HPLC-NMR-MS to the identification and confirmation of quercetin glycosides and phloretin glycosides in apple peel. *Anal. Chem.* **2000**, *72*, 1793–1797. https://doi.org/10.1021/ac9912303
59. Awad, M.A.; de Jager, A.; van Westing, L.M. Flavonoid and chlorogenic acid levels in apple fruit: Characterisation of variation. *Sci. Hort.* **2000**, *83*, 249–263. https://doi.org/10.1016/S0304-4238(99)00124-7