Microchemical Journal 207 (2024) 111863

Contents lists available at ScienceDirect

Microchemical Journal

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

Comparative analysis of features extraction protocols for LC-HRMS untargeted metabolomics in mountain cheese 'identitation'

S. Pellacani [a], C. Citti [b], L. Strani [a,*], B. Benedetti [c], P.P. Becchi [d], V. Pizzamiglio [e], S. Michelini [e], G. Cannazza [f], A. De Juan [g], M. Cocchi [a], C. Durante [a]

[a] Department of Chemical and Geological Sciences, Università di Modena e Reggio Emilia, Via Campi 103, 41125 Modena, Italy
[b] Institute of Nanotechnology-CNR NANOTEC, CampusEcotekne,Via Monteroni, 73100 Lecce, Italy
[c] Dipartimento di Chimica e Chimica Industriale – DCCI, Università degli Studi di Genova, via Dodecaneso 31, 16146 Genova, Italy
[d] Department for Sustainable Food Process, Università Cattolica del Sacro Cuore, 29122 Piacenza, Italy
[e] Consorzio Formaggio Parmigiano Reggiano, via Kennedy 18, 42124 Reggio Emilia, Italy
[f] Department of Life Sciences, Università di Modena e Reggio Emilia, Via Campi 103, 41125 Modena, Italy
[g] Chemometrics Group. Universitat de Barcelona. Dept. of Chemical Engineering and Analytical Chemistry, Martí I Franquès, 1, 08028 Barcelona, Spain

ARTICLE INFO

Keywords:
MCR-ROI
LC-HRMS
Parmigiano Reggiano
Mountain
Metabolomics
Authenticity

ABSTRACT

This study presents a comprehensive metabolomic analysis of Parmigiano Reggiano samples to differentiate between those designated as Mountain Quality Certification (QC) and conventional Protected Designation of Origin (PDO). Despite following the same production protocol, these cheese varieties differ in the cows' feeding regimes and milk stable locations, with mountain-certified samples adhering to specific requirements regarding milk origin and feed composition. An untargeted approach with Liquid Chromatography-High Resolution Mass Spectrometry (LC-HRMS) was proposed to characterize the cheese metabolome. High-resolution LC-MS data can generate gigabyte-sized files, making data compression essential for manageable multivariate analysis and noise reduction. This study employs the Region of Interest-Multivariate Curve Resolution (ROI-MCR) protocol to achieve effective data compression and chromatographic resolution, thereby extracting the most informative features. This method was compared with a classical approach for feature extraction from chromatographic data, namely Compound Discoverer (CD) software. The features extracted by both methods were analysed through Principal Component Analysis (PCA) and ASCA (ANOVA Simultaneous Component Analysis). The comparison of ROI-MCR and CD approaches demonstrated that while both methods yielded similar overall conclusions, ROI-MCR provided a more streamlined and manageable dataset, facilitating easier interpretation of the metabolic differences. Both approaches indicated that amino acids, fatty acids, and bacterial activity-related compounds played significant roles in distinguishing between the two sample types.

1. Introduction

In 2012 the EU Regulation introduced the "Mountain Product" label to support agricultural activity and food production in disadvantaged areas [1]. The mountain denomination can be a value driver for food produced in mountain areas, making them immediately recognizable to consumers [2,3]. In fact, these products are often attributed specific valence, which typically results in a greater willingness to buy, potentially providing mountain producers with an adequate income and generally contributing to the permanence of agricultural activity and the overall vitality of mountain regions. Therefore, there is a growing interest and an increasing need to evaluate analytical methods capable of objectively identifying the identity characteristics of mountain products to ensure their origin and traceability. However, studies specifically focusing on the mountain denomination are not so numerous [4–9]. Among mountain products, Parmigiano Reggiano, recognized as a Protected Designation of Origin (PDO) product, is one of the most famous due to its distinctive flavor. While often thought of as a single product, the PDO denomination embraces products that differ for ripening time, cow variety, and may have additional denomination, such as organic or mountain label. In particular, the "Prodotto di Montagna – Progetto Territorio" represents a quality denomination for PR cheese that, in addition to the PDO, also respects the mountain denomination [10] and must comply with additional rules established by the producers'

* Corresponding author.
E-mail address: lostrani@unimore.it (L. Strani).

https://doi.org/10.1016/j.microc.2024.111863
Received 21 August 2024; Received in revised form 3 October 2024; Accepted 4 October 2024
Available online 9 October 2024

consortium [11]. In fact, all milk must come from stables in mountain areas and the cheese must be produced and matured in mountain dairies for up to 12 months. Then, at least 60 % of the cows' feed must originate from mountain regions. Eventually, the final product has to undergo a quality inspection at 24 months using the "hammer" test by the Consortium's experts and it must pass a rigorous sensory evaluation.

The main analytical approaches employed to characterize the mountain dairy products have been targeted [4], such as the study of the fatty acids profile, terpenes and volatile compounds in general (by liquid or gas chromatography-mass spectrometry), or the determination of stable isotope ratio of light elements, which varies with altitude. Fingerprinting techniques, such as UV–vis, mid, and near infrared spectroscopies have been mainly used to distinguish dairy products produced by animals undertaking different feeding regimes [4]. More recently untargeted metabolomic approaches have been applied [3,5,9], either based on nuclear magnetic resonance [5] or liquid chromatography-high resolution mavss spectrometry [3,6].

Untargeted metabolomic analysis presents several challenges from the data analysis point of view. First of all, the dimensionality of the data is huge (in the order of gigabytes) and several sources of noise can affect the data, so data reduction/compression is necessary, but this step should not negatively affect information retrieval, e.g. loosing low intensity peaks or lowering spectral resolution [12]. Also, the initial LC-HRMS data, which contain scans of unequally spaced masses, must be reconducted to matrices where the rows represent each scan (i.e., retention times) and the columns represent consistent mass values across all samples. The main approaches generally used to this aim are search for Region of Interest (ROI selection) [13] and binning. Then, features extraction, either by peak detection [14] or peak resolution strategies [15], allow obtaining a data sets holdings peaks areas for all potential detected/resolved metabolites. While aiming at the same results these strategies are quite different since the aim of peak resolution is identifying the pure components which are responsible for the occurrence of the features obtaining a pure spectrum (m/z) and elution profile by resolving coeluted peaks. In feature detection, peak alignment is necessary in order to search for corresponding peaks across distinct chromatographic runs and compare them between samples, and this step may introduce issues [12]. In the present study, we carried out an untargeted metabolomic analysis, based on liquid chromatography-high resolution mass spectrometry (LC-HRMS), of Parmigiano Reggiano cheese (PR), characterised by the Protected Designation of Origin (PDO) [16], referred to as conventional-PDO in the text. In particular, we aimed to characterise the mountain denomination of Parmigiano Reggiano "Prodotto di Montagna – Progetto Territorio" (Mountain-CQ in the text). In this study, we focused on comparing, on our case of study, a ROI strategy with a peak resolution approach, by applying Multivariate Curve Resolution (MCR), i.e. ROI-MCR protocol [17], with a commercial software Compound Discoverer [18] which adopts binning and a peak detection strategy. The different parameters settings and data analysis workflow are illustrated. Once features are retrieved the same preprocessing, data analysis and putative identification of salient features has been conducted for both data sets.

2. Materials and methods

2.1. Chemical reagents

Acetonitrile (ACN) and methanol (MeOH), ultra-pure water and formic acid LC-MS grade were purchased from Thermo Fischer Scientific (Waltham, Massachusetts, USA).

2.2. Sample collection

A total of 40 samples of Parmigiano Reggiano evenly distributed between the two designations examined, i.e. EU mountain label certification, meeting as well the additional criteria of "Prodotto di Montagna – Progetto Territorio" (labeled as "Mountain-CQ"), and conventional Parmigiano Reggiano PDO, (labeled as "conventional-PDO"), respectively, were used for this study.

To obtain the mountain certification, in addition to complying with the PDO protocol [1], the following requirements must be met (i) 100 % of the milk must come from mountain areas; (ii) more than 60 % of the cows' feed must come from mountain areas. In addition to the mountain designation itself, the specific requirements for 'Prodotto di Montagna – Progetto Territorio' [11] are as follows (iii) milk production and maturing for up to 12 months; (iv) qualitative selection at 24 months through an evaluation by the Consortium's experts using the "hammer" method; and (v) successful completion of a sensory evaluation.

The samples were collected directly at the dairies and then stored and prepared (see 2.3.1) at the laboratory of the producers' consortium ("Consorzio del Formaggio Parmigiano Reggiano"). Table S1 (Supplementary Materials) provides detailed information on each sample, including the ripening date, the ripening months, and the dairy province. The geographical distribution of the dairies from which the samples were taken is shown in Fig. S1, Supplementary Materials. Dairies producing conventional PDO are highlighted in green, while those adhering to the certified production "Prodotto di Montagna – Progetto Territorio" are highlighted in red.

2.3. Sample preparation and extraction procedure

The Parmigiano Reggiano samples analyzed consisted of cheese tips, each weighing about one kilogram and stored under vacuum. Initially, these cheese tips were stored in a cold store at temperatures of 4–6 °C. To ensure identical pretreatment of all samples prior to analysis, they were heated to a constant temperature of 20 °C before grinding. The cheese was then cut into approximately two-centimeter pieces to achieve a uniform size before grinding with a Grindomix GM200 (Retsch, Hann, Germany). After grinding, the grated sample was manually mixed and homogenized in a stainless-steel tube. The contents were then transferred to plastic bags and vacuum sealed for sample preparation. Each bag contained approximately 100 g of sample, sufficient for all analyses. These plastic bags were stored in a freezer at −20 °C and transported from the consortium headquarters to our laboratory freezer in a Styrofoam container filled with dry ice to maintain a constant temperature during transportation. Approximately 500 mg of the sample were taken from the plastic containers and transferred to a glass centrifuge tube. Then, using a calibrated pipette, 5 mL of the extractant — a mixture of 2:2:1 acetonitrile:methanol:water with 1 % formic acid — were added to the tube. The mixture was sonicated for 10 min after being shaken for one minute with a mechanical vortex (Falc Instruments, Italy) at a speed of 3000 rpm. Prior to LC-MS analysis, samples were filtered with PTFE filters (0.22 μm) and centrifuged at 5000 rpm for 10 min.

2.4. Ultra-high performance liquid chromatography-high resolution mass spectrometry analysis

Samples were analysed using the Thermo Fisher Scientific Vanquish Core (Thermo-Fisher Scientific, Waltham, MA, USA) as the UHPLC instrument coupled to a heated electrospray ionization system and a mass spectrometer, the Exploris 120 Orbitrap (UHPLC-HESI-Orbitrap). The UHPLC Vanquish Core instrument is equipped with a vacuum degassing system, a binary pump, an autosampler and a thermostable column compartment. The Poroshell 120 SB-C18, (3 × 100 mm, 2.7 μm particle size, Agilent, Milan, Italy) was used for chromatographic separation. The chromatographic separation was performed with water (A) and acetonitrile (B) with 0.1 % (v/v) formic acid. The elution conditions were set as follows: a linear gradient from 5 to 95 % B (0–20 min), maintaining this composition for 3 min, then a rapid change from 95 to 98 % B to wash the column, an isocratic elution with 98 % B (23.1–30 min) and a final re-equilibration step with 5 % B (30.1–36.0 min). The column was

thermostated at 30 °C. The following HESI source characteristics were used: Sheath gas 70 arbitrary units (au), auxiliary gas 5 au, sweep gas 0.5 au, ion transfer tube temperature 390 °C, evaporator temperature 150 °C and electrospray voltage 4.2 kV (positive mode) and 3.8 kV (negative mode). Analyses were acquired by using the Xcalibur software version 4.4 (Thermo-Fisher Scientific, Waltham, MA, USA) in full-scan (FS) and data-dependent (dd-MS2) modes with a fast positive–-negative polarity change and a resolving power of 60,000 full width at half maximum (FWHM) at *m/z* values of 200 for FS mode and 30,000 for dd-MS2 mode. The isolation window for filtering the precursor ions was set to *m/z* values of 1.2, and a progressive step collision energy was used to fragment the precursor ions. Only positive mode ions data was considered for further data elaboration.

Extracts were analyzed on three different days and loaded into the autosampler in a random order w.r.t. to Mountain CQ and conventional PDO samples; six instrumental blanks were also recorded, two for each day. In addition, the extraction of one sample per category was repeated four times and performed in random order, so that a total of 48 specimens were acquired. In this way, the reproducibility of both the extraction and the chromatographic analysis could be monitored during the analysis days.

2.5. Data analysis

2.5.1. ROI-MCR protocol

The data from a high-resolution LC-MS instrument can result in file sizes reaching gigabytes. Compressing the data is thus a fundamental step prior to multivariate data analysis, both to avoid resorting to high performance computing systems and to reduce data noise. Different methodologies have been proposed to this purpose [19]. In this work, Region of Interest (ROI) search combined with Multivariate Curve Resolution – Alternative Least Square (MCR-ALS) [15,20], i.e. ROI-MCR, were exploited to achieve data compression and chromatographic resolution, so to extract the informative features. Initially, 6 blanks and 46 samples were analyzed together using the ROI GUI [21].

Prior to MCR-ROI, the chromatograms were cut along retention times (Rt), retaining only the first 820 datapoints of each chromatogram, since the last Rt region is due to the recondition of the column (Fig. S2, Supplementary Materials).

The Region of Interest approach requires the definition of three input parameters: a threshold for signal-to-noise ratio, an *m/z* tolerance deviation to assign a signal to the same attributed *m/z* value, and the minimum number of consecutive signals at a particular *m/z* value to define a chromatographic peak). The parameter values were established through a comprehensive analysis of literature Refs. [21–24] and conducting several trials on the dataset. Regarding the signal-to-noise-ratio threshold, a range of values, spanning from 0.1 % to 1 % of the maximum ion peak intensity, was systematically examined. Setting the threshold too low led to the retrieval of numerous impurities and noisy peaks, while a threshold set too high resulted in the exclusion of minor characteristic metabolites. In this study, it was determined that a threshold of 0.5 % of the maximum ion peak intensity proved to be the optimal choice. The *m/z* error is usually set to 0.005 Da for Orbitrap analysis so the choice was based on literature works [22,24]. The minimum number of occurrences was tailored to the chromatographic characteristics of the instrument. Specifically, for high-performance liquid chromatography (HPLC), a range of 7 to 12 points is recommended, while in the case of ultra-high-performance liquid chromatography (UHPLC), 2 to 10 points appear to be optimal. A detailed examination of the chromatograms uncovered the presence of narrow and short-lived peaks. Consequently, the minimum number of occurrences was set at 2 per sample to appropriately capture these features. The *m/z* values meeting these criteria are depicted in an augmented matrix featuring elution times for each sample on rows and ROI intensities on columns (MSroi matrix). Every ROI has then been examined carefully to make sure that it depicts a chromatographic peak and is not present in any of the blanks (see Fig. S3, Supplementary Materials). Following this inspection, 114 ROIs were retained, forming an MSroi matrix (data matrix size: 42692 x 114). MCR-ALS has then been applied considering in the column wise augmented data set only the samples without the blanks, yielding a new MSroi matrix (37766 x 114).

2.5.1.1. Multivariate Curve Resolution-Alternating least Suares (MCR-ALS). MCR-ALS [20] resolves elution profiles and spectra of distinct sample constituents by decomposing the initial measurement data according to Eq. (1) (a graphical representation of the decomposition performed by MCR-ALS is shown in Fig. 1).

$$D = CS^T + E \quad (1)$$

where **D** (*I* x *J*) is the final MSroi matrix obtained from the ROI analysis (as described in 2.5.1). **D** is a column-wise augmented matrix, where the MSroi blocks collected from the chromatograms performed on all samples are placed one on top of each other. The range of *m/z* channels adopted in the augmented matrix **D** includes the total amount of relevant *m/z* values found in the chromatograms analysed. In the column-wise augmented matrix **D**, the columns hold the elution profiles at a given *m/z* channel (j = 1,… *J*) for all samples and the rows hold the ROIs signals at each chromatographic retention point (i = 1,… *I*), as shown in Fig. 1. The two factor matrices, **C** and $\mathbf{S}^T$, contain the elution profiles of the N (n = 1,…, N) resolved components (single chemical constituents) for all chromatograms and their related pure mass spectra, respectively. Matrix **E** (*I* x *J*) holds the residuals, i.e. the unmodelled part of **D**. To reduce the rotational ambiguity non-negativity constraints on both **C** and **S** were enforced. Initial pure mass spectra profile estimates (**S**) were derived by a SIMPLISMA-based method [25]. A 51-components model was deemed optimal based on lack of fit assessments [22] (lof = 3.0011, R^2 = 99.9044), and inspection of elution and *m/z* profiles of each component. The 51 resolved components are reported in Table S2 (Supplementary Materials) with their putative identification, *m/z* value, and retention time. The peak areas for each of the 51 components for each sample are calculated by integrating the elution profiles in each **C** column (a single component) obtained by taking the rows corresponding to the related sample. Thus a 48 x 51 feature matrix has been obtained.

2.5.1.2. Putative metabolites identification. Metabolites associated with the ROI-MCR components were putatively identified by using the library present in the Compound Discoverer software [18], i.e. inserting the *m/z* and retention time for each resolved feature in the search. Furthermore, the fragmentation spectrum of the main mass peak (MS/MS data) were matched with online databases such as HMDB, BMDB, FoodDB, MCDB. Finally, manual MS/MS spectra interpretation was aided by mzCloud (https://www.mzcloud.org/). As an example, the putative identification of phenylalanine is reported in Fig. S4, Supplementary Materials. The metabolites reported in this work have an identification confidence equal to level 3 (tentative candidate) [26].

2.5.2. Compound Discoverer procedure

Compound Discoverer (CD) software [18], version 3.2 by Thermo Fisher Scientific, was also employed for comparative purposes. Fig. 2 illustrates the steps undertaken by the CD software in untargeted analysis modality from raw chromatogram to features extraction. Although the specifics of each node in the data analysis pipeline and its algorithms are not always transparent due to confidentiality issues, the process is outlined through a simplified block diagram, with operations executed sequentially. Each node is customizable, allowing parameter adjustments based on the LC-MS chromatogram characteristics.

The analysis begins (first node) by loading the chromatograms of all the analyzed samples in the format produced by the instrument and readable with Xcalibur (Thermo Fisher Scientific). These files contain all the information about how the runs were acquired and the mass spectra recorded (both MS1 and MS/MS).

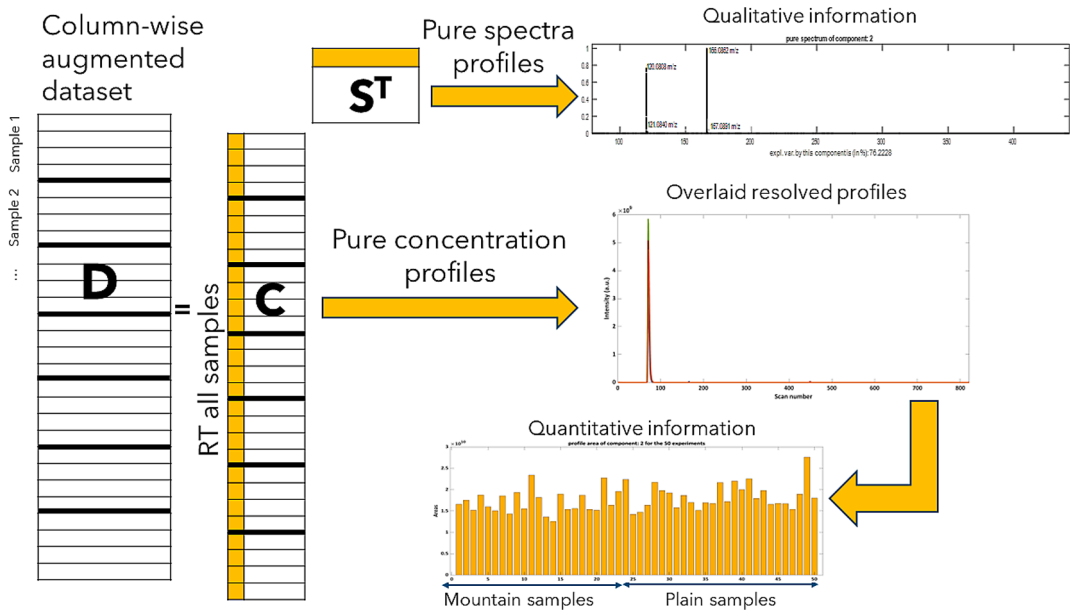

Fig. 1. Graphical representation of the MCR's decomposition (Eq. (1) of the multiset (D) obtained by the ROI analysis. The pure spectra profiles allow the putative identification of the compounds, meanwhile the pure concentration profiles allow determining the relative concentration of the compounds in each sample.

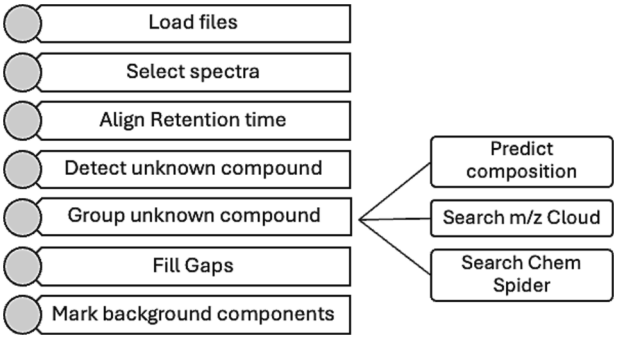

Fig. 2. Schematic block diagram of pipeline implemented in Compound Discoverer.

The next node "Select Spectra" allows the *m/z* spectra to be filtered based on scan number, polarity, degree of fragmentation, peak intensity. Moreover, it is possible to retrieve spectra within a specific mass range. In the case of the threshold, it is important to note that this does not refer to the single ion intensity, as in the case of ROI, but to the sum of the intensities at the centroid of all *m/z* signals in a scan. In this case, the threshold was set equal to 0 to import all scans with intensities above the signal-to-noise ratio (set equal to 1.5).

The subsequent node "Align Retention time" allows the alignment of the chromatograms using the ChromAlig algorithm [27]. Briefly, this is a two steps algorithm: in the first step chromatographic profiles, or more precisely chromatographic surfaces (unfolding the Rt x m/z surface for a sample to a 1D vector), are pre-aligned determining a time offset that maximizes the overlap between two chromatographic profiles (maximizing correlation in Fourier domain), usually the first sample is taken as reference; in the second step, for computational efficiency masses are binned to integer values and times in bin of 0.1 sec, then the algorithm calculate the correlation matrix of the full mass scans between the two pre-aligned chromatographic profiles (i.e. each sample is confronted with the reference one) from which the optimal alignment path is calculated by dynamic warping. However, to reduce the computational complexity only a portion of the correlation matrix is used in this step, since the chromatograms were pre-aligned in the first step. In the present study, the maximum input file shift was set equal to 1 min, and the mass tolerance for feature matching was set equal to 5.0 ppm.

The "Detect Unknown Compounds" node is used to detect the chromatographic peaks within the extracted ion chromatograms (XICs). Here there are several parameters that can be customized as desired by the user. Below are reported the parameters setting used in this study:

- mass tolerance for trace XICs was set equal to 5 ppm;
- minimum base peak height in the XIC traces was set equal to 10^6 FHWM;
- minimum number of consecutive elution times to define a peak was set equal to 2, as in the MCR-ROI analysis;
- in addition, there is a whole set of parameters that filter the chromatographic peaks based on their quality [18]. The vendor recommends keeping the default values for these parameters, which was done in this work.
- finally, this node also allows the user to specify the adducts that may be present in the samples. The following adducts were chosen for this analysis: $[M + H]^{+}$, $[M + NH_4]^{+1}$, $[M + Na]^{+1}$.

The "Group Unknown Compounds" node allows combining the unknown compounds (all the compounds detected so far) in all input files based on their molecular weight and retention time. A mass tolerance of 5 ppm and a retention time tolerance of 0.2 min were used. At this point, for the detected peaks, the integration of areas is accomplished. Areas are determined based on the most common adduct ion in the input files. It may happen that a chromatographic peak revealed by the "Detect Unknown Compounds" node is present in some input files and missing in others. In this case, the "Fill the Gaps" node is used, "the missing" peak can be "re-detected" by lowering the intensity threshold; in the case it is still not detected a "simulated peak" can be used. This is obtained by fitting a Gaussian peak for the expected *m/z* range, however if the filled area is lower than the detection limit, the detection limit value is used to fill the gap (i.e. the gap is filled by noise).

The "Mark Background Compounds" node is used to highlight compounds that are also found in the blanks so that they will not be considered in the final output. In addition, this node chooses the most representative MS1 scan to be used in the next node "Predict Compositions".

The "Predict Compositions" node is used to predict the formulas of

unknown compounds using a mass tolerance of 5.0 ppm. The "Search ChemSpider" node is used to search the mass spectra database for unknown compounds with a certain tolerance (5.0 ppm). In this work, Bovine Metabolome Database (BMDB, https://bovinedb.ca/), FoodDB (https://foodb.ca/), Human Metabolome Database (HMDB, https://hmdb.ca/), and Milk Composition Database (MCDB, https://mcdb.ca/) were selected.

The "mzCloud" node compares MS fragmentation spectra derived from the raw spectra with those of the reference databases within the software (mzCloud https://www.mzcloud.org/)). In the output, the identified compounds are organized by molecular weight, retention time, and measured areas per sample.

2.5.3. Principal component analysis of features matrix

Principal Component Analysis (PCA) [28] was applied both to the feature matrices holding the peak areas of the resolved components elution profile (MCR-ROI analysis) of dimension 48 x 51, and the feature matrix holding the peaks areas obtained by the CD procedure of dimension 48 x 1684.

Prior to PCA both data sets were preprocessed by applying first, normalization along rows (each row holds the peaks areas of the resolved features for a single sample) dividing by the sum of the absolute area values (1-norm) to remove size/bulk effect and then columns autoscaling.

2.5.4. ANOVA Simultaneous component analysis (ASCA)

The ASCA method [29] was applied to the feature matrices (peak areas) obtained by both ROI-MCR and CD, to assess the potential significance of the factors: denomination (mountain/conventional), ripening month (<27 / ≥ 27), and their interaction.

The ASCA algorithm performs an ANOVA, splitting the data variability into the contributions of individual factors and interactions:

$$\boldsymbol{X}_c = \boldsymbol{X} - 1m^T = \boldsymbol{X}_a + \boldsymbol{X}_b + \boldsymbol{X}_{ab} + \boldsymbol{X}_{res} \quad (2)$$

Here, $\mathbf{X}$ denotes the scaled data matrix, m^T is the mean profile of the samples, $\mathbf{X}_a$ and $\mathbf{X}_b$ represent matrices associated with the main effects (in this case, associated with denomination and ripening), $\mathbf{X}_{ab}$ correspond to the interaction effect and $\mathbf{X}_{\text{res}}$ contains the residuals. Each matrix was subsequently analyzed by a distinct PCA model, and Eq. (2) can be re-expressed as:

$$\boldsymbol{X}c = \boldsymbol{T}_a\boldsymbol{P}_b^T + \boldsymbol{T}_b\boldsymbol{P}_b^T + \boldsymbol{T}_{ab}\boldsymbol{P}_{ab}^T + \boldsymbol{X}_{res} \quad (3)$$

where $\mathbf{T}$ contains the scores and $\mathbf{P}$ the loadings of each PCA model, with the maximum number of principal components for each model being equal to the number of levels minus one.

The significance of each design factor or interaction effect was evaluated using permutation tests involving 1000 randomizations [30].

2.6. Software

Xcalibur 3.0 was used as LC-MS spectra acquisition software, while raw data were processed using Compound Discoverer, both from Thermo Fisher (Thermo Fisher Scientific, Waltham, MA, USA). The ROI-MCR and MCR-ALS (2.0) software were downloaded from the official developers' website (https://mcrals.wordpress.com/theory/mcr-als/) and are implemented in the MATLAB environment (Mathworks, Natick, Massachusetts, USA). They have been used in their GUI (Graphical User Interface) version, which allows us to work with a user-friendly interface: command-line versions can also be downloaded for both [20,21]. PLS_Toolbox software (version 9.1, Eigenvector Research Inc., Wenatchee, WA) was used for PCA analysis. ASCA was performed using routines developed by Dr. F. Marini from the University of Roma La Sapienza (Italy), which were kindly made available.

3. Results

3.1. Preprocessing of extracted features

Fifty-one components were extracted from the analysis by ROI-MCR. These were inspected one by one to ensure that they correctly described a feature (a chromatographic peak). A putative assignment was then made based on the m/z and retention time describing the feature by comparison with the Compound Discoverer output. In addition, the MS/MS spectra were manually interpreted using mzCloud. The feature assignments with the corresponding m/z and retention times are listed in Table S2. The peak area matrix (48 x 51) was subjected to preprocessing before applying PCA. The preprocessing step is crucial to obtain valid and reliable results. In this work, as described in 2.5.1.2, two preprocessing steps were applied: Normalization along the rows and autoscaling along the columns. A look at the correlation matrix between the raw extracted features, i.e. peaks area values, (Fig. S5A, Supplementary Materials) shows that there is a strong correlation between most of the variables (correlation = 1 red, anti-correlation = -1 blue. Fig. S5A). This phenomenon is referred to as the "size effect" or dilution effect [31]. The "size effect" may be due to several issues: unwanted systematic errors/bias and experimental variance, or specifically the reduction of systematic variation or bias in the data due to instrument or sampling issues (e.g. sources of experimental variation, inhomogeneity of samples, differences in sample preparation, ion suppression). To eliminate this effect, normalization was applied. It can be observed that normalization significantly reduced the correlation among the variables (Fig. S5B, Supplementary Materials) being effective in removing the dilution effect. Subsequent autoscaling was applied to allow all metabolites to potentially contribute equally to the model.

The same effect has been observed for CD procedure extracted features (Fig. S6, Supplementary Materials).

3.2. Principal component analysis on ROI-MCR features

Principal component analysis (PCA) is used as an exploratory multivariate technique to inspect the data as such and to determine whether Parmigiano Mountain QC labelled samples can potentially be distinguished from conventional PDO ones. A PCA model with 4 principal components explaining 53 % of the variance was inspected. Fig. 3A shows the scores plot of the first component (PC1) versus the second component (PC2). Further components were examined but they were not informative about distinction by category (Fig. S7, Supplementary Materials). The PCA plots were inspected also with respect to ripening months and producers, but no grouping or specific trends were highlighted. The replicates of both classes (blue triangle for the Mountain CQ replicates, light blue triangles towards the bottom for the conventional PDO replicates) are quite tightly close in PC1 and more dispersed along PC2. It is important to note that these samples are replicates of the method, i.e. it is reasonable to have them not strictly overlayed. In addition, colouring the scores plots by preparation days and acquisition days the absence of systematic unwanted variation due to experimental conditions was confirmed.

In Fig. 3A, it is possible to distinguish along PC1 between samples labelled "Prodotto di Montagna – Progetto Territorio" (red diamonds) and conventional PDO samples (green squares). The labelled samples are situated in regions of PCA space associated with the opposite designation. These samples were studied in detail to determine the factors contributing to their behaviour. Among the Mountain-CQ samples that fall within the conventional sample space, sample 26 is from a dairy at 219 m above sea level, while the others are at higher elevations. Looking at the individual characteristics, samples 27 and 29 they have a higher content of feature labelled according to its calculated MW 143.0946, which is proposed to be stachydrine (but its mzCloud best match score is less than 80) than the others. Stachydrine has been reported as characteristic of alfalfa [32], which is one of the most

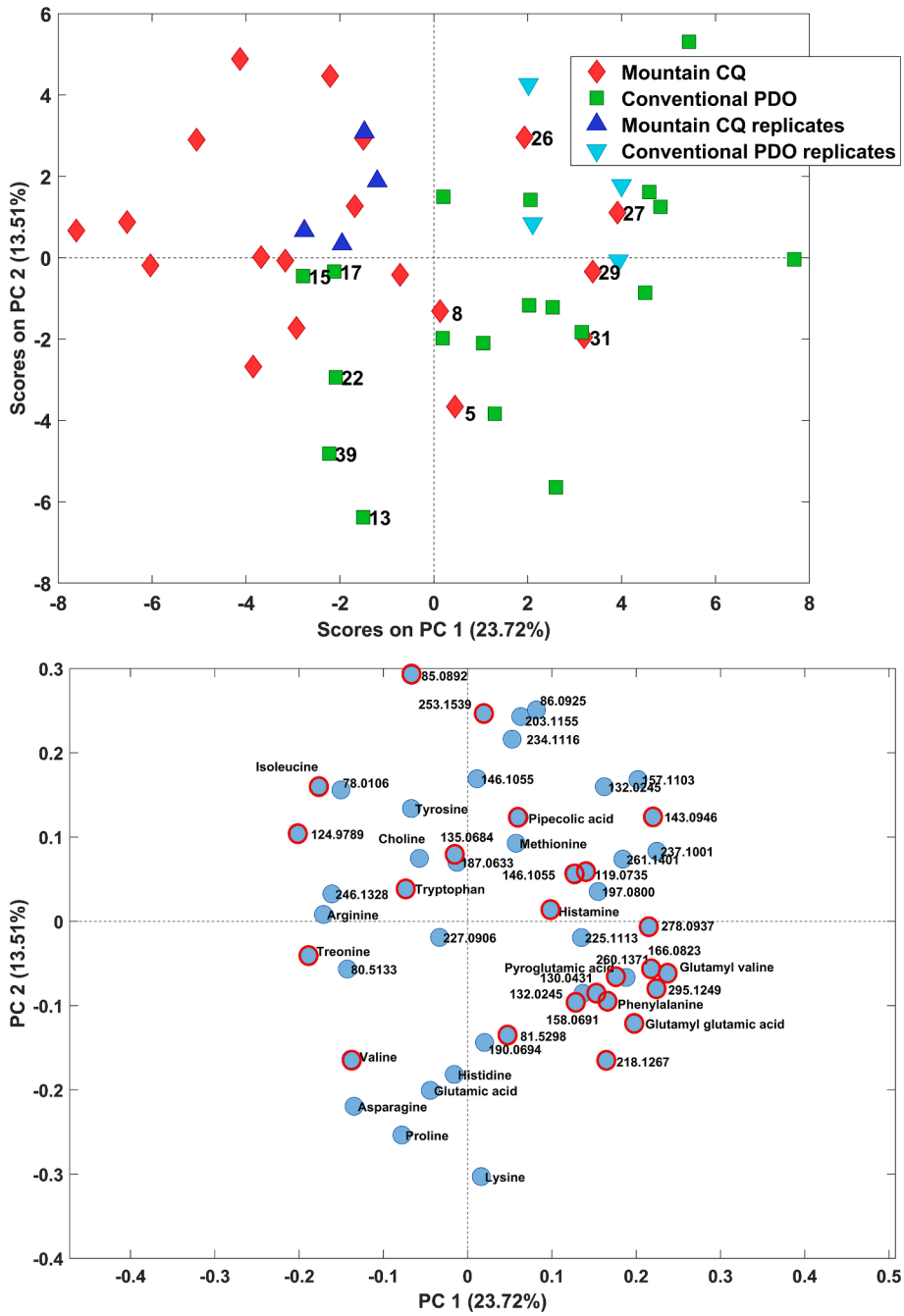

Fig. 3. PCA on the areas resolved by ROI-MCR. A) PC1 vs PC2 scores plot. B) Loadings plot of the PCA on the areas resolved by ROI-MCR. Encircled in red the components that showed a significant loading value for the difference between Mountain-CQ and conventional-PDO according to ASCA analysis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

commonly used forage in Parmigiano Reggiano production, differences in its content may be due to different proportions of forage in the cow diet depending also on time spent on pasture. Samples 31 and 14 have a higher content of glutamil-leucine and a lower content of arginine than the others of their category.

Noteworthy, most of them were also mislocated in a model obtained by PCA of NMR spectra [5]. On the other hand, conventional-PDO samples mislocated are characterized by higher content of valine, asparagine and proline.

The loadings plot (Fig. 3B) shows which analytes are most responsible for the distinction between the two classes. Analytes that could not be identified with Compound Discoverer are indicated with the exact mass (i.e. 86.0964). In general, the metabolic pattern is mainly characterized by amino acids, short-chain carboxylic acids and their derivatives, as also reported in the literature [3,6]. Loadings cluster mostly at positive values, indicating a probably greater richness of the metabolic pattern of the conventional samples. Specifically, the analytes with more negative loadings and thus indicating a higher amount in Mountain QC denomination samples are Leucine-isoleucine, arginine and threonine. Arginine and threonine were reported to characterize mountain CQ samples in a previous NMR metabolomics study [5]. In addition, arginine was reported to be related to the degradation of casein during ripening [33]. Conventional samples are mainly characterized by high contents of glutamic acid derivatives (pyroglutamic acid, glutamyl-valine, glutamyl-glutamic acid) and “stachydrine” which is linked to bacterial activity [34]. Free pyroglutamic acid is related to the ripening

time of Parmigiano Reggiano [33]. However, in our study, the ripening age varies only moderately and is not related to this difference (as shown in Fig. S8, Supplementary Materials, which is coloured according to the ripening months). All glutamyl peptides, such as glutamyl-valine, are products or by-products of the glutathione cycle (GSH) in organisms [35]. Glutamic acid, valine, asparagine and proline at negative PC1 loadings, are mainly indicating higher concentrations of these compounds in the conventional PDO samples numbered 22, 39 and 13, as remarked previously. Valine proved to be a biomarker for non-fermented diary [36]. Asparagine and proline are mainly formed by proteolysis during cheese ripening and are fundamental for cheese flavour [37].

3.2.1. ASCA analysis

In order to support the category distinction observed in PCA, we used the ASCA method on areas obtained by ROI-MCR by considering two factors: the denomination (Mountain-CQ and Conventional) and the ripening time considered at two levels: <27 months and >=27 months, and their interaction. We did not expect a significant difference due to ripening because all the products are aged more than 24 months and the aging range is quite narrow; nonetheless, decoupling this effect could allow a clearer evaluation of the denomination distinction and assessment of the most significant features to explain it.

The results are displayed in Table 1, which includes the explained variance and p-values for each factor and their interaction. In this table are also reported the results related to ASCA applied on the CD extracted features, which will be discussed in section 3.3.1. As expected, only the factor "denomination" resulted statistically significant ($p < 0.001$), whereas factor "ripening time" and the interaction showed a p-value higher than 0.05.

Fig. 4 displays the scores plot of Simultaneous Component Analysis (SCA) for the effect matrix corresponding to the factor "denomination", including projected residuals. The denomination effect has two levels, therefore, the SCA model is represented by a single component (SC1), which explains 100 % of the variance. The scores plot confirmed a significant difference between the two levels of the factor "denomination." Nearly all "Mountain" samples have negative scores, while all conventional samples have positive scores, emphasizing the substantial difference between these two levels.

The variables, i.e. the compounds, mainly responsible to this difference are highlighted in bold in Table S2, and encircled in red in Fig. 3B, confirming valine, leucine, isoleucine, tryptophan, as mainly characterizing Mountain-CQ samples, while phenylalanine and glutamic acid derivatives as characteristic of conventional-PDO. Also using only the subset of significant variables, according to ASCA analysis (highlighted in bold in Table S2) a better separation of the two denominations in scores space was obtained (Fig. S9, Supplementary Materials).

3.3. PCA on Compound Discoverer extracted features

Principal component analysis was also applied to all compounds detected by Compound Discoverer (CD), considering the values of the peak areas measured for all samples (giving a 48 x 1684 features matrix). The same pre-processing procedures as described in Section 3.1 were applied. Metabolites detected by the software were named and

Table 1
Explained variance and probability values for main factors and their second order interaction for both datasets.

Dataset	ROI-MCR		CD	
Parameter	Explained Variance (%)	p-value	Explained Variance (%)	p-value
Denomination	24.01	<0.001	6.72	<0.001
Ripening time	0.94	0.7	0.86	0.95
Interaction	2.49	0.11	2.93	0.1

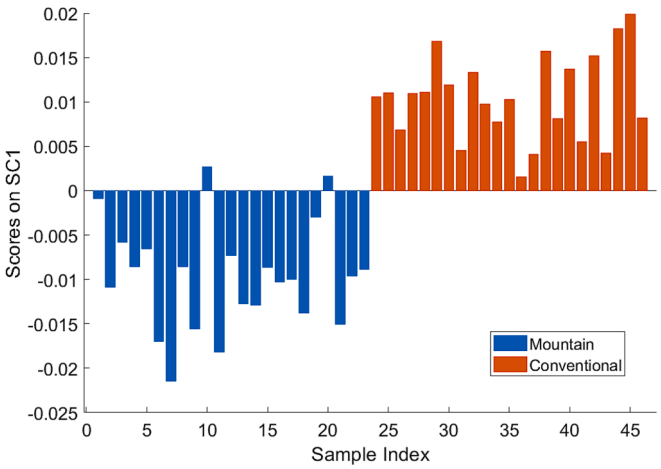

Fig. 4. SCA on the effect matrix "denomination". Scores plot (SC1) with projected residuals related to ROI-MCR dataset.

colored according to chemical classes (see legend of Fig. 5), whereas unidentified compounds were labeled as ND. Only the metabolites putatively identified (reported in Table S3, Supplementary Material) through the reference databases were used for the subsequent simplified graphical representation of the loadings.

In the scores plot (Fig. 5A), the Mountain CQ replicates are distributed in a restricted space, while there seems to be more variability among the conventional PDO replicates. The denomination separation is mainly observable along the first component, with the Mountain CQ samples attaining negative values on PC1, and the conventional PDO samples positive values. The mountain CQ samples are more clustered than the conventional PDO samples, and only two samples are quite far from all others blending with the conventional PDO samples (i.e. number 29 and 31, also mislocated in PCA based on ROI-MCR features). On the other hand, seven lowland samples are mislocated w.r.t. to their category showing negative scores: 17, 22, 15, 40, 28 and 13 (samples 15, 17 and 22 also mislocated in the PCA model based on ROI-MCR features).

The databases HMDB, FoodDB (https://foodb.ca/) and Classy Fire (https://classyfire.wishartlab.com/) were used to categorize the many analytes found into the different chemical classes, listed in the legend of Fig. 5B. Inspection of the loadings plot (Fig. 5B) shows that many of the analytes have PC1's loadings value close to zero, suggesting that most of the metabolomic pattern is shared by the two classes. The region occupied by the Mountain-CQ samples (negative scores) is characterized by a high number of unidentified compounds (red diamonds) and amino acids (yellow circles, loadings at negative PC1 values). Conversely, the positive PC1 direction associated with the conventional PDO samples, is characterized by a greater amount of putatively identified compounds. The main metabolites putatively identified are essential amino acids (phenylalanine, methionine, and leucine), non-essential amino acids (proline, glutamic acid, tyrosine and arginine) and other compounds such as short-chain fatty acids, which are important intermediates in certain metabolic pathways. According to the literature [38], the combination of extraction solvents used can precipitate proteins and simultaneously extract a wide range of metabolites ranging from highly polar compounds, such as polycarboxylic acids and phosphorylated species, to hydrophobic compounds, such as phospholipids and fatty acids/amides. Given the large number of metabolites detected by CD a different visualization was used to simplify the interpretation of the loadings, as follow: first only metabolites with loadings value above a given threshold (more extreme in the PC1 vs PC2 plot) were selected, then a spider plot [39], was used to represent them (Fig. 6). The spider plot itself does not consider the sign of the loading values; therefore, it was split in two sub-plots identifying the most influential compounds,

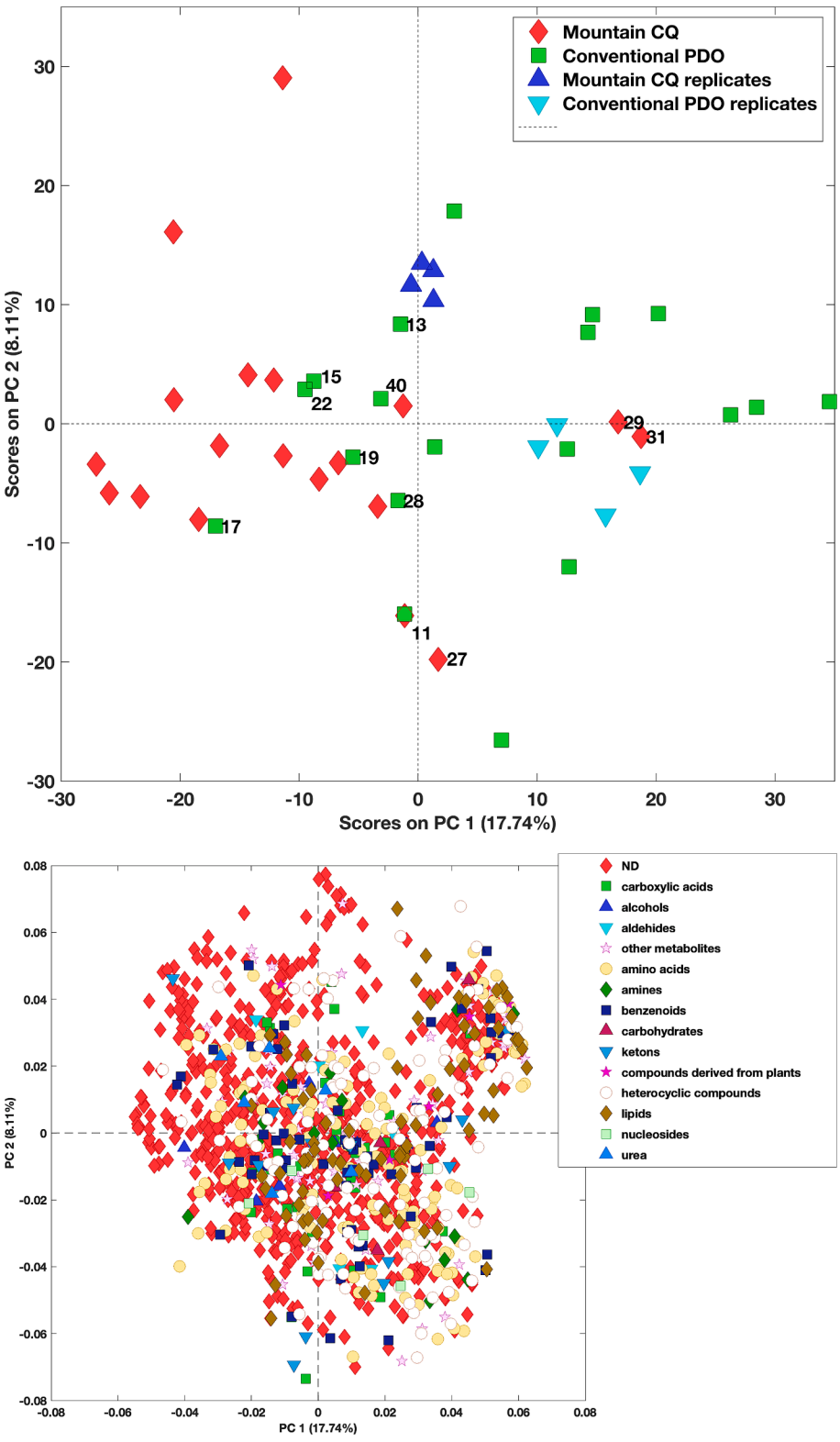

Fig. 5. PCA of the peak areas of the Compound Discoverer's output. A) Scores plot of PC1 vs PC2. B) Loadings plot of PC1 vs PC2. The dotted green lines represent the threshold used for the following spider plot visualization. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

attaining negative (Fig. 6A) and positive (Fig. 6B) PC1 loadings, respectively. The classes of amino acids, lipids and heterocyclic compounds are the most represented in both cases. For Mountain-CQ samples (Fig. 6A) amino acids are most relevant, especially isoleucine, arginine, threonine, serine together with some dipeptides and tripeptides, as already observed in the ROI-MCR analysis (Section 3.2). In terms of lipid composition, medium-chain saturated fatty acids are more present in conventional PDO samples (Fig. 6B, see Table 2 for the correspondence between numbers and analytes), together with some benzenoids, as well as a number of heterocyclic compounds.

3.3.1. ASCA analysis

ASCA was applied to the CD extracted features in the same way as described in section 3.2.1 ASCA analysis for ROI-MCR dataset. Results of the permutation test are shown in Table 1, where it can be observed that, also in this case, "denomination" resulted the only statistically significant factor ($p < 0.001$), with a higher variance explained (6.72 %) than the other main effect and interaction (lower than 3 %), as confirmed by the scores plot of the effect for the factor "denomination", including projected residuals (Fig. 7).

However, the difference between the "denomination" effect and the others was more pronounced in the ROI-MCR dataset, compared to CD dataset. The number of mountain and conventional samples that are confounded is higher than the one obtained by the ROI-MCR dataset. The compounds primarily responsible for this difference are highlighted in bold in Table S3 (Supplementary Material) and confirmed as relevant the same metabolites highlighted by the spider plots. Also when using only the subset of significant variables, according to ASCA analysis (highlighted in bold in Table S3), contrary to ROI-MCR, the separation of the two denominations in scores space did not improve much (Fig. S10, Supplementary Materials).

3.4. Summary of results

Exploratory data analysis showed that is possible to distinguish Mountain-CQ from conventional-PDO based on their metabolic profiles. Most of the mislocated samples in PCA are the same by using MCR-ROI resolved components or CD detected compounds, but the latter tends to mislocate more conventional-PDO ones. ASCA analysis showed that the denomination effect is larger and the samples distinction clearer by using MCR-ROI results.

Table 3 reports only the metabolites that were putatively identified and were detected by both approaches. There is substantial agreement on the positioning of these in the loadings plot except for Pyroglutamic acid. Putative identification (which means that a full match in *m/z* cloud library with a best match score higher than 80 was attained) was possible for twenty of the fifty-one MCR components and ten were significant (according to ASCA analysis) to distinguish Mountain CQ from conventional-PDO categories; however, only four were also found significant for the CD dataset. Phenylalanine, Histamine, Glutamyl-valine, Pyroglutamic, Glutamyl-glutamic and pipecolic acids are characterizing the conventional-PDO while a higher content of the essential amino acids Valine, Isoleucine, Threonine and Tryptophan characterize Mountain-CQ product.

The observed differences can be attributable to the distinct microbial composition of the natural whey starters used in the mountains and plains. These differences may be influenced by several factors, including temperature, milk composition, and especially the milk microbiome. Cows grazing on pasture come into contact with fresh vegetation, which

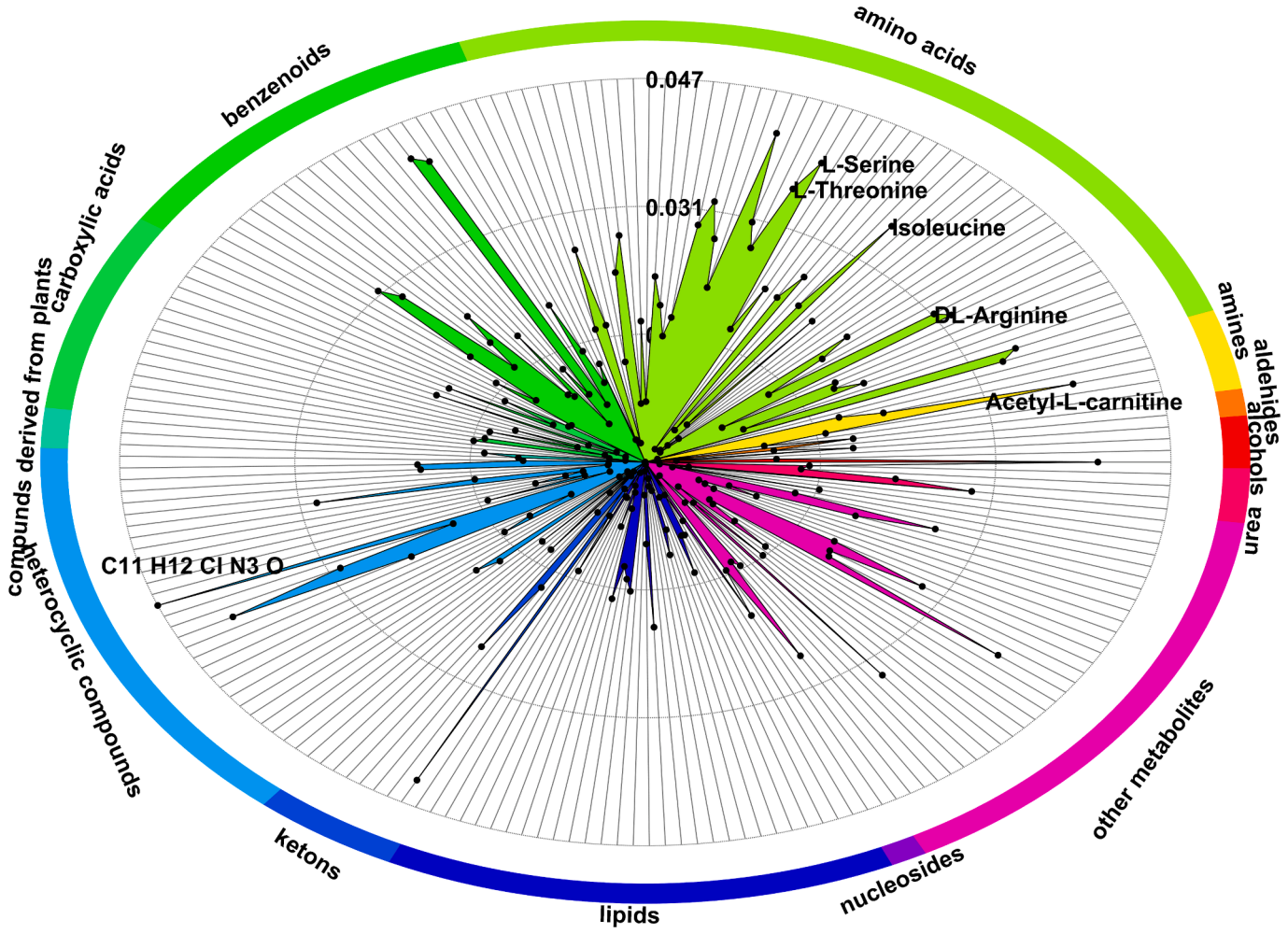

Fig. 6. Spider plot on PC1 loadings. The rays form equal angles to each other; each ray represents one of the variables, and the distance from the center is proportional to the value of the loadings. A) Spider plot obtained from negative PC1 loadings (with values less than −0.03). B) Spider plot obtained from positive PC1 loadings (with values greater than 0.047). The correspondence between the number reported in Figure 6B and the analytes found with Compound Discoverer are listed in Table 2.

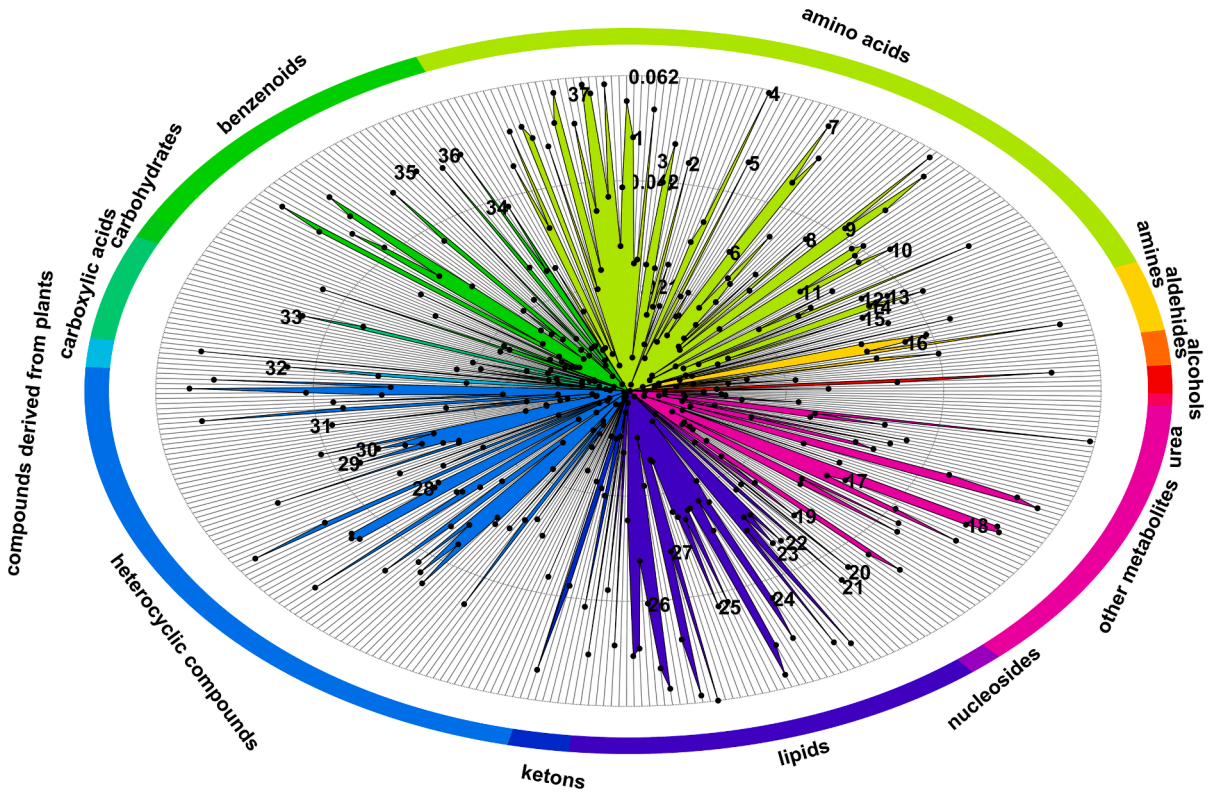

Fig. 6. (*continued*).

Table 2
Correspondence between the number reported in Fig. 6B and the analytes found in with Compound Discoverer.

Number in Fig. 6B	Correspondent analyte	Number in Fig. 6B	Correspondent analyte
1	C9H17NO3S	20	Pinolenic acid
2	C21H22N4O3	21	C17H32O2
3	N-Acetyl-DL-glutamic acid	22	C19H28O5S
4	C9H11NO2	23	C9H13N3O5
5	L-Methionine	24	Ethyl myristate
6	Histamine	25	C10H20O3
7	Hexanoylglicyne	26	C19H34O4
8	C7H13NO2	27	2-(8-Hydroxy-4a,8-dimethyldecahydro-2-napthalenyl) acrylic acid
9	C10H19NO3	28	C12H10N2
10	C22H43NO3	29	C21H25N5O2
11	C14H18N2O5	30	C11H19N3O
12	2-Hydroxyphenylalanine	31	C18H20N2O3
13	2-Aminooctanedioc acid	32	C21H23FN4O2
14	2-Aminoadipic acid	33	C10H15N3O3
15	2-(acetylamino)-4-(methylthio)butanoic acid	34	C19H20N4O2
16	C5H11NO2	35	C13H17NO3
17	C10H14N2O4	36	C19H20N4O2
18	C25H41NO3	37	C16H24O5S
19	Uracil		

likely contributes to a more diverse microbiome [40,41].

3.5. Comparative discussion of the two approaches

Applying both Compound Discoverer and ROI-MCR to the same dataset allowed for a thorough examination of their strengths and weaknesses, laying the groundwork for an initial comparison. From the

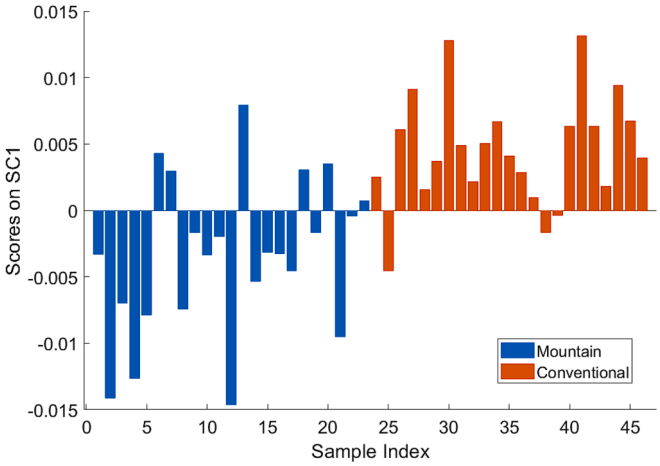

Fig. 7. SCA on the effect matrix "denomination". Scores plot (SC1) with projected residuals related to CD dataset.

methodological point of view, as reported in [11], the most relevant differences are: (i) chromatographic alignment is not needed in the MCR-ROI protocol as well as no binning is applied either in the *m/z* or retention time; (ii) some initial settings are common to both approaches, i.e. signal-to-noise ratio, maximum peak intensity and *m/z* error, but overall, less parameters have to be set with respect to the CD pipeline where there are additional settings, e.g. in the "Alignment" and "Detect Unknown Compounds" nodes; (iii) the ROI selection step in ROI-MCR allows higher compression of the full scan data, which indeed is highly sparse, while in the CD pipeline there is not an analogous compression step. The number of peaks detected (hence of final features retained) is governed by the several parameters present in the "Detect Unknown Compounds" node and fill the gap option (i.e. Jaggedness quality factor, Modality quality factor, Zig-Zag quality factor, etc. [18]. Generally, a

Table 3
Metabolites putatively identified, common to MCR-ROI and CD. A positive PCA loadings sign indicates metabolites that are more abundant/characteristic of conventional-PDO w.r.t. to Mountain-CQ, and *vice versa* for negative loadings sign (the higher the number of + or – the more extreme in the loadings plot the variables are). The last column reports if they were significant for differentiation according to the ASCA permutation test.

MCR component n°	Putative identification	PCA Loadings sign (MCR)	PCA Loadings sign (CD)	Significant ASCA
1	Proline	−	−-	
2	Phenylalanine	+	+	✓ (MCR-ROI)
4	Valine	−-	−-	✓ (MCR-ROI)
5	Methionine	+	++	✓ (CD)
6	Lysine	0	−0	
7	Isoleucine	−−	−−	✓ (MCR-ROI, CD)
8	Tyrosine	−	−	
10	Pyroglutamic acid	++	−	✓ (MCR-ROI)
12	Glutamic acid	−0	−-	
14	Arginine	−-	−-	
15	Glutamyl valine	++	++	✓ (MCR-ROI)
17	Histamine	+	++	✓ (MCR-ROI, CD)
19	Choline	−	−-	
20	Glutamyl glutamic acid	++	+	✓ (MCR-ROI)
22	Pipecolic acid	+	0	✓ (MCR-ROI)
30	Histidine	−0	−-	
32	Treonine	−−	−−	✓ (MCR-ROI, CD)
41	Asparagine	−-	−-	
46	Tryptophan	−-	0	✓ (MCR-ROI, CD)

quite higher number of features is retained; (iv) MCR is a spectral unmixing method, so the peaks area integration is operated on "pure" elution profile bypassing the overlapped peaks issue. The fact of working with resolved chromatographic peaks also results on profiles less affected by noise. In the CD pipeline (in the "Detect Unknown Compounds" node) to overcome this issue peaks area integration is based on a single ion (most common adduct). Nonetheless, the preceding peak detection step may be affected by peaks overlapping.

Summarizing, while both methods demand a certain level of user expertise, Compound Discoverer is notably more intricate, necessitating a deeper understanding of numerous nodes and parameters. In contrast, ROI-MCR requires defining only three parameters per Region of Interest (ROI) and specifying the number of components per Multivariate Curve Resolution (MCR). Another significant distinction lies in the transparency of the methodologies. ROI-MCR is based on an open-source routine, accessible to all, whereas Compound Discoverer is proprietary software, limiting access to detailed information for legal confidentiality reasons. Consequently, adjusting parameters in ROI-MCR is more straightforward, offering clarity regarding their functions. In Compound Discoverer, user intervention is often restricted, with default parameter values frequently retained. Analysing outputs is also more challenging due to the abundance of features obtained. Conversely, ROI-MCR facilitates individual feature analysis, aiding in decision-making regarding their retention based on peak quality.

A notable advantage of Compound Discoverer is its ability to interface with databases automatically, enabling preliminary and putative identifications. It excels in comprehensive, untargeted characterizations, accommodating even low-intensity analytes. On the other hand, the ROI-MCR approach focuses on feature extraction and resolution, which was requiring a manual comparison of mass spectra with databases for feature identification, however very recently an automatic match with some databases has been developed [42]. Conversely, if the objective is to extract and differentiate features efficiently, ROI-MCR proves to be swifter and more effective, yielding datasets with a manageable number of compounds suitable for sample differentiation based on various indicators. While Compound Discoverer outputs extensive compound lists, ROI-MCR generates datasets with a more manageable size, streamlining analysis and interpretation processes.

In the present study, we pose attention in setting the same initial parameters, for the one that are common/have the same meaning and in general to make comparable choices. Also, we use the same database for putative identification and interpreted only metabolites whit high match score in the mas spectra library. Indeed, both methods yielded highly similar results, underscoring the robustness of their outcomes.

As already remarked the CD extracted features were much higher (1684), however these include a number of features that refer to the same metabolite, after manual pruning 1257 were selected, and as reported in Table S3, only for 169 there was a full match with the mzCloud library and 97 were putatively identified.

At variance, with as little as 51 features, conventional samples and mountain-CQ samples could be distinguished using the ROI-MCR approach. This greatly accelerates and simplifies the process of determining potential biomarkers. To this aim ASCA is a more suited method than analysis based on p-values from multiple statistical comparison test; 26 and 41 compounds resulted significant for the MCR-ROI and CD dataset, respectively (the putatively identified are in common and discussed in 3.4).

4. Conclusion

In this work, the metabolomic pattern of Parmigiano Reggiano samples was analysed in an untargeted manner by LC-HRMS, the results highlighted possible distinctive compositional traits for Mountain-CQ PR product when confronted with conventional PDO PR one. This is a promising result for the valorisation of the mountain label, also considering that these are very close product that share the same production protocol except for being produced in the mountain area (100 % of the milk and at least 60 % of the cow feed).

The untargeted LC-HRMS analysis allows a thorough characterisation of the cheese metabolome, nonetheless its main drawback is a complex and high-dimensional output, which represents a computational challenge, and, depending on the methodologies used, may lead to a very large number of detected analytes whose significance cannot be assessed solely on the basis of univariate significance tests, albeit mightily corrected for multiple comparisons, to this aim ASCA proved suitable. In this paper, two approaches were compared: ROI-MCR protocol and Compound Discoverer pipeline. The latter is more prone to furnish a huge list of metabolites, not implementing any unmixing/resolution.

The convergence of the results that we obtained by the two methodologies (which strengthen them) is not always granted in metabolomic analysis. It is worth noticing, that particular care was posed in using similar thresholds and considering only features that gave a reasonable matching score with the mzCloud library.

Moreover, the results of LC-MS analysis are in accordance with the previous findings obtained by NMR analysis on the same dataset [5], and some are also in common with finding reported in few recent studies which reported metabolomic analysis of PR [3,6]. However, it is worth noticing that samples considered there spanned a wide ripening period (a range of three years) and/or contained also a percentage of rind. Here a balanced number of Mountain-CQ and Conventional-PDO PR samples were considered, spanning in a representative way the whole production area and showing a very contained ripening age difference. Obviously, analysis of standards will be required to increase confidence in the identification of the analytes. In addition, given the presence of analytes characterizing the mountain samples that have not yet been identified, it

will certainly be possible in the future to perform more targeted experiments on the analytes of interest, for example, with the use of an inclusion list that allows MS experiments to focus on specific *m/z*. Finally, works is in progress to expand the dataset so to reach a sufficient number of samples to perform supervised classification.

CRediT authorship contribution statement

S. Pellacani: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **C. Citti:** Writing – review & editing, Resources, Methodology, Investigation, Data curation. **L. Strani:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **B. Benedetti:** Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation. **P.P. Becchi:** Methodology, Investigation, Data curation. **V. Pizzamiglio:** Writing – review & editing, Supervision, Resources, Data curation. **S. Michelini:** Writing – review & editing, Investigation, Data curation. **G. Cannazza:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Conceptualization. **A. De Juan:** Writing – review & editing, Validation, Software, Investigation, Conceptualization. **M. Cocchi:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **C. Durante:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research has been supported by University of Modena and Reggio Emilia - Fondazione Modena through FAR Mission Oriented 2021 funds (project: MOUNTAIN-ID). A.J. acknowledges funding from the Spanish government through the project referenced PID2023-146465NB-I00.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.microc.2024.111863.

References

[1] P.O. of the E. Union, Publication of an application for approval of a minor amendment in accordance with the second subparagraph of Article 53(2) of Regulation (EU) No 1151/2012 of the European Parliament and of the Council on quality schemes for agricultural products and foodstuffs, Publ. Off. EU (2018). https://op.europa.eu/en/publication-detail/-/publication/7886031c-3ea3-11e8-b5fe-01aa75ed71a1 (accessed October 1, 2024).
[2] E.L.M. Artinelli, F.R.D.E.C. Anio, R.E.E. Milia, Exploring the impact of the "Mountain Product" label guarantee on the attitude-intention path, in: 2023. https://www.semanticscholar.org/paper/Exploring-the-impact-of-the-%E2%80%9CMountain-Product%E2%80%9D-on-Artinelli-Anio/719b1c65271f489ae9f17effaef79d94c5dcc291 (accessed August 8, 2024).
[3] P. Paolo Becchi, G. Rocchetti, F. Vezzulli, M. Lambri, L. Lucini, The integrated metabolomics and sensory analyses unravel the peculiarities of mountain grassland-based cheese production: The case of Parmigiano Reggiano PDO, Food Chem. 428 (2023) 136803, https://doi.org/10.1016/j.foodchem.2023.136803.
[4] M. Ungureanu-Iuga, I. Surdu, D. Necula, Characteristics of mountain vs. lowland dairy products, Int. J. Food Sci. Technol. 59 (2024) 4359–4373, https://doi.org/10.1111/ijfs.17150.
[5] N. Cavallini, L. Strani, P.P. Becchi, V. Pizzamiglio, S. Michelini, F. Savorani, M. Cocchi, C. Durante, Tracing the identity of Parmigiano Reggiano "Prodotto di Montagna - Progetto Territorio" cheese using NMR spectroscopy and multivariate data analysis, Anal. Chim. Acta 1278 (2023) 341761, https://doi.org/10.1016/j.aca.2023.341761.
[6] P.P. Becchi, G. Rocchetti, P. García-Pérez, S. Michelini, V. Pizzamiglio, L. Lucini, Untargeted metabolomics and machine learning unveil quality and authenticity interactions in grated Parmigiano Reggiano PDO cheese, Food Chem. 447 (2024) 138938, https://doi.org/10.1016/j.foodchem.2024.138938.
[7] V. Maciuc, C. Pânzaru, M. Ciocan-Alupii, C.-G. Radu-Rusu, R.-M. Radu-Rusu, Comparative assessment of the nutritional and sanogenic features of certain cheese sorts originating in conventional dairy farms and in "mountainous" quality system farms, Agriculture 14 (2024) 172, https://doi.org/10.3390/agriculture14020172.
[8] L. Zhang, P. Wang, S. Li, D. Wu, Y. Zhong, W. Li, H. Xu, L. Huang, Differentiation of Mountain- and Garden-Cultivated Ginseng with Different Growth Years Using HS-SPME-GC-MS coupled with chemometrics, Molecules 28 (2023) 2016, https://doi.org/10.3390/molecules28052016.
[9] S. Segato, G. Galaverna, B. Contiero, P. Berzaghi, A. Caligiani, A. Marseglia, G. Cozzi, Identification of Lipid Biomarkers To Discriminate between the Different Production Systems for Asiago PDO Cheese, J. Agric. Food Chem. 65 (2017) 9887–9892, https://doi.org/10.1021/acs.jafc.7b03629.
[10] Regulation - 583/2009 - EN - EUR-Lex, (n.d.). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32009R0583 (accessed July 16, 2024).
[11] Seals and marks, (2024). https://www.parmigianoreggiano.com/product-guide-seals-and-marks#5 (accessed August 8, 2024).
[12] E. Gorrochategui, J. Jaumot, S. Lacorte, R. Tauler, Data analysis strategies for targeted and untargeted LC-MS metabolomic studies: Overview and workflow, TrAC Trends Anal. Chem. 82 (2016) 425–442, https://doi.org/10.1016/j.trac.2016.07.004.
[13] R. Stolt, R.J.O. Torgrip, J. Lindberg, L. Csenki, J. Kolmert, I. Schuppe-Koistinen, S. P. Jacobsson, Second-order peak detection for multicomponent high-resolution LC/MS data, Anal. Chem. 78 (2006) 975–983, https://doi.org/10.1021/ac050980b.
[14] R. Tautenhahn, C. Böttcher, S. Neumann, Highly sensitive feature detection for high resolution LC/MS, BMC Bioinf. 9 (2008) 504, https://doi.org/10.1186/1471-2105-9-504.
[15] R. Tauler, Multivariate curve resolution applied to second order data, Chemom. Intell. Lab. Syst. 30 (1995) 133–146, https://doi.org/10.1016/0169-7439(95)00047-X.
[16] Delegated regulation - 665/2014 - EN - EUR-Lex, (n.d.). https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32014R0665 (accessed August 8, 2024).
[17] R. Tauler, E. Gorrochategui, J. Jaumot, R. Tauler, A protocol for LC-MS metabolomic data processing using chemometric tools, Protoc. Exch. (2015), https://doi.org/10.1038/protex.2015.102.
[18] Compound Discoverer Software - IT, (n.d.). https://www.thermofisher.com/uk/en/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-software/multi-omics-data-analysis/compound-discoverer-software.html (accessed August 8, 2024).
[19] L. Hao, J. Wang, D. Page, S. Asthana, H. Zetterberg, C. Carlsson, O.C. Okonkwo, L. Li, Comparative Evaluation of MS-based Metabolomics Software and Its Application to Preclinical Alzheimer's Disease, Sci. Rep. 8 (2018) 9291, https://doi.org/10.1038/s41598-018-27031-x.
[20] J. Jaumot, A. de Juan, R. Tauler, MCR-ALS GUI 2.0: New features and applications, Chemom. Intell. Lab. Syst. 140 (2015) 1–12, https://doi.org/10.1016/j.chemolab.2014.10.003.
[21] M. Pérez-Cova, C. Bedia, D.R. Stoll, R. Tauler, J. Jaumot, MSroi: A pre-processing tool for mass spectrometry-based studies, Chemom. Intell. Lab. Syst. 215 (2021) 104333, https://doi.org/10.1016/j.chemolab.2021.104333.
[22] E. Gorrochategui, J. Jaumot, R. Tauler, ROIMCR: A powerful analysis strategy for LC-MS metabolomic datasets, BMC Bioinf. 20 (2019), https://doi.org/10.1186/s12859-019-2848-8.
[23] F.Y. Yamamoto, C. Pérez-López, A. Lopez-Antia, S. Lacorte, D.M. de Souza Abessa, R. Tauler, Linking MS1 and MS2 signals in positive and negative modes of LC-HRMS in untargeted metabolomics using the ROIMCR approach, Anal. Bioanal. Chem. 415 (2023) 6213–6225, https://doi.org/10.1007/s00216-023-04893-3.
[24] C. Pérez-López, B. Oró-Nolla, S. Lacorte, R. Tauler, Regions of Interest Multivariate Curve Resolution Liquid Chromatography with Data-Independent Acquisition Tandem Mass Spectrometry, Anal. Chem. 95 (2023) 7519–7527, https://doi.org/10.1021/acs.analchem.2c05704.
[25] A. de Juan, J. Jaumot, R. Tauler, Multivariate Curve Resolution (MCR). Solving the mixture analysis problem, Anal. Methods 6 (2014) 4964–4976, https://doi.org/10.1039/C4AY00571F.
[26] E.L. Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ruff, H.P. Singer, J. Hollender, Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence, Environ. Sci. Technol. 48 (2014) 2097–2098, https://doi.org/10.1021/es5002105.
[27] R.G. Sadygov, F. Martin Maroto, A.F.R. Hühmer, ChromAlign: a two-step algorithmic procedure for time alignment of three-dimensional LC−MS chromatographic surfaces, Anal. Chem. 78 (2006) 8207–8217, https://doi.org/10.1021/ac060923y.
[28] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemom. Intell. Lab. Syst. 2 (1987) 37–52, https://doi.org/10.1016/0169-7439(87)80084-9.
[29] A.K. Smilde, J.J. Jansen, H.C.J. Hoefsloot, R.-J.-A.-N. Lamers, J. van der Greef, M. E. Timmerman, ANOVA-simultaneous component analysis (ASCA): a new tool for

analyzing designed metabolomics data, Bioinforma. Oxf. Engl. 21 (2005) 3043–3048, https://doi.org/10.1093/bioinformatics/bti476.
[30] M. Anderson, C.T. Braak, Permutation tests for multi-factorial analysis of variance, J. Stat. Comput. Simul. 73 (2003) 85–113, https://doi.org/10.1080/00949650215733.
[31] J. Sun, Y. Xia, Pretreating and normalizing metabolomics data for statistical analysis, Genes Dis. 11 (2024) 100979, https://doi.org/10.1016/j.gendis.2023.04.018.
[32] R.G. Buttery, D.G. Guadagni, L.C. Ling, Flavor compounds. Volatilities in vegetable oil and oil-water mixtures. Estimation of odor thresholds, J. Agric. Food Chem. 21 (1973) 198–201, https://doi.org/10.1021/jf60186a029.
[33] H. Ochi, Y. Sakai, H. Koishihara, F. Abe, T. Bamba, E. Fukusaki, Monitoring the ripening process of Cheddar cheese based on hydrophilic component profiling using gas chromatography-mass spectrometry, J. Dairy Sci. 96 (2013) 7427–7441, https://doi.org/10.3168/jds.2013-6897.
[34] B. Yanibada, U. Hohenester, M. Pétéra, C. Canlet, S. Durand, F. Jourdan, J. Boccard, C. Martin, M. Eugène, D.P. Morgavi, H. Boudra, Inhibition of enteric methanogenesis in dairy cows induces changes in plasma metabolome highlighting metabolic shifts and potential markers of emission, Sci. Rep. 10 (2020) 15591, https://doi.org/10.1038/s41598-020-72145-w.
[35] J. Yang, W. Bai, X. Zeng, C. Cui, Gamma glutamyl peptides: The food source, enzymatic synthesis, kokumi-active and the potential functional properties – A review, Trends Food Sci. Technol. 91 (2019) 339–346, https://doi.org/10.1016/j.tifs.2019.07.022.
[36] K.J. Li, K.J. Burton-Pimentel, E.M. Brouwer-Brolsma, E.J.M. Feskens, C. Blaser, R. Badertscher, R. Portmann, G. Vergères, Evaluating the robustness of biomarkers of dairy food intake in a free-living population using single- and multi-marker approaches, Metabolites 11 (2021), https://doi.org/10.3390/metabo11060395.
[37] F. Masotti, J.A. Hogenboom, V. Rosi, I. De Noni, L. Pellegrino, Proteolysis indices related to cheese ripening and typicalness in PDO Grana Padano cheese, Int. Dairy J. 20 (2010) 352–359, https://doi.org/10.1016/j.idairyj.2009.11.020.
[38] J. Ivanisevic, Z.-J. Zhu, L. Plate, R. Tautenhahn, S. Chen, P.J. O'Brien, C. H. Johnson, M.A. Marletta, G.J. Patti, G. Siuzdak, Toward 'omic scale metabolite profiling: a dual separation-mass spectrometry approach for coverage of lipid and central carbon metabolism, Anal. Chem. 85 (2013) 6876–6884, https://doi.org/10.1021/ac401140h.
[39] M.A. Rasmussen, E. Maslova, T.I. Halldorsson, S.F. Olsen, Characterization of Dietary Patterns in the Danish National Birth Cohort in Relation to Preterm Birth, PLoS One 9 (2014) e93644.
[40] E. Neviani, A. Levante, M. Gatti, The Microbial Community of Natural Whey Starter: Why Is It a Driver for the Production of the Most Famous Italian Long-Ripened Cheeses? Fermentation 10 (2024) 186, https://doi.org/10.3390/fermentation10040186.
[41] D. Bentivoglio, S. Savini, A. Finco, G. Bucci, E. Boselli, Quality and origin of mountain food products: the new European label as a strategy for sustainable development, J. Mt. Sci. 16 (2019) 428–440, https://doi.org/10.1007/s11629-018-4962-x.
[42] C. Perez-Lopez, A. Ginebreda, J. Jaumot, F.Y. Yamamoto, D. Barcelo, R. Tauler, MSident: Straightforward identification of chemical compounds from MS-resolved spectra, Chemom. Intell. Lab. Syst. 245 (2024) 105063, https://doi.org/10.1016/j.chemolab.2024.105063.