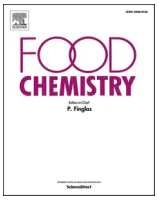
Food Chemistry 456 (2024) 139986

Contents lists available at ScienceDirect

Food Chemistry

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



^{1}H NMR profiling and chemometric analysis for ripening and production characterization of Grana Padano cheese

Valentina Maestrello [a,b], Pavel Solovyev [a,*], Angelo Stroppa [c], Luana Bontempo [a], Pietro Franceschi [a]

[a] Fondazione Edmund Mach (FEM), via E. Mach 1, 38098 San Michele all'Adige, Italy
[b] University of Trento, via Mach 1, 38098 San Michele all'Adige, (TN), Italy
[c] Consorzio Tutela Grana Padano, Via XXIV Giugno 8, 25015, San Martino Della Battaglia, Desenzano del Garda, BS, Italy

ARTICLE INFO

Keywords:
Ripening
Cheese
Metabolomics
NMR
Chemometrics

ABSTRACT

Grana Padano (GP) cheese is a renowned PDO Italian cheese whose nutritional characteristics and market price are influenced by the ripening stage. In this work, it was demonstrated that the combined use of untargeted ^{1}H NMR profiling and chemometric analysis can be used as a powerful tool to quantitatively characterize GP ripening and production, focusing on both aqueous and lipid fractions. An initial exploratory analysis revealed substantial variations in the aqueous fraction attributable to aging time, year and season of production. Multivariate analysis was adopted to show these differences, mainly attributable to amino acids. In contrast, the lipid fraction analysis highlighted the impact of production season on fatty acid unsaturation, influenced by feed variations. As regards the production process, this study focuses on the variations induced by bactofugation. In this respect, the aqueous fraction was found to be extensively influenced by this centrifugation step, affecting compounds crucial to organoleptic characteristics.

1. Introduction

Grana Padano (GP) cheese is one of the most renowned cheeses in the world and it is protected by the Protected Denomination of Origin label (PDO). It is a high-quality food product, owing to its unique amino acid and nutrient profiles and its sensorial attributes which, from a commercial point of view, make it highly priced. These characteristics evolve during ripening; hence the aging time defines its quality and economic value. On the market, the cheese is sold with 3 different aging times: Grana Padano aged between 9 and 16 months, with a delicate taste; aged more than 16 months, with a stronger tastiness; and Riserva aged more than 20 months, considered of the highest quality (Consorzio Tutela Grana Padano, 2017); each of them with different prices.

For economic reasons, GP production is controlled by strict specification rules (*The PDO*, 2023), which can be updated only if the changes do not impact on the peculiar and discriminative characteristics of the final product. Remarkably, however, plans and rules for GP protection are mostly documentary and empirical. This constitutes a limitation for the objective evaluation of updated production steps. In this respect, in recent years an additional step is under evaluation by the *Consortium,* i.e. bactofugation. This centrifugation step at high pressures allows for the separation of spores and thermos-resistant microorganisms from milk (Farkye, 2004), with potential positive impacts on production safety and quality improvement. The removal of those microorganisms could nonetheless influence the characteristics of cheeses, because some important flavor compounds could not be produced anymore due to the lack of microorganism-induced processes (Gésan-Guiziou, 2010).

Focusing on the protection of Grana Padano cheese, in order to understand, benchmark and protect the quality of such an important commodity, advanced chemical characterization technologies could provide key quantitative evidence. A demonstration of that is the inclusion of the analysis of the profile of amino acids, minerals and isotopes in the "Grana Padano PDO production specification rules". Among the diverse analytical techniques applicable to food profiling, nuclear magnetic resonance spectroscopy (NMR), often combined with chemometric analysis (Balthazar et al., 2021), has proven to be highly effective. This is attributed to its high throughput, chemical selectivity, and reproducibility.

NMR has already been used to study the metabolite profile of dairy products in general (Scano, Cusano, Caboni, & Consonni, 2019), as well

* Corresponding author.
E-mail addresses: valentina.maestrello@unitn.it (V. Maestrello), pavel.solovyev@fmach.it (P. Solovyev), a.stroppa@granapadano.com (A. Stroppa).

https://doi.org/10.1016/j.foodchem.2024.139986
Received 31 January 2024; Received in revised form 23 May 2024; Accepted 3 June 2024
Available online 4 June 2024

as to investigate changes during ripening times of Cheddar (Chen et al., 2021), Fiore Sardo (Piras et al., 2013) and Parmigiano Reggiano (Consonni & Cagliani, 2008). Other studies adopted different "declinations" of NMR spectroscopy; for example, HRMAS NMR was used, outlining the fact that it is a quick tool for the evaluation of ripening steps of Parmigiano Reggiano cheese (Shintu & Caldarelli, 2005), while another study adopted fast-field cycling relaxometry on Parmigiano Reggiano cheese for the evaluation of changes regarding the physico-chemical properties (Conte et al., 2021). NMR was also adopted to study the variation in the amino acid profile and the behavior of water during ripening with a targeted approach (de Angelis et al., 2000). To date, just a few other works have investigated the ripening process of Grana Padano cheese, mainly focusing on specific compounds or classes of compounds. For example, the behavior of casein and specific peptides was studied during aging with chromatography and electrophoresis (Masotti, Hogenboom, Rosi, De Noni, & Pellegrino, 2010). Likewise, pyroglutamic acid (Mucchetti et al., 2000), phosphopeptides (Ferranti et al., 1997), *Clostridium* spores (Morandi, Silvetti, & Brasca, 2022) and LAB strains (Pogačić et al., 2013) were studied in relation to ripening, to develop new methods for its assessment. Other studies adopted thermogravimetric techniques to understand the behavior of water with the matrix and try to assess the aging index (de Angelis et al. 1999). Only a recent study involved magnetic resonance imaging (MRI) as a tool to monitor ripening stages keeping the sample intact and examine the cheese in its entirety (Mulas, Anedda, Longo, Roggio, & Uzzau, 2016). In addition to changes due to ripening time, changes due to the influence of the season of the year and factory plant production have been studied, e.g. the volatile organic compounds profile was studied in relation to Trentingrana cheese (Ricci et al., 2022).

This work aims to investigate whether an untargeted NMR-based metabolomics approach could be used not only to detect the compositional modifications of Grana Padano during ripening, but also to try to quantitatively evaluate the impact of bactofugation on the final characteristics of the product. The focus is on an extensive number of samples covering different ripening times, seasons and years of production, considering both the aqueous and the lipid fraction.

The results demonstrate that untargeted NMR metabolomics can be effectively used to objectively evaluate the Grana Padano production and ripening processes, adding a science-based parameter to the empirical assessment currently used. This could be a support for the *Consortium* both for the standardization of quality assessment and for the benchmarking of "new" production steps.

2. Materials and methods

2.1. Samples

For the study of ripening times, 110 real PDO Grana Padano (Grana Padano) samples (collected by *Consorzio Tutela Grana Padano*) with different aging times (between 9 and 29 months) were considered. For some of those authentic samples, a second "experimentation" sample was produced with the same milk that underwent the "bactofuge" step, collecting 217 samples in total.

2.2. Experimental procedures

The experimental procedure followed the path reported in the paper of (Andersen, Bosetti, Mancini, & Bontempo, 2022). The Grana Padano samples were cut into pieces and then grated. After the lyophilization, 100 mg were weighted and added at 900 μL de-ionized H_2O and 100 μL D_2O (99.9% isotopic purity containing 0.03% 3-(Trimethylsilyl) propionic-2,2,3,3-d_4 acid sodium salt or TMSP-d_4). Samples were centrifuged at 12 rpm for 15 min and then filtered to separate the solid particles from the liquid, and 600 μL filtrate was loaded into 5 mm NMR tubes. The sample preparation for the lipid fraction was the same, except for the addition of 900 μL of $CDCl_3$. No pH adjustments were needed: only the peak assigned to histamine shifted, but it did not cause any problems in the analyses. Acquisition of the spectra was carried out using Bruker Avance Neo spectrometer with a 600 MHz base frequence (5 mm sample tubes) and SampleXpress 60-position autosampler (Bruker BioSpin GmbH, Rheinstetten, Germany). The software used was Topspin 4.1.3 software in the automation mode with Icon NMR 5.2.3.

2.2.1. NMR experiments

The aqueous experiments were acquired through *noesygppr1d* pulse sequence with automatic adjustment of water signal suppression frequency; the size of the spectrum was 15 ppm, time domain consisted of 64 K data points, scans numbered 128 and dummy scans 2, the time for relaxation delay was 10 s, and the receiver gain for all spectra was set at 2.25. The lipid experiments were acquired with simple *zg* pulse sequence, the size of the spectrum was 12 ppm, time domain consisted of 64 K data points, scans numbered 64 and dummy scans 4, the time for relaxation delay was 10 s, and the receiver gain for all spectra was set at 4. A second experiment was performed for the lipid fraction: a *noesygppr1d.wvm* pulse program, in order to suppress the main signals of the previous experiment, with the same parameters as the previous one, except for NS of 128, D1 of 3 s and RG set at 101. Spectra were processed in the TopSpin software with apk0.NOE phase correction and the size of real spectrum set to 131,072 (128 K, 2xTD) data points.

2.2.2. Data processing

Peak assignment was performed manually, based on literature data and with the aid of AssureNMR software (Bruker BioSpin GmbH, Rheinstetten, Germany), using HMDB database (Wishart et al., 2022). Quantitative analysis in the aqueous fraction was performed with AssureNMR software through an external standard method using the so-called ERETIC technique (electronic reference to access in vivo concentrations), based on the PULCON principle (Hong et al., 2013). The identification of lipids was performed with the aid of literature data (Ralli & Spyros, 2023) and the quantification of the integrals of the signals was calibrated to the integral area of alpha carboxyl protons (2.31 ppm) of all fatty acids (Alexandri et al., 2017) in percentages. The repeatability of the method was confirmed by periodically acquiring one manufacturer standard (2 mM sucrose solution) against another one, with the same experiments described above. All the quantified signals have a signal-to-noise ratio higher than 10.

2.3. Statistical analysis

The spectra were binned in segments of 0.04 ppm. The bucket table was obtained with the AssureNMR software, and all statistical analyses were performed with R (R Core Team, 2021). The information obtained from the two NMR experiments performed on the lipid fraction were merged for the statistical analysis by substituting the zero intensities of the suppressed regions with the intensities of the peaks from the experiment with no suppression. All data management and plots were obtained with tidyverse (Wickham et al., 2019), tidymodels (Kuhn & Wickham, 2020), mdatools (Kucheryavskiy, 2020) and broom (Robinson, Hayes, & Couch, 2024) packages. PCA plots were obtained with FactoMineR (Lê, Josse, & Husson, 2008) and FactoExtra (Ramadan, Kamel, Taha, El-Shabrawy, & Abdel-Fatah, 2020) packages.

All values in the bucket tables were mean-centered and Pareto scaled to reduce the relative impact of high-intensity buckets (Kandasamy et al., 2020), keeping the overall structure of the dataset intact (van den Berg, Hoefsloot, Westerhuis, Smilde, & van der Werf, 2006). Initially, PCA was performed to visualize the factors that had the greatest impact on the dataset variance and identify potential outliers. The results of this multivariate analysis were visualized in terms of score plots, which can mostly show clustering of the data.

The experimental design was very complex though, due to the numerous variables that could influence the metabolome of the samples: production year, aging time, production year. Actually, the samples

were not produced in the same year and season, and as showed in the PCA (Fig. 1), these factors had an impact on the sample profiles. Since the separation of the groups for each variable (season, year and aging) was only partial, the quantification of the separation of the groups was performed following an approach based on LDA (Longobardi et al., 2015) and reported in Supplementary material.

To separate the influences of these factors on the dataset, multivariate and univariate analyses were combined: multivariate analysis made it possible to exploit the correlation between variables and to see their influence on the complete dataset, while univariate analysis allowed us to highlight the variables that were more responsive to the individual experimental design factors.

In all cases, the dataset was firstly reduced to a lesser number of buckets with the method of medians to eliminate those referred to noise that would be irrelevant for the statistics. This method consists in the calculation of the median of the intensity of each bucket among the spectra and selecting those buckets with the highest values along the distribution curves (see Supplementary Material, Supplementary Fig. A.1). After this filtering step, the intensity of each bucket was modeled as a function of season and year of production by a *glm* (considering the interaction between the season and year of production). The univariate *glms* made it possible:

- to remove the effect of nuisance variables (year and season) from the dataset, "concentrating" the effect of aging on the residuals of individual models.
- to identify the buckets that were more sensitive to the season and the year of production (low *p* values for each individual factor). These buckets were then annotated as reported above.

After the removal of the effects of year and season, Partial Least-Square Regression (PLSR) was then applied to the residual matrix to construct a regression model able to link NMR profile and age for the individual samples. In particular, the matrix of the residuals of individual *glms* was used as a predictor of aging time. In order to avoid overfitting (a well-known limitation of PLSR), the dataset was calibrated and validated (the dataset was split 80:20). Firstly, the calibration set was used to optimize the number of components with a leave-one-out method. The test set was then used to evaluate the unbiased model performances in terms of root mean square error (RMSE). As a further validation step, the same model was applied to a series of "random" datasets (20) where the response vector (aging time) was randomly associated with the samples. If the RMSE error of these random models was higher than the one obtained with the correct y labels, this was considered a clear indication that the NMR profiles were informative about the age of individual samples. The most important variables were obtained from the ranking of regression coefficients of the PLS model, because they assess the importance of the variables that are mostly able to "separate" the different classes (Debik, Sangermani, Wang, Madssen, & Giskeødegård, 2022).

As far as bactofugation was concerned, PLS regression was replaced by PLS-DA. Even in this case, the dataset was split into calibration and test set (80:20). The number of components was tuned with the leave-one-out method and the model performance assessed in terms of specificity, sensitivity and overall classification error (Kandasamy et al., 2020). The significant buckets were extracted with the analysis of regression coefficients, as done with PLSR.

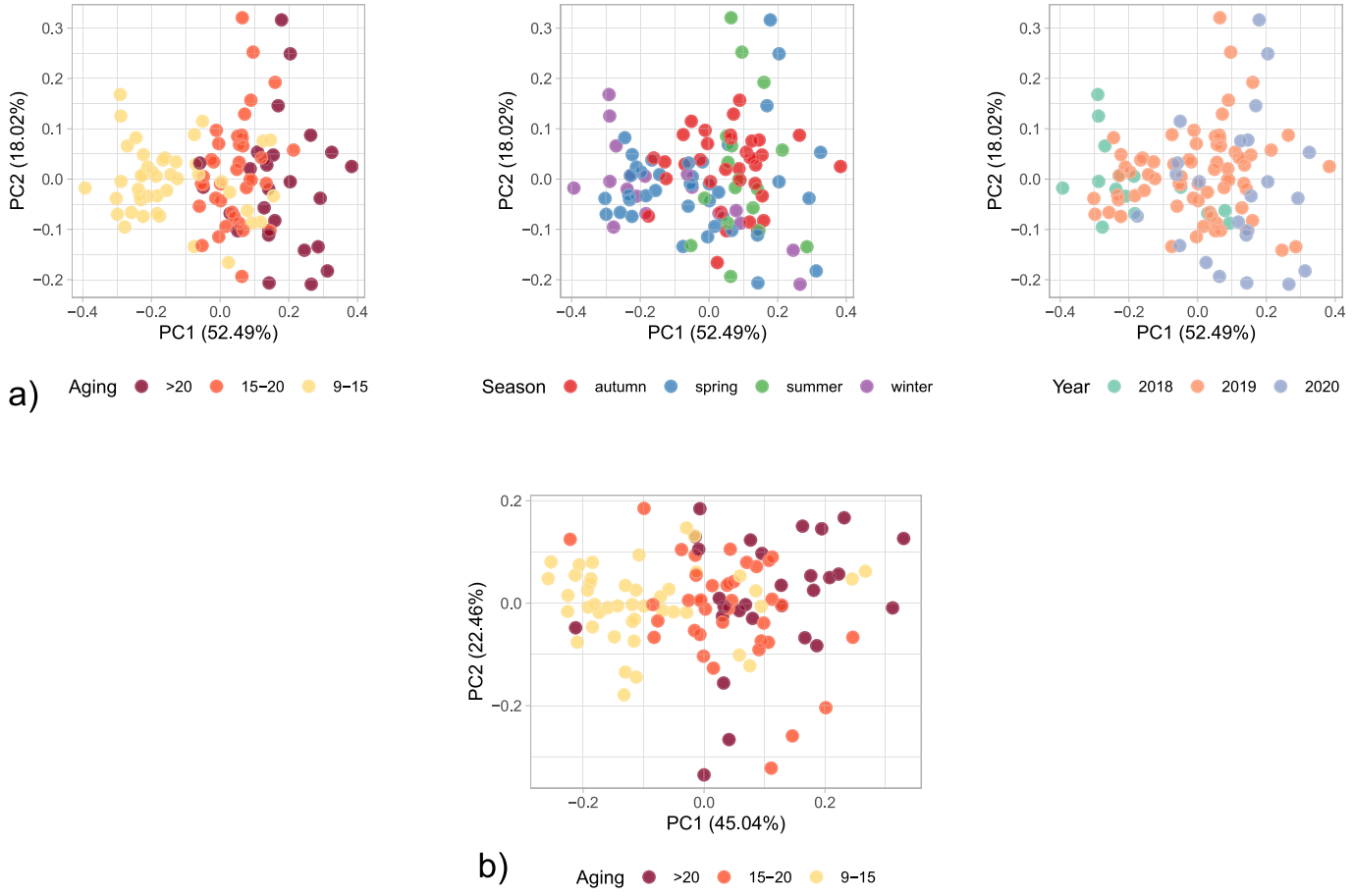


Fig. 1. a) PCA plot of the aqueous fraction. The figures are the same plot; both samples are colored for three different variables: ripening time, production season, and production year; b) PCA plot obtained with residuals of glm model, aqueous fraction. Samples are colored according to their commercial aging segment.

3. Results and discussion

3.1. Aging

3.1.1. Aqueous fraction

The score plot of the initial PCA on the aqueous fraction is shown in Fig. 1a. The plot highlights the role of the different design factors on the PCA plane, which accounts for 70% of total variability.

The three panels are used to illustrate the contribution of the three major study factors to the distribution of the points in the PCA plane. The separation of the three aging times is evident, even if some partial partitioning of the different years and seasons can be also observed. In order to disentangle the effects of the different factors, multivariate analyses were combined with univariate modeling (see Material and Methods). The results of the glm indicated that some free amino acids (FAA) were influenced by both year and season of production (valine, threonine, alanine, isoleucine, lysine, proline, serine and phenylalanine), and the amounts of glycerol and pyroglutamate similarly varied according to these factors (Supplementary Fig. A.2). In particular, the results on acetic acid, lysine and serine are in agreement with a previous study (Careri, Spagnoli, Panari, Zannoni, & Barbieri, 1996). In fact, as reported in various studies, the seasonal and pasture variation have an influence on the raw milk and thus on the final product (Ricci et al., 2022).

Looking at individual factors, the compounds mostly influenced by the year of production were methionine and asparagine and the organic acids, acetic acid, glutamic acid and aspartic acid, while the production season mostly influenced the amount of sucrose. This result is not unexpected due to the different feed: during winter and autumn, the feed is mostly from silage, whereas in spring and summer the feed is mostly grain corn, ending up in different compositions of the feed itself.

The matrix obtained with the residuals of the glm was used to perform PCA and analyze the distribution of samples without the influence of year and season of production (Fig. 1b).

The score plot clearly shows that aging time is still the main driver of sample variability, since the three age groups are partially separated along PC1. As described in Materials and Methods, PLSR regression was applied to the residual matrix to identify the aging “biomarkers”. The PLS score plot is shown in Fig. 2 and clearly highlights the age trend along the first latent variable.

Compounds with the larger regression coefficients are shown in Supplementary Fig. A.3, and in terms of chemical class, they were identified either as organic acids or as amino acids.

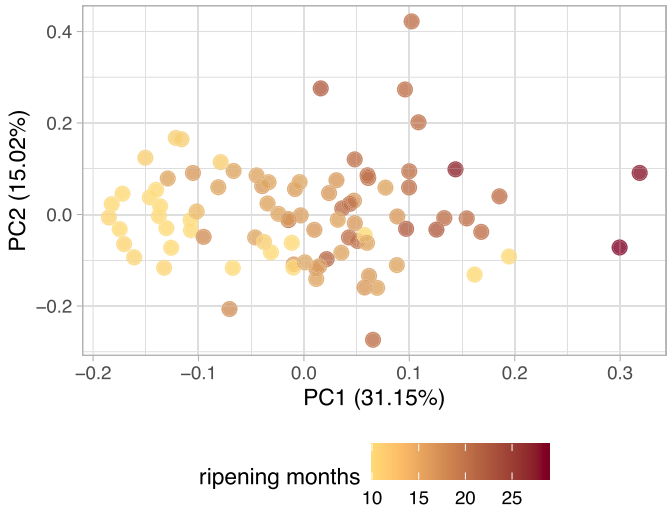


Fig. 2. PLSR model obtained with residuals of the glm model; aqueous fraction colored according to the aging time.

In general, lipolysis and proteolysis are important steps in the maturation process, because they produce free amino acids and free fatty acids, which influence the aroma and flavor of the final product. Water content also has an important role, giving a strong contribution to texture and stability (Mulas et al., 2016). Those mainly responsible for these reactions are Lactic Acid Bacteria (LAB) (Piras et al., 2013): they induce several microbial pathways, which lead to different metabolic compositions of the final product (Rocchetti et al., 2018) and they are different for each type of hard cheeses, being influenced by farm environment and milking parlor environment (Gatti, Bottari, Lazzi, Neviani, & Mucchetti, 2014).

As far as the individual trends of organic acids are concerned, acetate tended to increase as a product of proteolysis (Eugster, Fuchsmann, Schlichtherle-Cerny, Bütikofer, & Irmler, 2019), while pyruvate remained constant along the time of aging (Careri et al., 1996); instead, lactic acid, which is expected to increase (McSweeney, 2004), had a weird trend, probably because it could be bound to proteins or crystallized, making its analytical detection difficult.

In the case of amino acids, valine and threonine increased with ripening time, as expected for known markers of old cheeses (Consonni & Cagliani, 2008). Other amino acids whose content is related to ripening time grew, such as lysine (Ciriello, Cataldi, Crispo, & Guerrieri, 2015), alanine, methionine, serine and phenylalanine (de Angelis et al., 2000). Proline, which is abundant in casein, increased, and aspartic acid, consequentially on the degradation of asparagine (Masotti et al., 2010), followed the same trend. Pyroglutamic acid is known to be high in old cheeses (Mucchetti et al., 2000), whereas tyrosine decreased (Consonni & Cagliani, 2008), probably due to the formation of crystals after the unraveling of proteins. At odds with other investigations (Consonni & Cagliani, 2008), isoleucine was found to increase with ripening time: an explanation could be that the dataset is really wide in composition and the feeding was different from other studies, because in our work pasture was not considered. Remarkably, most of the compounds that h increased during ripening have important roles as aroma precursors, confirming the higher organoleptic characteristic of aged cheeses (Niro et al., 2017).

3.1.2. Lipid fraction

Following the same approach adopted for the aqueous fraction, PCA was used in the exploratory analysis for the lipid fraction. The score plot of the complete dataset is presented in Supplementary Fig. A.4 and it clearly shows the presence of a group of outlying samples that dominates the dataset variability. These outliers showed higher intensities for signals assigned to linolenic acid (2.82 ppm, Boccia, Cusano, Scano, & Consonni, 2020), but upon metadata inspection it was not possible to identify informative links among them. In order to investigate the impact of study factors on this fraction, the 7 outlying samples were excluded from the subsequent analysis.

The updated PCA plot without outliers is reported in Fig. 3a, where the PC1 x PC2 plane accounts for 50% of the overall dataset variability. As before, the three plots highlight the contribution of the study factors on sample distribution. At odds with what was observed for the aqueous fraction, year and season of production here are the major factors influencing the NMR profile. As before, univariate glm was used to pinpoint compounds most affected by these two factors. It is important to highlight, however, that in the case of the lipid fraction there is a lower level of chemical detail, due to the fact that fatty acids (FA) can only be characterized at the level of classes of protons (e.g. methylenic protons, methylic protons, protons of acyl moieties).

Considering individual glms, as far as the year of milk production is concerned, no coherent trend was visible, because the specific compound assignment is not possible in the lipid fraction and specific characteristics cannot be recognized. The factor of the year is difficult to control because it depends on climatic conditions, and the management system of the feeding (Araújo et al., 2012) that changes year by year.

The production season, though, had an important role, especially for

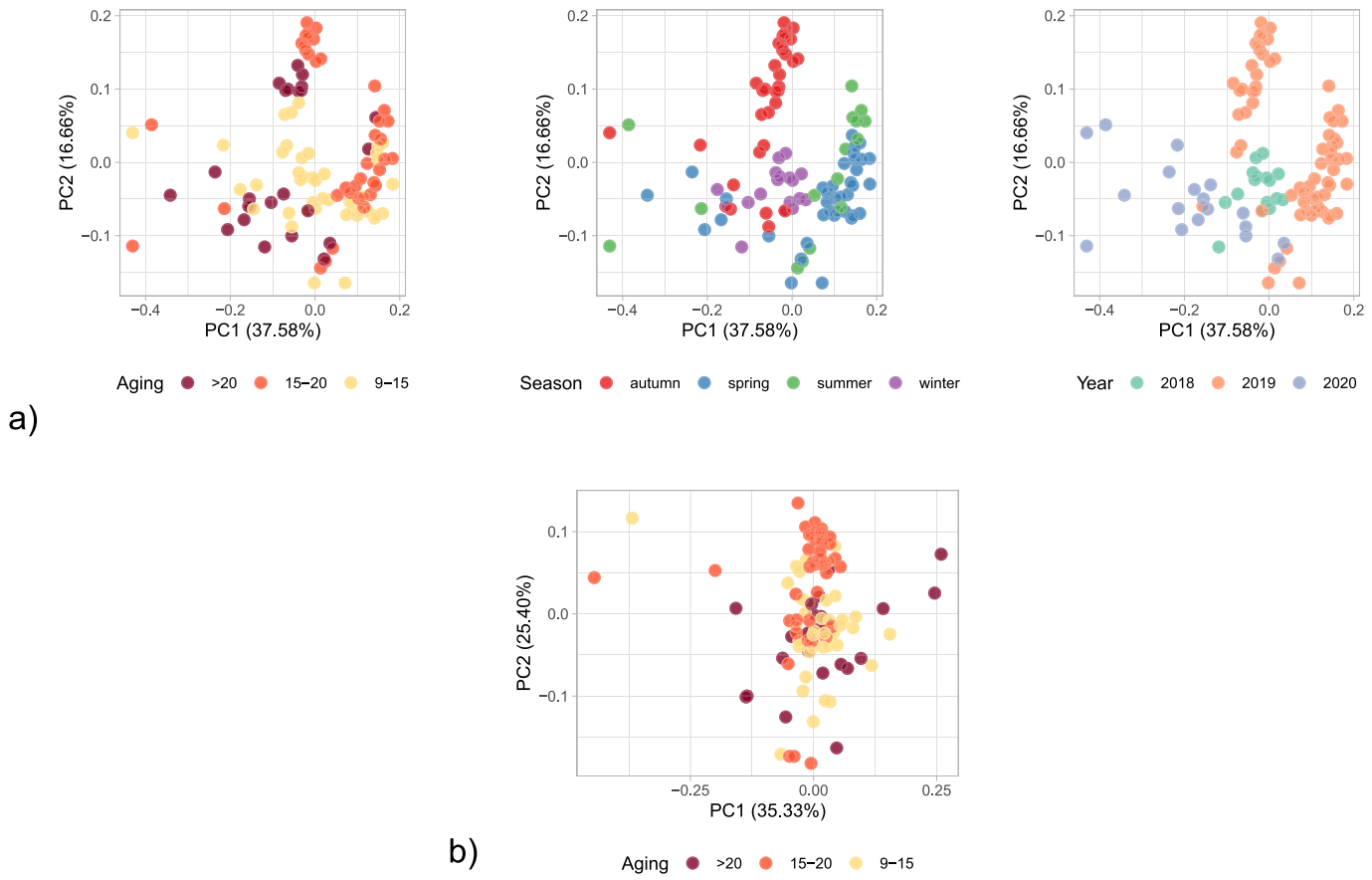


Fig. 3. a) PCA plot of the lipid fraction, no outliers. Samples are colored according to the different metadata: production season and year and aging time; b) PCA plot obtained with the residuals of glm, no outliers. The colors are according to their commercial aging segment.

unsaturation of the FAs. Unlike other studies that reported an increase in unsaturation during summer (Prandini, Sigolo, Cerioli, & Piva, 2009), in this study there was a decrease of unsaturation (saturation and unsaturation lipids are strongly correlated because they interconvert) during spring and summer, due to the different feed (Supplementary Fig. A.5).

An explanation could be that in other studies cows went to pastures in summer, whereas for this work cows stayed lowland with a change in the feed because, usually, starch comes from grain corn in summer and from silage during winter, and this causes an increase in saturated FA and the amount of conjugated linoleic acids (CLA) during summer.

However, the focus of this study was on highlighting the differences during ripening. As it can be seen from the score plot of the PCA (Fig. 3b), even when removing the influences of year and season of production, the ripening months are not clearly distinguished, but this only indicates that ripening is not the major source of variability in residuals. To assess whether the lipid profile and ripening are somehow associated, the PLSR was performed, as it was previously described (Supplementary Fig. A.6).

In this case, the RMSE was higher than the one obtained for the aqueous fraction, revealing a weaker association of the lipid profile to the aging time. The permutation test, however, showed that the model performed better than a random one, indicating that some form of correlation still existed. As already discussed, it was not possible to pinpoint specific compounds, but the larger regression coefficients were the ones associated with the total amount of CLA, 1,2-DAG and TAG. The observed trends are consistent with the presence of enzymatic processes acting on triglycerides (Malacarne et al., 2009) (Supplementary Fig. A.7).

3.2. *Bactofuge*

3.2.1. *Aqueous fraction with and without bactofugation step*

The PCA score plot of the dataset containing GP and bactofugated samples is shown in Fig. 4. The PC1xPC2 plane accounts for a large part of the variance and the separation between the two classes of samples is

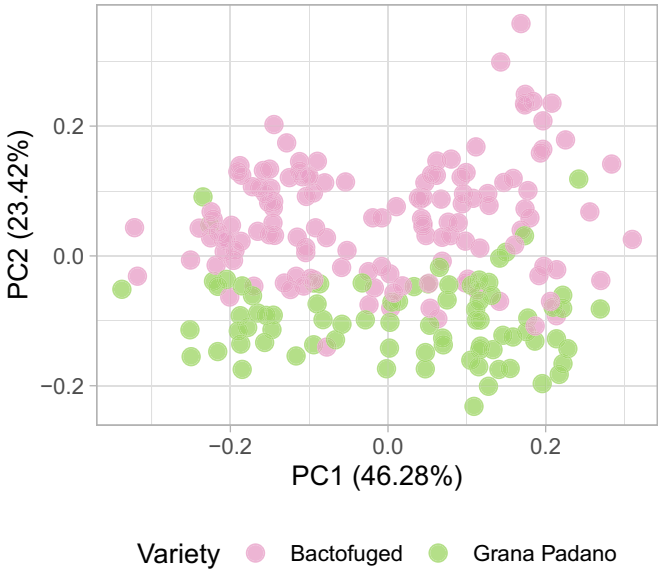


Fig. 4. PCA plot, aqueous fraction, colored according to the "traditional GP vs Bactofuged GP" class.

clear, indicating that the presence of this additional step strongly affected the composition of the aqueous fraction.

As discussed in the Material and Methods section, PLS-DA was used to identify the most relevant "biomarkers". As expected, the model had very good sensitivity and specificity (over 0.9), performing a good separation of the two classes with a very low misclassification rate (3 out of 45). The regression coefficients were used to identify the most significant compounds and the profile of the annotated buckets are shown in Fig. 5.

The samples produced with the "traditional" procedure resulted in higher organic acids (acetate, formate, pyruvate, lactate) and glycerol and lower amounts of threonine, tyrosine and pyroglutamate (Fig. 5).

The higher amounts of acids are in line with the higher amounts of microorganisms and spores that are thermos-resistant and not completely eliminated by the production process, while tyrosine is an important compound for the formation of granules that give the typical characteristic to the paste of Grana Padano cheese. Anyway, an unexpected result is the lower amount of pyroglutamic acid in samples produced with the traditional process: its formation is mostly enzymatic (Mucchetti et al., 2000) and is accelerated by thermophilic lactic acid bacteria (Mucchetti, Locci, Massara, Vitale, & Neviani, 2002). It thus seems that LAB can perform better without any other microorganism or spore.

3.2.2. *Lipid fraction with and without bactofugation step*

As in the case of the GP aging dataset, the lipid fraction was not extensively affected by the bactofugation process, but, interestingly, the PCA highlights the different time of production (Fig. 6).

In fact, by coloring the samples according to their month of production, it is possible to detect three different clusters that are related to three different seasons: spring/summer, autumn, winter. These results were discussed in the previous section.

Since the production season was the major impact factor on the distribution of samples, a glm was performed, similarly to the ripening analysis. After removing the influence of season and year, with the matrix of the residuals a PLS-DA was performed to see whether a discrimination was now possible. The PLS-DA model has a lower specificity compared to the one obtained with the aqueous fraction (around 0.7) and a rather high misclassification error (16 misclassified out of 45); nonetheless, it was still possible to correctly classify the larger part

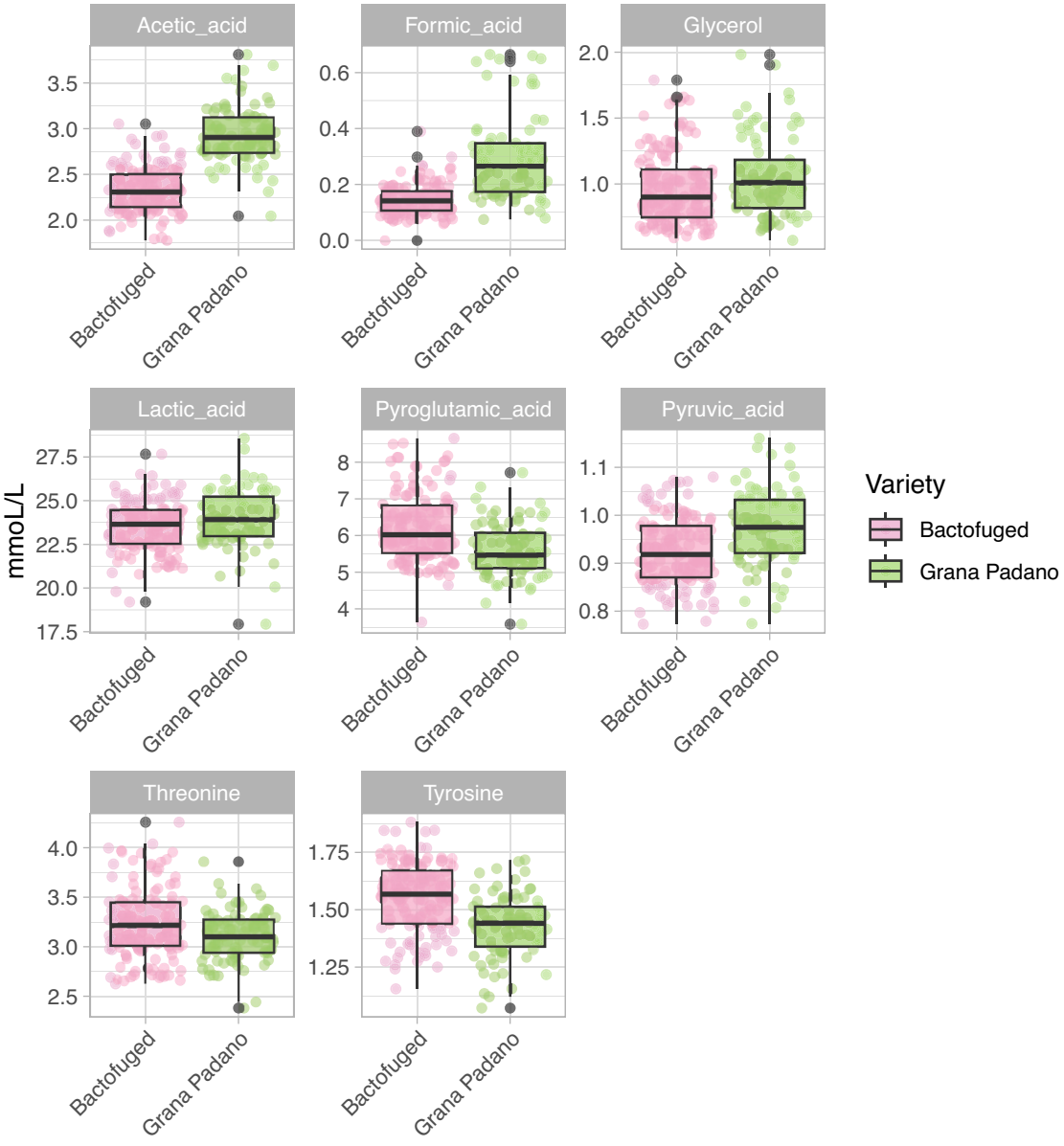


Fig. 5. quantification of significant compounds from the PLSR model, aqueous fraction.

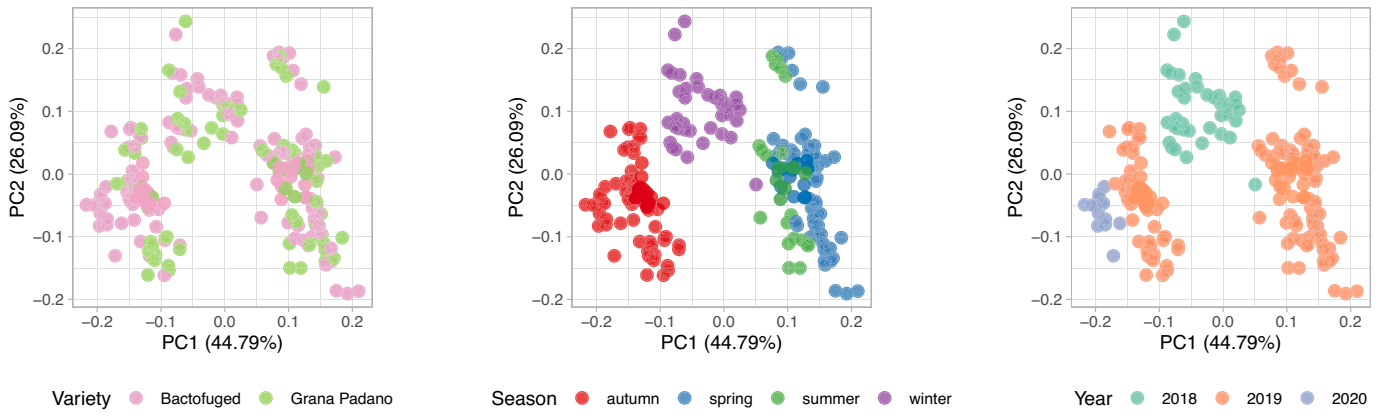


Fig. 6. PCA plot of lipid fraction, colored according to production season and the "Grana Padano traditional and bactofuged" class.

of the test set. In terms of biomarker identification, the samples produced with the bactofugation step had a lower amount of 1,2-diacylglycerols (Supplementary Fig. A.8). A suggested explanation could be that lipases produced by milk psychrotrophs are no longer present in the bactofuged milk, and the lysis of triglycerides is accordingly lower (Ribeiro, & Peruzi, Bruzaroski, Tamanini, Lobo, Alexandrino, Conti, Alfieri, & Beloti, 2019).

4. Conclusions

In this work, two important aspects regarding cheese economic value have been tackled: ripening processes and production process. The aqueous and lipid fractions have been studied with NMR spectroscopy following an untargeted approach. The statistical treatment was complex due to the different factors influencing the profiles and different models were applied, considering both univariate and multivariate approaches. Our investigation demonstrated that aqueous fraction was influenced by ripening time and several possible biomarkers were identified, mainly associated with amino acids and organic acids linked to processes induced by LABs. The lipid fraction, instead, was mainly influenced by the season of production, and this can be linked to the fact that, depending on the season and weather conditions, the cow feed changes with strong impacts on milk composition. Despite this fact, a reliable PLSR model on this fraction highlighted minor evolutions of the lipid profile regarding the unsaturation level during aging, most likely linked to lipolysis. The second part of our study involved the assessment of the quantitative effects of bactofugation on the cheese compositional profile. It was possible to highlight the differences in profile for both aqueous and lipid fractions between samples produced in the "traditional" way and samples produced with the additional centrifugation step, giving a strong and scientific evaluation basis to the *Consortium* for a decision for or against its inclusions in the procedural guidelines. In fact, samples produced with the "traditional" process were found to have higher amounts of organic acids and a low amount of tyrosine, which are important compounds regarding the organoleptic characteristics. This information can thus be useful in the decision process on whether or not to include this step in the specification rules. The research hypothesis to adopt an untargeted multivariate statistical approach with a simple NMR analysis to investigate the metabolome changes influenced by ripening and production processes was satisfied in this work. This study lays the groundwork for the inclusion of untargeted metabolomics in official procedure for high quality cheese protection and could lead the path for other studies to test its application and performances on other cheese matrices.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2024.139986.

Funding sources

The authors acknowledge the financial support provided by ADP funded by the Autonomous Province of Trento, Italy, and the Program of the Autonomous Province of Trento (Italy) with EU co-financing (Fruitomics), Grant/Award Number: FESR 2014–2020.

CRediT authorship contribution statement

Valentina Maestrello: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Formal analysis, Conceptualization. **Pavel Solovyev:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Angelo Stroppa:** Resources, Project administration. **Luana Bontempo:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Pietro Franceschi:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, 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

The data that has been used is confidential.

Acknowledgment

The authors acknowledge the support of the Consortium for the Protection of Grana Padano.

References

Alexandri, E., Ahmed, R., Siddiqui, H., Choudhary, M. I., Tsiafoulis, C. G., & Gerothanassis, I. P. (2017). High resolution NMR spectroscopy as a structural and analytical tool for unsaturated lipids in solution. *Molecules, 22*(10), 1663. https://doi.org/10.3390/molecules22101663

Andersen, J., Bosetti, M., Mancini, A., & Bontempo, L. (2022). Improvement of Caciotta-like cheese nutritional value by means of enrichment with blackcurrant (Ribes nigrum) and cornelian cherry (Cornus mas). *Frontiers in Nutrition, 9*, 1023490. https://doi.org/10.3389/fnut.2022.1023490

de Angelis, C. S., Curini, R., Delfini, M., Brosio, E., D'Ascenzo, F., & Bocca, B. (2000). Amino acid profile in the ripening of Grana Padano cheese: A NMR study. *Food Chemistry, 71*(4), 495–502. https://doi.org/10.1016/S0308-8146(00)00192-8

Araújo, K. B. S., Rangel, A. H. N., Fonseca, F. C. E., Aguiar, E. M., Simplício, A. A., Novaes, L. P., & Júnior, D. M. L. (2012). Influence of the year and calving season on production, composition and mozzarella cheese yield of water buffalo in the state of

Rio Grande Do Norte, Brazil. *Italian Journal of Animal Science, 11*(1), Article e16. https://doi.org/10.4081/ijas.2012.e16

van den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K., & van der Werf, M. J. (2006). Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics, 7*, 142. https://doi.org/10.1186/1471-2164-7-142

Boccia, A. C., Cusano, E., Scano, P., & Consonni, R. (2020). NMR lipid profile of Milk from alpine goats with supplemented hempseed and linseed diets. *Molecules, 25*(7), 1491. https://doi.org/10.3390/molecules25071491

Careri, M., Spagnoli, S., Panari, G., Zannoni, M., & Barbieri, G. (1996). Chemical parameters of the non-volatile fraction of ripened Parmigiano-Reggiano cheese. *International Dairy Journal, 6*(2), 147–155. https://doi.org/10.1016/0958-6946(94)00056-5

Chen, Y., MacNaughtan, W., Jones, P., Yang, Q., Williams, H., & Foster, T. (2021). Selection of potential molecular markers for cheese ripening and quality prediction by NMR spectroscopy. *LWT, 136*(2), Article 110306. https://doi.org/10.1016/j.lwt.2020.110306

Ciriello, R., Cataldi, T. R. I., Crispo, F., & Guerrieri, A. (2015). Quantification of l-lysine in cheese by a novel amperometric biosensor. *Food Chemistry, 169*, 13–19. https://doi.org/10.1016/j.foodchem.2014.07.141

Consonni, R., & Cagliani, L. R. (2008). Ripening and geographical characterization of Parmigiano Reggiano cheese by ^{1}H NMR spectroscopy. *Talanta, 76*(1), 200–205. https://doi.org/10.1016/j.talanta.2008.02.022

Consorzio Tutela Grana Padano. (2017). Ageing - Grana Padano. Retrieved from https://www.granapadano.it/en-en/the-art-of-production/an-iconic-product/ageing/ Accessed September 2023.

Debik, J., Sangermani, M., Wang, F., Madssen, T. S., & Giskeødegård, G. F. (2022). Multivariate analysis of NMR-based metabolomic data. *NMR in Biomedicine, 35*(2), Article e4638. https://doi.org/10.1002/nbm.4638

Eugster, E., Fuchsmann, P., Schlichtherle-Cerny, H., Bütikofer, U., & Irmler, S. (2019). Formation of alanine, α-aminobutyrate, acetate, and 2-butanol during cheese ripening by Pediococcus acidilactici FAM18098. *International Dairy Journal, 96*, 21–28. https://doi.org/10.1016/j.idairyj.2019.04.001

Farkye, N. Y. (2004). Cheese technology. *International Journal of Dairy Technology, 57*(2–3), 91–98. https://doi.org/10.1111/j.1471-0307.2004.00146.x

Ferranti, P., Barone, F., Chianese, L., Addeo, F., Scaloni, A., Pellegrino, L., & Resmini, P. (1997). Phosphopeptides from Grana Padano cheese: Nature, origin and changes during ripening. *The Journal of Dairy Research, 64*(4), 601–615. https://doi.org/10.1017/s0022029997002392

Gatti, M., Bottari, B., Lazzi, C., Neviani, E., & Mucchetti, G. (2014). Invited review: Microbial evolution in raw-milk, long-ripened cheeses produced using undefined natural whey starters. *Journal of Dairy Science, 97*(2), 573–591. https://doi.org/10.3168/jds.2013-7187

Gésan-Guiziou, G. (2010). 15 - removal of bacteria, spores and somatic cells from milk by centrifugation and microfiltration techniques. In M. W. Griffiths (Ed.), *Improving the safety and quality of Milk* (pp. 349–372). Woodhead Publishing.

Hong, R. S., Hwang, K. H., Kim, S., Cho, H. E., Lee, H. J., Hong, J. T., & Moon, D. C. (2013). Survey of ERETIC2 NMR for quantification. *Journal of the Korean Magnetic Resonance Society, 17*(2), 98–104. https://doi.org/10.6564/jkmrs.2013.17.2.098

Kandasamy, S., Yoo, J., Yun, J., Kang, H. B., Seol, K.-H., & Ham, J.-S. (2020). ^{1}H HRMAS-NMR based metabolic fingerprints for discrimination of cheeses based on sensory qualities. *Saudi Journal of Biological Sciences, 27*(6), 1446–1461. https://doi.org/10.1016/j.sjbs.2020.04.043

Kucheryavskiy, S. (2020). Mdatools – R package for chemometrics. *Chemometrics and Intelligent Laboratory Systems, 198*(103937), Article 103937. https://doi.org/10.1016/j.chemolab.2020.103937

Kuhn, M., & Wickham, H. (2020). Tidymodels: A collection of packages for modeling and machine learning using tidyverse principles. Boston, MA, USA. Available online: https://www.tidymodels.org/.

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software, 25*, 1–18. https://doi.org/10.18637/jss.v025.i01

Longobardi, F., Casiello, G., Ventrella, A., Mazzilli, V., Nardelli, A., Sacco, D., Catucci, L., & Agostiano, A. (2015). Electronic nose and isotope ratio mass spectrometry in combination with chemometrics for the characterization of the geographical origin of Italian sweet cherries. *Food Chemistry, 170*. https://doi.org/10.1016/j.foodchem.2014.08.057

Malacarne, M., Summer, A., Franceschi, P., Formaggioni, P., Pecorari, M., Panari, G., & Mariani, P. (2009). Free fatty acid profile of Parmigiano–Reggiano cheese throughout ripening: Comparison between the inner and outer regions of the wheel. International Dairy Journal, 19(10), 637–641. Doi: https://doi.org/10.1016/j.idairyj.2009.04.004.

Masotti, F., Hogenboom, J. A., Rosi, V., De Noni, I., & Pellegrino, L. (2010). Proteolysis indices related to cheese ripening and typicalness in PDO Grana Padano cheese. *International Dairy Journal, 20*(5), 352–359. https://doi.org/10.1016/j.idairyj.2009.11.020

McSweeney, P. L. H. (2004). Biochemistry of cheese ripening. *International Journal of Dairy Technology, 57*(2–3), 127–144. https://doi.org/10.1111/j.1471-0307.2004.00147.x

Morandi, S., Silvetti, T., & Brasca, M. (2022). Content and spatial distribution of dairy-related Clostridium spores in Grana Padano cheese during the ripening period. *LWT, 167*, Article 113850. https://doi.org/10.1016/j.lwt.2022.113850

Mucchetti, G., Locci, F., Gatti, M., Neviani, E., Addeo, F., Dossena, A., & Marchelli, R. (2000). Pyroglutamic acid in cheese: Presence, origin, and correlation with ripening time of Grana Padano cheese. *Journal of Dairy Science, 83*(4), 659–665. https://doi.org/10.3168/jds.S0022-0302(00)74926-5

Mucchetti, G., Locci, F., Massara, P., Vitale, R., & Neviani, E. (2002). Production of pyroglutamic acid by thermophilic lactic acid bacteria in hard-cooked mini-cheeses. *Journal of Dairy Science, 85*(10), 2489–2496. https://doi.org/10.3168/jds.S0022-0302(02)74331-2

Mulas, G., Anedda, R., Longo, D. L., Roggio, T., & Uzzau, S. (2016). An MRI method for monitoring the ripening of Grana Padano cheese. *International Dairy Journal, 52*, 19–25. https://doi.org/10.1016/j.idairyj.2015.08.011

Niro, S., Succi, M., Tremonte, P., Sorrentino, E., Coppola, R., Panfili, G., & Fratianni, A. (2017). Evolution of free amino acids during ripening of Caciocavallo cheeses made with different milks. *Journal of Dairy Science, 100*(12), 9521–9531. https://doi.org/10.3168/jds.2017-13308

Piras, C., Cesare Marincola, F., Savorani, F., Engelsen, S. B., Cosentino, S., Viale, S., & Pisano, M. B. (2013). A NMR metabolomics study of the ripening process of the Fiore Sardo cheese produced with autochthonous adjunct cultures. *Food Chemistry, 141*(3), 2137–2147. https://doi.org/10.1016/j.foodchem.2013.04.108

Pogačić, T., Mancini, A., Santarelli, M., Bottari, B., Lazzi, C., Neviani, E., & Gatti, M. (2013). Diversity and dynamic of lactic acid bacteria strains during aging of a long ripened hard cheese produced from raw milk and undefined natural starter. *Food Microbiology, 36*(2), 207–215. https://doi.org/10.1016/j.fm.2013.05.009

Prandini, A., Sigolo, S., Cerioli, C., & Piva, G. (2009). Survey on conjugated linoleic acid (CLA) content and fatty acid composition of Grana Padano cheese produced in different seasons and areas. *Italian Journal of Animal Science, 8*(4), 531–540. https://doi.org/10.4081/ijas.2009.531

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available online: https://www.R-project.org/.

Ralli, E., & Spyros, A. (2023). A study of Greek Graviera cheese by NMR-based metabolomics. *Molecules, 28*(14), 5488. https://doi.org/10.3390/molecules28145488

Ramadan, A., Kamel, A., Taha, A., El-Shabrawy, A., & Abdel-Fatah, N. A. (2020). A multivariate data analysis approach for investigating daily statistics of countries affected with COVID-19 pandemic. *Heliyon, 6*(11), Article e05575. https://doi.org/10.1016/j.heliyon.2020.e05575

Ribeiro, J. C., júnior, Peruzi, G. A. S., Bruzaroski, S. R., Tamanini, R., Lobo, C. M. O., Alexandrino, B., … Beloti, V. (2019). Short communication: Effect of bactofugation of raw milk on counts and microbial diversity of psychrotrophs. *Journal of Dairy Science, 102*(9), 7794–7799. https://doi.org/10.3168/jds.2018-16148

Ricci, M., Gasperi, F., Endrizzi, I., Menghi, L., Cliceri, D., Franceschi, P., & Aprea, E. (2022). Effect of dairy, season, and sampling position on physical properties of Trentingrana cheese: Application of an LMM-ASCA model. *Foods (Basel, Switzerland), 11*(1), 127. https://doi.org/10.3390/foods11010127

Robinson, D, Hayes, A, & Couch, S. broom: Convert Statistical Objects into Tidy Tibbles. R package version 1.0.6. https://github.com/tidymodels/broom. https://broom.tidymodels.org/.

Rocchetti, G., Lucini, L., Gallo, A., Masoero, F., Trevisan, M., & Giuberti, G. (2018). Untargeted metabolomics reveals differences in chemical fingerprints between PDO and non-PDO Grana Padano cheeses. *Food Research International, 113*, 407–413. https://doi.org/10.1016/j.foodres.2018.07.029

Scano, P., Cusano, E., Caboni, P., & Consonni, R. (2019). NMR metabolite profiles of dairy: A review. *International Dairy Journal, 90*, 56–67. https://doi.org/10.1016/j.idairyj.2018.11.004

The PDO. (2023, January 23). Grana Padano. https://www.granapadano.it/en-en/the-art-of-production/the-pdo/. (Accessed October 2023).

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of the Open Source Software, 4*, 1686. https://doi.org/10.21105/joss.01686

Wishart, D. S., Guo, A., Oler, E., Wang, F., Anjum, A., Peters, H., … Gautam, V. (2022). HMDB 5.0: The human metabolome database for 2022. *Nucleic Acids Research, 50*(D1), D622–D631. https://doi.org/10.1093/nar/gkab1062