applied sciences

MDPI

Article

Non-Destructive Characterization of Italian Local *Brassicaceae* Cultivars Using ATR-FT-IR and Chemometrics

Luciano Di Martino [1], Alessandra Biancolillo [2,*], Claudia Scappaticci [2], Martina Foschi [2] and Angelo Antonio D'Archivio [2]

[1] Majella Seed Bank-Parco Nazionale della Majella, Via Badia 28, 67039 Sulmona, Italy; luciano.dimartino@parcomajella.it
[2] Dipartimento di Scienze Fisiche e Chimiche, Università degli Studi dell'Aquila, Via Vetoio, 67100 L'Aquila, Italy; claudia.scappaticci@graduate.univaq.it (C.S.); martina.foschi@univaq.it (M.F.); angeloantonio.darchivio@univaq.it (A.A.D.)
* Correspondence: alessandra.biancolillo@univaq.it

Abstract: Brassicaceae is a family of vegetables found all over the world that has been attracting the attention of researchers due to its rich chemical composition and potential health benefits (antioxidant and anti-inflammatory, as well as antimutagenic activity and potential anticarcinogenic effects). In Italy, various Brassicaceae varieties are commercially available, including traditional local cultivars, which have unique features and genetic diversity. As a result, there is a growing need to protect and recognize these landraces to preserve biodiversity. In this study, non-destructive tools such as Attenuated Total Reflectance-Fourier Transform-Infrared Spectroscopy (ATR-FT-IR) and chemometrics were employed to investigate eight distinct Brassicaceae landraces. The collected data were analyzed using a class modeling approach (Soft Independent Modeling of Class Analogy) and a discriminant classification method (Partial Least Squares Discriminant Analysis) to assess similarities and dissimilarities among the samples, all cultivated in an experimental field under the same pedoclimatic conditions. Remarkably, the combination of IR spectra and chemometric tools allowed accurate classification of the samples according only to their genetic background and despite their inclination to hybridization. The study highlights and demonstrates the importance and applicability of this specific non-destructive method for assisting the management and preservation of the genetic resources related to the local varieties of Brassicaceae.

Keywords: brassicaceae; spectroscopy; ATR-FT-IR; chemometrics; classification; SIMCA; PLS-DA

Citation: Di Martino, L.; Biancolillo, A.; Scappaticci, C.; Foschi, M.; D'Archivio, A.A. Non-Destructive Characterization of Italian Local *Brassicaceae* Cultivars Using ATR-FT-IR and Chemometrics. *Appl. Sci.* **2024**, *14*, 1277. https://doi.org/10.3390/app14031277

Academic Editor: Roberto Romaniello

Received: 31 December 2023
Revised: 31 January 2024
Accepted: 1 February 2024
Published: 3 February 2024

1. Introduction

The Brassicaceae are winter vegetables grown worldwide; Italy cultivates a wide range of Brassicaceae varieties, including local variants adapted to withstand low temperatures and harsh winter conditions in mountainous regions. Some Brassicaceae cultivars are planted in association with other crops as they may help repel or disrupt the life cycle of certain harmful insects. This specific ability is often attributed to glucosinolates, which, at the moment the plant is damaged, are subject to hydrolyzation by myrosinase enzymes and transformed into bioactive isothiocyanates, volatile compounds known to have broad-spectrum antimicrobial and insecticidal activities. Thus, the cultivation of Brassicaceae has proved helpful in reducing the need for pesticides or other chemicals and is considered an ecologically friendly crop [1]. These species are of considerable interest due to their richness in beneficial substances for humans. Indeed, they are known for their antioxidant properties [2–4], thanks to the presence of vitamin C and flavonoids, and anti-inflammatory activity, for the presence of glucosinolate [5,6], which also imparts cardioprotective, antimutagenic, and anticancer properties [7–10].

Several studies can be found in the literature regarding the characterization and quantification of glucosinolates in Brassicaceae since their monitoring is essential to inform about

the availability of these compounds in the end product and to predict the potential health and nutritional benefits of this functional food. Both targeted and fingerprinting methods, such as liquid chromatography-mass spectrometry (LC-MS), gas chromatography–mass spectrometry (GC-MS), and vibrational spectroscopy, were employed for this goal [11,12]. The latter technique is particularly attractive for large-scale analysis and industry conditions, especially when combined with chemometrics, since it is relatively fast, inexpensive, non-destructive, and does not require complex sample preparation [13]. Several studies have focused on optimizing extraction processes on this remarkable matrix, which can be actually considered a functional food [14]; Baky et al. [15] characterized, using headspace solid-phase microextraction (HS-SPME) technique and gas chromatography-mass spectrometry (GC-MS), volatile and nonvolatile profiles of six Brassicaceae cultivar leaves (*Brassica oleracea* (cabbage), *Brassica oleracea* var. *italica* (broccoli), *Brassica oleracea* var. *oleracea* (cauliflower), *Brassica rapa* (turnip), *Raphanus sativus* L. (radish), and *Nasturtium officinale* (watercress)) providing a comparative study between edible and inedible parts of Brassicaceae vegetables and suggesting potential uses and valorization practices of this functional food. Lucarini and co-workers [16] evaluated, using Nuclear Magnetic Resonance (NMR) and systematic investigation, the specific effect of cultivation (organic and conventional fertilization methods) on the secondary metabolites of *Brassica oleracea* and *Brassica rapa* samples grown in the same pedoclimatic and soil conditions.

The Brassicaceae family includes more than 3700 species that commonly undergo interspecific hybridization, and the management and monitoring of this family in terms of genetic identity is a relevant topic due to the economic, ecological, and health implications. In this context, visible and near-infrared spectroscopy and chemometrics were employed to effectively discriminate Genetically Modified (GM) and non-GM *Brassica napus*, *B. rapa*, and F1 hybrids (*B. rapa*/GM *B. napus*) [17].

The interest in studying Brassicaceae also stems from the aim of preserving local and traditional varieties of Brassicaceae; in this work, we have considered typical cultivars from Abruzzo (Central Italy), known as Mugnoli of Pettorano sul Gizio (Abruzzo, Central Italy), in order to conserve its genetic biodiversity. Preserving these local varieties is important because their cultivation is not widely spread due to their ease of hybridization and gradual rather than simultaneous harvesting, which leads companies to discard them when selecting varieties to cultivate. The Mugnoli of Pettorano sul Gizio are local varieties of Brassicaceae obtained via the hybridization of cabbage, belonging to the *Brassica oleracea* L. family, with turnip, which is part of the *Brassica rapa* L. family. Therefore, it is a plant species that exhibits characteristics of both the cabbage and the turnip from which it originated [18]. Other common Abruzzo cultivars are: Guardiagrele turnip, curly kale from Lama dei Peligni, "Rapa senza testa", "Cima dell'Osento". These species cultivated in Abruzzo are the subject of our study, which is based on the characterization and classification of these local varieties to protect and enhance their value [19]. The Guardiagrele turnip is a variety of *Brassica napus* cultivated in the town of Guardiagrele (Abruzzo, Central Italy), from which it takes its name. The curly kale of Lama dei Peligni (Abruzzo, Central Italy) is a species belonging to the family *Brassicaceae oleracea* var. *sabellica*. Among the *Brassica rapa* species is "Rapa senza testa", rich in numerous leaves without a central stem. Lastly, "Cima dell'Osento" is cultivated near the Osento River in Abruzzo (Central Italy). These local varieties were analyzed and compared alongside two commercial Brassicaceae varieties: "Cima 90° San Marzano" and "Cima Grande".

These varieties were analyzed using infrared spectroscopy [20]; this technique, being rapid, non-destructive, and green, allowed us to obtain results quickly while keeping the analyzed sample intact.

Data were processed using chemometric approaches such as Soft Independent Models of Class Analogy (SIMCA) and Partial Least Squares-Discriminant Analysis (PLS-DA) to highlight similarities and differences between the considered species despite their different family origins and their ability to hybridize [21,22]. These chemometric algorithms made it possible to classify the various sample classes with good efficiency [23,24], enriching the

study conducted so far on these landraces, which are interesting due to their rich chemical composition, nutritional value, and genetic heritage [25–27].

2. Materials and Methods

2.1. Brassicaceae Samples

The study was conducted on eight out of nine Brassicacea cultivars that were grown, collected, and analyzed in an experimental field. All the cultivars are reported in Figure 1; however, Class Γ was excluded from this investigation because the cultivation was unsuccessful, and obtaining a significant number of samples was impossible. The landraces analyzed in this study were cultivated within the same experimental field in Pattorano sul Gizio (Abruzzo, Central Italy), divided into four sub-fields (Figure 2). Each sub-field was further divided into nine rows, one for each cultivated species. The order of the nine rows was then randomized for each sub-field. Growing plant species in the same experimental field allows the appreciation of different characteristics among the considered varieties without the influence of climate and soil. The cultivated and analyzed Brassicaceae varieties, along with their respective class designations, were as follows: Mugnoli broccoli A designated as "Class A", Mugnoli broccoli B designated as "Class B", Mugnoli broccoli C identified as "Class C", Guardiagrele turnip designated as "Class D", curly kale from Lama dei Peligni labeled as "Class E", Rapa senza testa designated as "Class Γ", Cima dell'Osento labeled as "Class F", Cima 90° San Marzano identified as "Class G", and finally Cima Grande designated as "Class H" [19].

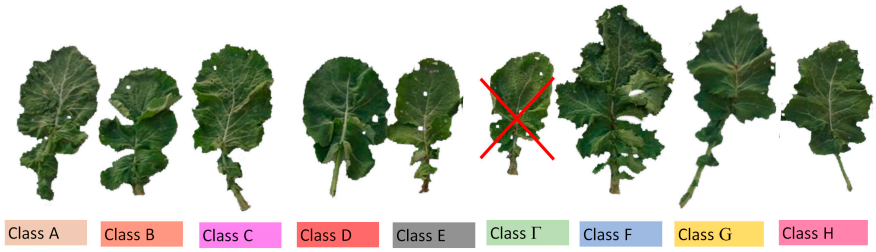

Figure 1. Pictures of the leaves of the analyzed Brassicaceae varieties. Denomination of the Class: "Class A"—Mugnoli broccoli A, "Class B"—Mugnoli broccoli B, "Class C"—Mugnoli broccoli C, "Class D"—Guardiagrele turnip, "Class E"—curly kale from Lama dei Peligni, "Class Γ"—Rapa senza testa, "Class F"—Cima dell'Osento, "Class G"—Cima 90° San Marzano, "Class H"—Cima Grande. Class Γ was excluded from further investigation because the cultivation was unsuccessful.

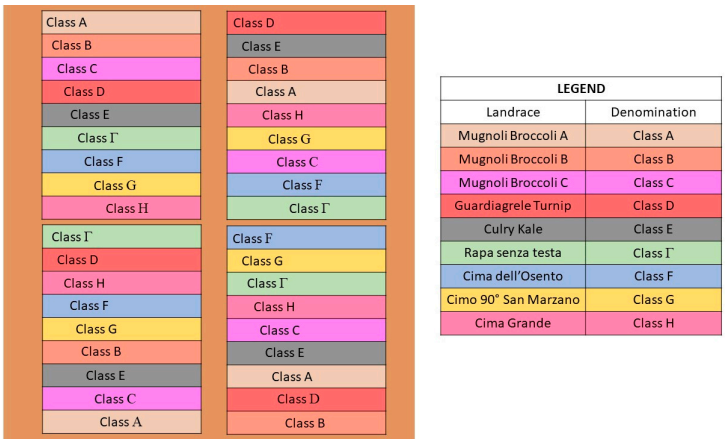

Figure 2. Representation of the cultivation of nine Brassicaceae varieties in the experimental field. The field was divided into four sub-fields, and the rows of plants were randomly arranged within each sub-field.

The conservation activity of the different Mugnoli accessions concerns the cultivation and multiplication of seeds (not commercially available) by local growers who retain the accessions by passing them from one cultivation to another each year, contributing to the preservation of this local variety.

Among the considered cultivars, curly kale from Lama dei Peligni is an Abruzzo landrace easily found commercially, but only Class G and Class H are commonly marketed in all Italian Marketplace; on the other hand, the others are all local niche varieties. All seeds were planted on 8 September 2022, harvested on 6 February 2023, stored in a refrigerator, and analyzed using infrared spectroscopy within three days of harvesting.

2.2. IR Spectra Collection

The infrared spectra were recorded on a PerkinElmer Spectrum Two™ (PerkinElmer, Waltham, MA, USA) FT-IR spectrometer, consisting of a deuterated triglycine sulfate (DTGS) detector and a PerkinElmer Universal Attenuated Total Reflectance (uATR) accessory equipped with a single bounce diamond crystal. Spectra were registered in the 4000–400 cm^{-1} spectral range with 4 cm^{-1} instrumental resolution, and sixteen scans averaged per spectral replicate. The background signal was collected with the crystal exposed to the air. ATR-FT-IR spectra were acquired after cleaning the leaves with kitchen paper to remove any soil residue. All samples were analyzed within three days of harvest, with the leaves stored in the refrigerator. The diamond was cleaned using methanol and precision wipes after each sample analysis. Care was taken to ensure complete evaporation of methanol before placing a new sample on the ATR device. A total of 60 IR spectra were acquired for Class A, 58 IR spectra for Class B, and 45 IR spectra for each of Classes C, D, E, F, G, and H.

2.3. SIMCA

Soft Independent Modeling of Class Analogies (SIMCA) is a class-modeling algorithm based on identifying similarities between samples belonging to the same category. Thus, an unknown sample will be assigned to a specific class if, when projected onto the PC model of the Class, it exhibits sufficient similarity with the samples belonging to that specific class [28–30]. SIMCA algorithm involves Principal Component Analysis (PCA) to adequately approximate the typical variability of the Class of interest and to assess the similarity [31]. Subsequently, a sample will be accepted or rejected by a class model based on its distance from the class space, which is calculated using Equation (1):

$$d = \sqrt{\left(\frac{T_i^2}{T_{0.95}^2}\right)^2 + \left(\frac{Q_i}{Q_{0.95}}\right)^2} = \sqrt{\left(T_{i,red}^2\right)^2 + \left(Q_{i,red}\right)^2} \quad (1)$$

The Mahalanobis distance of the sample from the center of the class model space is indicated by T^2, while Q represents its orthogonal distance from the model. These values are divided by $T_{0.95}^2$ and $Q_{0.95}$, which constitute the 95th percentiles of the distributions of T^2 and Q. This allows obtaining the reduced distances T_{red}^2 and Q_{red}, respectively [32]. Due to normalization, the limit values of T^2 and Q are both equal to 1; a sample will be accepted by the model if $d \leq \sqrt{2}$, otherwise, it will be rejected.

SIMCA models each category independently; as a result, a sample may be accepted by a single class, confused by being accepted by multiple classes, or excluded by all considered class models [33]. SIMCA algorithm provides values for sensitivity, specificity, and efficiency. Where sensitivity represents the percentage of samples belonging to a particular class correctly accepted by the class model, while specificity is the percentage of samples not belonging to a class that the model correctly excludes. Efficiency is the geometric mean of sensitivity and specificity.

2.4. PLS-DA

Partial Least Squares-Discriminant Analysis [34,35] is a chemometric algorithm that enables discriminant classification and is useful for analyzing highly correlated data, such as spectroscopic ones. Class membership is represented by a binary "Dummy" matrix $\boldsymbol{Y}$, with dimensions NxF, where N is the number of samples, and F is the number of classes [36]. Therefore, each object will have a value of 1 in its corresponding Class (column) of membership and a value of 0 for the other columns. The model can be mathematically expressed by Equation (2).

$$\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{E} \quad (2)$$

where $\boldsymbol{Y}$ is the dummy matrix, $\boldsymbol{X}$ is the training data matrix, and $\boldsymbol{B}$ and $\boldsymbol{E}$ are the regression coefficients and the residuals matrix, respectively.

From the resolution of Equation (1), a predicted vector $\boldsymbol{Y_{pred}}$ is obtained, which will not be binary but real-valued. Linear discriminant analysis (LDA) was applied to the predicted responses $\boldsymbol{Y_{pred}}$ [37,38] to classify the samples. In this case, according to Bayes' rule, the samples were assigned to the Class for which a higher a posteriori probability arises. The $\boldsymbol{B}$ matrix and the classification rule (obtained from the training set) were involved in classifying unknown samples of the test set, which was made up of ~30% of selected samples for each Class, using the duplex algorithm [39].

3. Results

The average spectra of the considered classes are shown in Figure 3. Although a direct comparison of intensities is not feasible on unprocessed spectra, certain differences in the intensities of specific peaks can be recognized using simple visual inspection, especially in the fingerprint zone from 1100 cm^{-1} to 500 cm^{-1}. A detailed interpretation of the reported profiles is challenging, given the complexity of the matrix and the strong influence of water on the spectroscopic signals. In general, the reported profiles reflect well the expected composition of the samples, mainly constituted by water, proteins, sugars and carbohydrates, crude fiber components, as well as phenolics, chlorophylls and porphyrins, carotenoids, Vitamins (A, C, K), glucosinolates, choline, and betaine. Starting from the broad band centered at 3360 cm^{-1}, it is primarily attributed to the O–H stretch of water, overlapping with vibrational bands associated with the O–H stretching modes of sugars, phenolics, Vitamin C and N–H stretch in proteins and indole moiety of glucosinolates [40]. Peaks from ~2920 cm^{-1} to ~2860 cm^{-1} represent symmetric and antisymmetric C–H stretching vibrations, respectively. The signal at 1740 cm^{-1} is linked to the C=O stretching of chlorophylls and pectins [41]. The broad spectral band at 1640 cm^{-1} results from several combined effects of O–H bending, H–O–H bending of water, to the Amide I band of proteins and the C=N-sulphate moiety, which result in strong IR bands between 1630 and 1690 cm^{-1} [42].

The signal at 1515 cm^{-1} is attributed to the stretching of Amide II of proteins. However, within the range of 1500–1200 cm^{-1}, absorptions are associated with CH_2 cellulose bending modes and various pectin and chlorophyll vibration modes. Signals in the 1200–950 cm^{-1} range reflect asymmetric stretching of the C–O–C glycosidic linkage and C–C and C–O ring stretching modes of polysaccharides [43]. Again, in the range of ~1495 to ~1265, signals related to S=O of glucosinolates can be found. Furthermore, signals at 1600 and 1100 cm^{-1} hint at the existence of aromatic compounds, flavonoids, and phenols.

The IR spectra collected on the various samples were subjected to different classification problems, which were solved using both SIMCA and PLS-DA.

In particular, both approaches were used to test whether it is possible to classify Class Mugnoli (constituted by all the individuals belonging to Class A, B, and C) vs. all the other investigated ecotypes (Classes D, E, F, G, and H modeled as four individual categories).

Afterward, the same classification problems were re-evaluated, considering the three accessions of Mugnoli as individual classes, in order to test whether the variability among them was high enough to require their individual modeling.

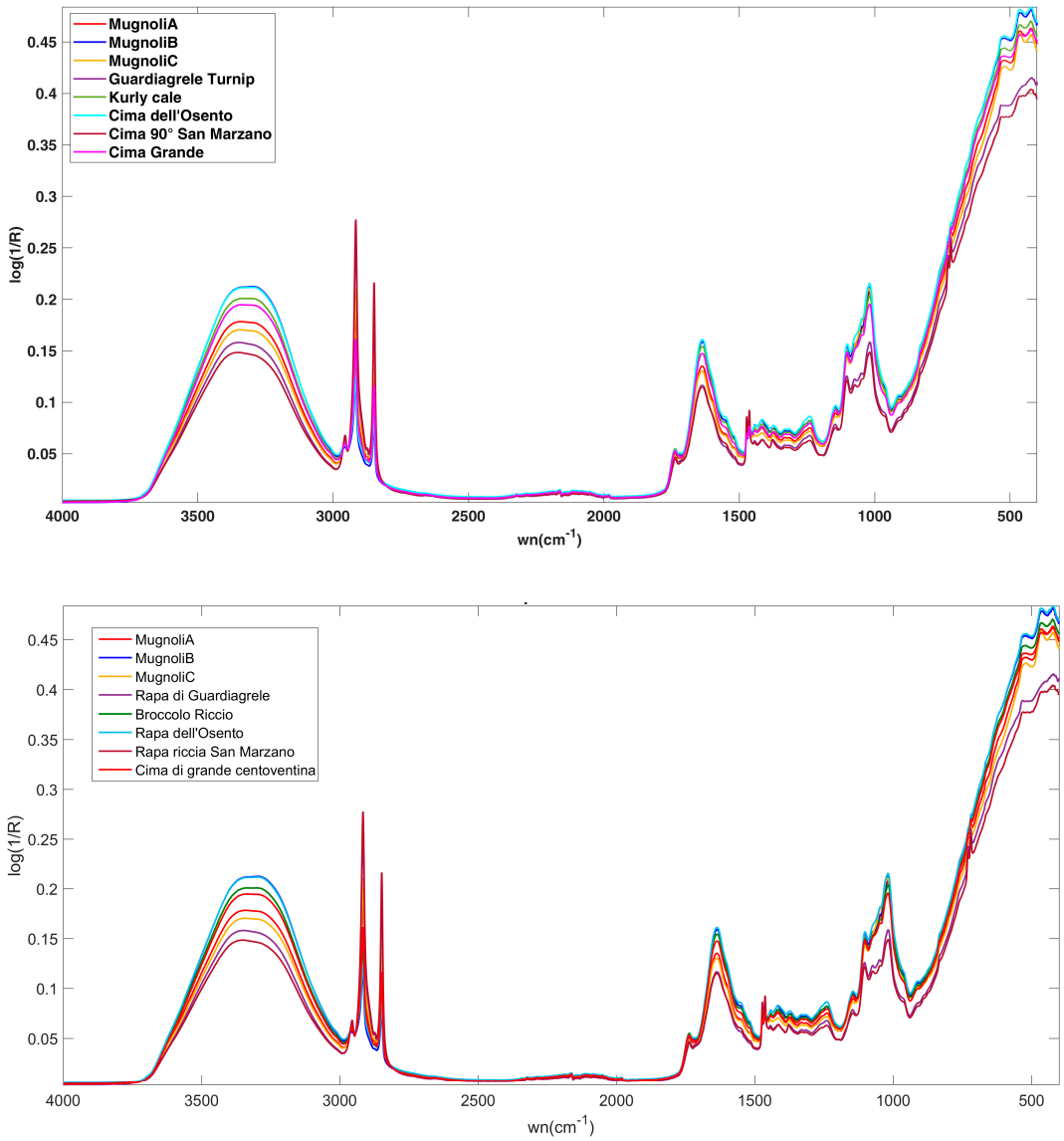

Figure 3. MIR spectra of the collected and analyzed samples, averaged according to the class belonging.

Furthermore, in order to find out significant differences between Mugnoli landrace and the Brassicaceae most commonly found on the market (Class E, G, and H), PLS-DA has been used to investigate whether it is possible to discriminate between Class Mugnoli (constituted by all the individuals belonging to Class A, B and C) vs. the three Classes of Commercial Brassicaceae, i.e., the individual classes of E, G, and H.

More details on the PLS-DA and SIMCA models are reported below in the respective paragraphs.

Regardless of the classifier used, different spectral preprocessings were tested on data:

- Mean centering (MC)
- Standard Normal Variate (SNV) (+MC)
- First Derivative (D1) (+MC)
- Second Derivative (D2) (+MC)
- SNV + D1 (+MC)
- SNV + D2 (+MC)

As it is possible to see from the list above, combinations of preprocessing methods were tested (e.g., SNV + D1), meaning that they were applied to the data in the same order

they are listed. Even if it will not always be specified, data were mean-centered prior to the creation of any model.

Consequently, six different models (for each classification problem and for each classifier) were calculated using a 7-fold cross-validation procedure. The outcome of the internal validation was then used (as it will be described below) to define the optimal data preprocessing.

3.1. Results of SIMCA Modeling

As anticipated, at first, Class Mugnoli (including samples belonging to accessions A, B, and C) has been modeled with respect to all the other investigated ecotypes. The cross-validated efficiencies (%CV_{eff}), expressed as the geometric average of sensitivity and specificity, together with the optimal number of PCs to be retained, are reported in Table 1.

Table 1. SIMCA analysis: a model of Class Mugnoli vs. all the other ecotypes. Cross-validated efficiencies (%CV_{eff}) and the optimal number of PCs retained in each model.

Preprocessing	MC	SNV	D1	D2	SNV + D1	SNV + D2
%CV_{eff}	77.3	79.4	77.1	75.3	68.5	48.5
PCs	15	14	14	14	14	14

Table 1 shows that the models built on bare mean-centered data and on data after the first derivative provide the highest comparable efficiencies. Consequently, the model built on signals after the first derivative has been considered optimal, as it requires lower PCs. This model was then used to predict the test set, providing a sensitivity of 82.0% and a specificity of 70.7%.

It is evident that the observed results are not completely satisfying. This could be attributed to possibly high variability among the three accessions, rendering the Class less well-defined and compromising the predictive capabilities of the model. Consequently, an attempt was made to model each of the three classes separately to assess whether this approach could lead to an improvement in predictive performance. The results are summarized in Table 2.

Table 2. SIMCA analysis: models of the three classes of Mugnoli (A, B, and C) vs. all the other ecotypes. Cross-validated efficiencies (%CV_{eff}) and the optimal number of PCs retained in each model.

Class	Preprocessing	MC	SNV	D1	D2	SNV + D1	SNV + D2
A	%CV_{eff}	81.77	80.99	80.57	76.16	72.37	62.61
	PCs	8	8	6	8	6	9
B	%CV_{eff}	79.26	78.12	83.28	81.95	83.8	70.44
	PCs	6	5	6	5	6	8
C	%CV_{eff}	80.68	80.35	78.56	83.82	80.25	71.58
	PCs	5	5	6	6	6	6

The results presented in Table 2 reveal that, comparing the models' outcomes pretreatment by pretreatment, an improvement when modeling the classes separately can be appreciated. This confirms the need to treat distinct accessions as separate entities. The chosen calibration models for predicting the test set were, for Class A, Class B, and Class C, those built on mean-centered data, signals preprocessed by SNV and first derivative, and spectra preprocessed by second derivative, respectively. These led to the predictive sensitivities and specificities reported in Table 3.

Table 3. SIMCA analysis: models of the three classes of Mugnoli (A, B, and C) vs. all the other ecotypes. External validation results of the optimal models expressed in terms of Sensitivity and Specificity.

Class	Preprocessing	Sensitivity	Specificity
A	Only MC	95.0	81.6
B	SNV + D1	89.5	82.7
C	D2	93.3	83.3

The outcomes reported in Table 3 are evidently superior to those obtained when modeling Mugnoli as a single class and, overall, satisfactory. This underscores the significance of considering the diverse accessions as independent entities, thereby allowing for more nuanced and effective predictive modeling.

This clearly reveals that the three accessions exhibit significant differences despite belonging to the same variety and being cultivated in the same experimental plot. In this case, the explanation could be related to the intense hybridization processes inherent to these plants.

A graphical representation of this outcome is shown in Figure 4. In Figure 4, the T^2 vs. Q plots of models of Class Mugnoli A, Class Mugnoli B, and Class Mugnoli C are shown in three subplots. In the figures, filled and empty symbols represent test and training samples, respectively. The dashed black line represents the acceptance cut-off; all the samples within this threshold are accepted by the model and, therefore, predicted as belonging to the modeled category, whereas all the others are rejected. Inspecting the plot makes it possible to appreciate that, in the case of the model of Class Mugnoli A, a great part of the samples belonging to this category fall inside the threshold, resulting in high sensitivity. At the same time, a relatively significant number of test samples (~20%) belonging to other categories, in particular Class B and Class H (Cima Grande), have been erroneously accepted by the model. In the case of the model of Class B, the test sample misclassification has mainly involved other Mugnoli (Class A) and Class F (Cima dell'Osento), which the model has unproperly accepted. Eventually, the model of Class C is the one providing the highest accuracy in prediction. In this case, the greatest part of the erroneously accepted samples belongs to Class Mugnoli A or Class Mugnoli B.

3.2. Results of PLS-DA Modeling

As for the SIMCA algorithm, PLS-DA models have been constructed using mean-centered data, data preprocessed using SNV, first derivative (D1), second derivative (D2), and combinations of the preprocessings (SNV + D1 and SNV + D2), all mean centered before constructing the models. Calibration models were optimized using cross-validation and by selecting optimal parameters such as the best data preprocessing and the optimal number of latent variables (LVs) based on the cross-validated correct classification rate (%CCRcv) for the Class or Classes of interest (Mugnoli A, B, C).

Initially, the Mugnoli Class ABC was obtained by fusing the three Classes of Mugnoli accessions (Class A, Class B, and Class C) and was modeled against all other and separated ecotypes (Class D, E, F, G, and H). Table 1 shows the optimal number of LVs and the cross-validation correct classification rate (%CCRcv) of the Mugnoli class (Class ABC) for each pretreatment.

From Table 4, it can be deduced that the best model is the one obtained by preprocessing the data with the second derivative.

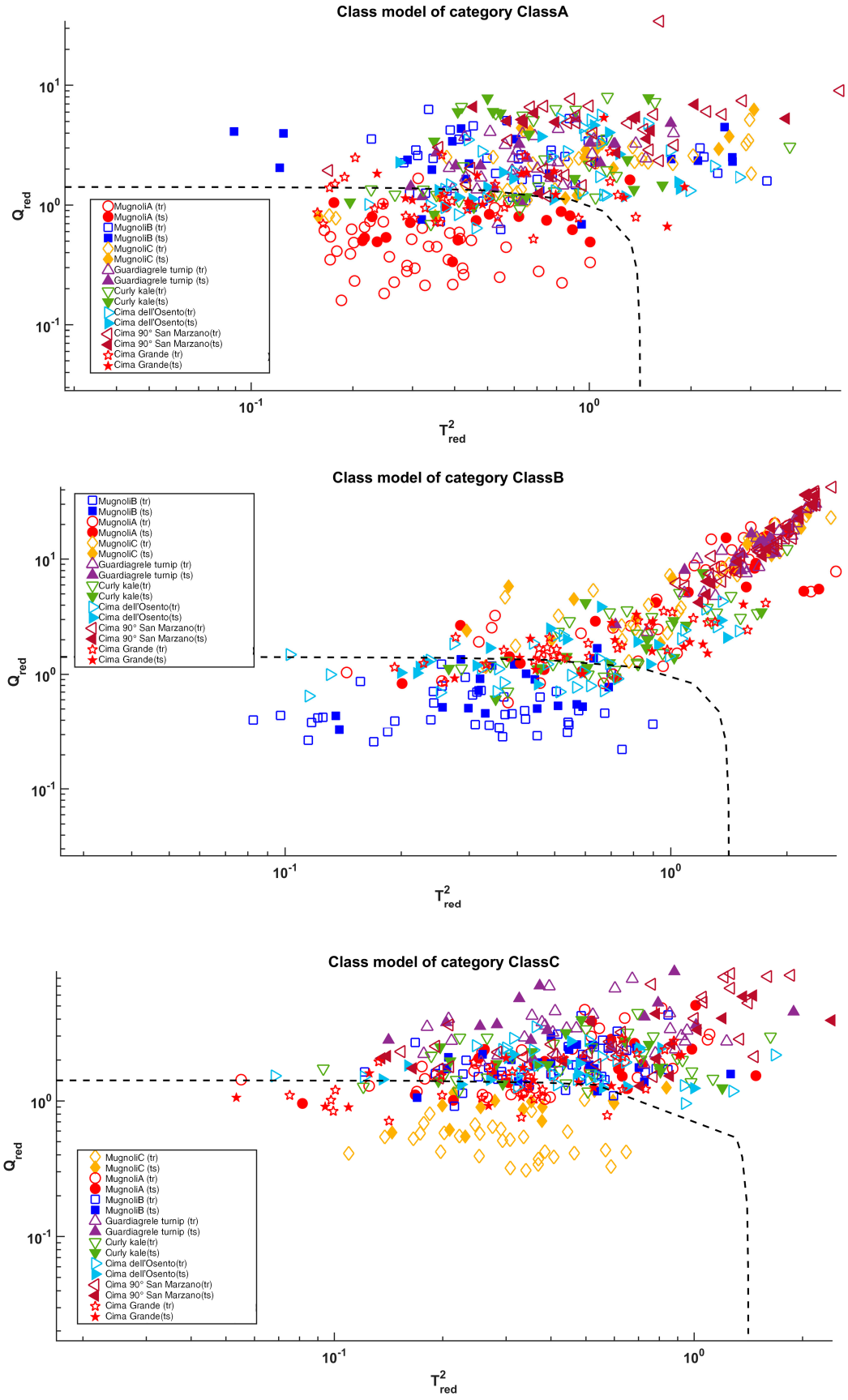

Figure 4. SIMCA analysis: SIMCA results for the three Mugnoli classes as a projection onto T^2_{red} vs. Q_{red} space of the training (full symbols) and test (empty symbols) samples.

Table 4. The table shows the correct cross-validation classification rate (%CCRcv) for the Class of Mugnoli (Class ABC) against the other ecotypes, relative pretreatment, and the optimal number of latent variables (LVs).

Preprocessing	MC	SNV	D1	D2	SNV + D1	SNV + D2
%CCRCV	76.99	82.30	81.42	89.38	71.68	86.73
LVs	16	17	17	15	13	17

Table 5 shows the corresponding prediction percentage values (%CCRpred) obtained by applying the optimal model to the test set samples.

Table 5. The table shows the results obtained in prediction on the test set by solving the classification problem, where the Mugnoli class (Class ABC) was classified against all other ecotypes.

Pretreatment	LVs	Class ABC % CCRpred	Class D % CCRpred	Class E % CCRpred	Class F % CCRpred	Class G % CCRpred	Class H % CCRpred
D2	15	84.00	100.00	73.33	80.00	93.33	73.33

As can be seen in Table 4, the model built on preprocessed data using the second derivative is capable of achieving, for our target class (Class ABC), a % CCRpred value of 84.00%

Then, the three Mugnoli classes, A, B, and C, were considered as three distinct classes and classified against all other and separated ecotypes (Class D, E, F, G, and H) in an attempt to improve predictive capabilities. Table 6 reports the values of the cross-validated correct classification rate (%CCRcv) and the optimal number of LVs for the models calculated using various preprocessing methods.

Table 6. The table shows the cross-validation correct classification rate of the models built by operating various pretreatments.

Preprocessing	LVs	Class A % CCRcv	Class B % CCRcv	Class C % CCRcv	Class D % CCRcv	Class E % CCRcv	Class F % CCRcv	Class G % CCRcv	Class H % CCRcv	Mean % CCRcv
MC	17	72.50	92.31	83.33	96.67	80.00	90.00	90.00	89.66	86.81
SNV	15	75.00	94.87	76.67	93.33	83.33	80.00	83.33	79.31	83.23
D1	15	72.50	97.44	86.67	96.67	73.33	90.00	90.00	86.21	86.60
D2	18	90.00	89.74	90.00	96.67	83.33	93.33	96.67	86.21	90.74
SNV + D1	15	82.50	84.62	83.33	96.67	86.67	90.00	96.67	86.21	88.33
SNV + D2	16	77.50	84.62	86.67	96.67	80.00	80.00	93.33	79.31	84.76

From Table 6, it is possible to value that the model built on data preprocessed with the second derivative provides a higher percentage of correct classification in cross-validation. For this reason, being the best model, it was used to predict the test set. In fact, as it can be observed in Table 7, where results in prediction are shown, that the model allows for the correct classification of test samples from Class A and Class B, showing a classification rate of 100.00% and a classification rate of 93.33% for Class C (one sample erroneously attributed to Class H). According to the model, a single sample of F is confused with Class C, whereas two samples of commercial Class H are attributed to Class F and a sample to Class E. Such high values of % CCRpred indicate that the spectral data are highly informative for discriminating between the species and even different Mugnoli accessions.

Table 7. Results obtained in prediction by solving the classification problem where the Mugnoli classes (Class A, Class B, and Class C) were considered separately against all the other ecotypes.

Pretreatment	LVs	Class A % CCRpred	Class B % CCRpred	Class C % CCRpred	Class D % CCRpred	Class E % CCRpred	Class F % CCRpred	Class G % CCRpred	Class H % CCRpred
D2	18	100.00	100.00	93.33	100.00	93.33	93.33	100.00	80.00

In agreement with the results obtained by applying the SIMCA algorithm, the values of %CCRcv and %CCRpred are better when the three classes of Mugnoli A, B, and C are considered as three distinct classes rather than a single class (Class ABC). This confirms that the local varieties of Mugnoli in Pettorano sul Gizio (Abruzzo, Central Italy) are a hybrid species; thus, the different accessions could be influenced by other cultivated Brassicaceae in their vicinity.

Finally, a more specific classification problem was addressed to assess the possibility of distinguishing typical Mugnoli from Pettorano sul Gizio (Abruzzo, Central Italy) from the most common among other commercial Brassicaceae such as Curly Kale (E), Cima 90° San Marzano (G) and Cima Grande (H). Therefore, only the three Mugnoli classes, Class A, Class B, and Class C, as well as Classes E, G, and H, were considered.

Table 8 reports the results related to the models obtained by preprocessing the data.

Table 8. In this table, the results obtained by PLS-DA models are reported.

Preprocessing	LVs	Class A % CCRcv	Class B % CCRcv	Class C % CCRcv	Class E % CCRcv	Class G % CCRcv	Class H % CCRcv	Mean % CCRcv
MC	18	82.50	94.87	90.00	100.00	90.00	93.10	91.75
SNV	15	85.00	94.87	83.33	93.33	76.67	86.21	86.57
D1	15	85.00	94.87	96.67	96.67	93.33	93.10	93.27
D2	17	90.00	94.87	90.00	96.67	90.00	100.00	93.59
SNV + D1	15	82.50	92.31	96.67	100.00	86.67	96.55	92.45
SNV + D2	17	95.00	92.31	93.33	100.00	90.00	89.66	93.38

Here, again, the model built using the second derivative turned out to be the best; therefore, it was validated using the external test set.

In this case (Table 9), satisfactory results can be noticed by applying the second derivative as data preprocessing. The three Mugnoli classes from Pettorano sul Gizio (Abruzzo, Central Italy), Class A, Class B, and Class C, were classified with a sensitivity of 100% for the first two classes (Class A and B) and 93.33% for Class C (a single sample attributed to Class A). While the classes of the most common commercially available Brassicaceae species were classified with a percentage of correct classification rate in prediction of 100.00% for the Culy Kale class (E), 93.33% for the Cima 90° San Marzano class (G) (a single sample confuse with the Class A), and 100.00% for the Cima Grande class (H).

Table 9. In this table, the prediction results on the test samples preprocessed using the second derivative are reported.

Pretreatment	LVs	Class A % CCRpred	Class B % CCRpred	Class C % CCRpred	Class E % CCRpred	Class G % CCRpred	Class H % CCRpred
D2	17	100.00	100.00	93.33	100.00	93.33	100.00

Applying Linear Discriminant Analysis on the predicted Y obtained from the last PLS-DA model (only Classes A, B, C, E, G, and H were considered), the graphical results shown in Figures 5 and 6 can be obtained. Specifically, it can be seen that Mugnoli of Class B, which partially overlaps with Curly Kale in Figure 5, is the most different from Mugnoli classes A and C, which are instead found at positive CV1 scores such as Cima Grande and Cima 90° San Marzano.

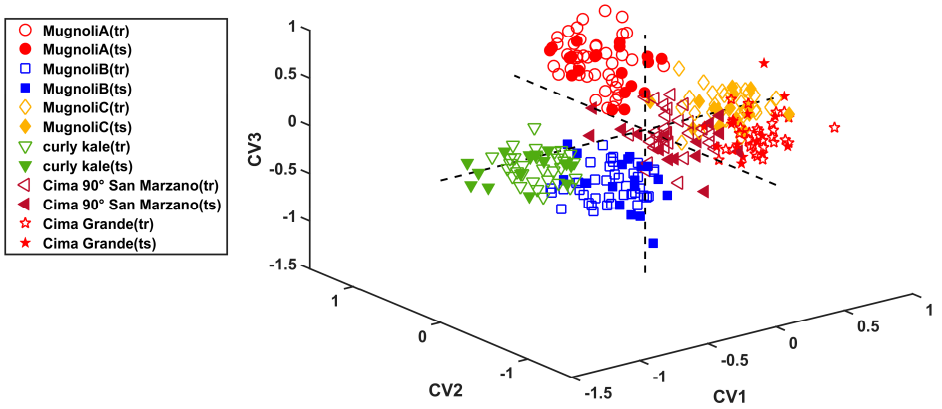

Figure 5. PLS-LDA analysis: projection of the samples on the first three Canonical Variate (CV1; CV2; CV3). Samples are distinguished according to the cultivar and in the training (full symbol) and test (empty symbols) set.

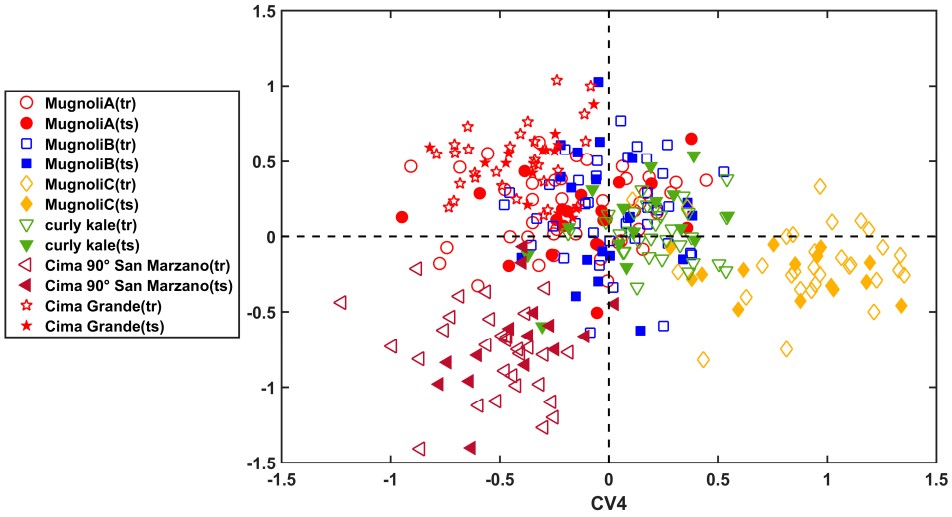

Figure 6. PLS-LDA analysis: projection of the samples on the last two Canonical Variates (CV4; CV5). Samples are distinguished according to the cultivar and in the training (full symbols) and test (empty symbols) set.

Figure 6 shows the class dispersions of Cima Grande, Cima 90° San Marzano, and Mugnoli C, which are quite distinct from each other. Meanwhile, Mugnoli A lies very close to Cima Grande and Mugnoli C, hence the few classification errors for the last considered model.

In order to identify the variables that contribute the most to the separation of the classes, VIP analysis [44] was applied to this model. Briefly, this approach allows the calculation of VIP scores for each variable and indicates its importance in the model's ability to discriminate among different classes. Since, by construction, the average of VIP values is 1, this can be set as a threshold; consequently, spectral features presenting a VIP index higher than this cut-off are those driving the classification model. A graphical representation of the outcome of the VIP analysis can be seen in Figure 7.

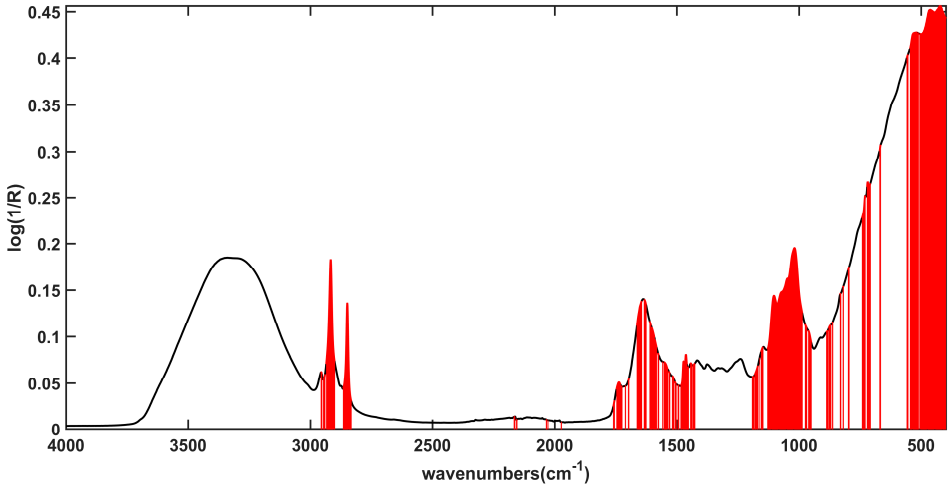

Figure 7. VIP analysis. Variables presenting a vip index > 1 are highlighted in red; the black solid line represents the average spectrum.

From this, it is clear that peaks at ~2900 and ~2850 cm^{-1}, ascribable to the symmetric and antisymmetric C–H stretching vibrations, are significant. Together with this, we can note the contribution of the range between ~1700 and ~1500, associated with the presence of chlorophylls and pectins [41], to the Amide I band of proteins and to C=N-sulphate moiety [42]. Eventually, the presence of significant variables in the range ~1200–900 indicates the relevancy of aromatic compounds, flavonoids, and phenols for the classification of the investigated classes.

4. Conclusions

In this study, an in-depth analysis of Brassicaceae, particularly focusing on local varieties from Abruzzo, Central Italy, known as Mugnoli of Pettorano sul Gizio, was conducted. The research aimed to characterize and classify three accessions of Mugnoli against local and commercial cultivars, all grown under the same pedoclimatic conditions. The design of the work allows the evaluate the genetic biodiversity by employing a non-destructive, fast, and cheap spectroscopic method aiming at assessing the possibility of using a fingerprinting technique to monitor and manage local varieties and accessions of the Brassicaceae family, which is very prone to interspecific hybridization in order to preserve the genetic heritage of this ancient and traditional varieties.

The study also explored the potential of distinguishing between Mugnoli varieties and commercially available Brassicaceae using chemometric approaches. Both SIMCA and PLS-DA models were employed, revealing that considering Mugnoli classes separately enhanced the predictive capabilities of the models. In conclusion, the study contributes valuable insights into the characterization, preservation, and classification of Brassicaceae varieties, with a specific focus on the unique Mugnoli varieties from Abruzzo. The combination of spectroscopic techniques and chemometric approaches proved effective in addressing the challenges associated with the genetic diversity and hybrid nature of Brassicaceae, providing a basis for future research and conservation efforts in the context of agricultural biodiversity and functional foods.

Author Contributions: Conceptualization, A.A.D. and L.D.M.; methodology, A.B.; software, C.S. and A.B.; validation, M.F. and C.S.; investigation, A.B. and C.S.; resources, L.D.M.; data curation, C.S. and M.F.; writing—original draft preparation, A.B. and C.S.; writing—review and editing, A.B. and M.F.; supervision, A.A.D. and L.D.M.; project administration, L.D.M.; funding acquisition, L.D.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Majella National Park within the Project "Tipizzazione di. specie vegetali endemiche, crop wild relatives e varietà agricole autoctone del Parco della Majella mediante metodi analitici ed approcci statistici multivariati" with the Dipartimento di Scienze Fisiche e Chimiche, Università degli Studi dell'Aquila.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

References

1. Flor-Weiler, L.B.; Behle, R.W.; Berhow, M.A.; McCormick, S.P.; Vaughn, S.F.; Muturi, E.J.; Hay, W.T. Bioactivity of brassica seed meals and its compounds as ecofriendly larvicides against mosquitoes. *Sci. Rep.* **2023**, *13*, 3936. [CrossRef]
2. Cai, Y.; Luo, Q.; Sun, M.; Corke, H. Antioxidant activity and phenolic compounds of 112 traditional Chinese medicinal plants associated with anticancer. *Life Sci.* **2004**, *74*, 2157–2184. [CrossRef]
3. Rahman, M.M.; Abdullah, A.T.M.; Sharif, M.; Jahan, S.; Kabir, M.A.; Motalab, M.; Khan, T.A. Relative evaluation of in-vitro antioxidant potential and phenolic constituents by HPLC-DAD of Brassica vegetables extracted in different solvents. *Heliyon* **2022**, *8*, e10838. [CrossRef]
4. Sakakibara, H.; Honda, Y.; Nakagawa, S.; Ashida, H.; Kanazawa, K. Simultaneous Determination of All Polyphenols in Vegetables, Fruits, and Teas. *J. Agric. Food Chem.* **2003**, *51*, 571–581. [CrossRef] [PubMed]
5. Jo, J.S.; Bhandari, S.R.; Kang, G.H.; Shin, Y.K.; Lee, J.G. Selection of broccoli (Brassica oleracea var. italica) on composition and content of glucosinolates and hydrolysates. *Sci. Hortic.* **2022**, *298*, 110984. [CrossRef]
6. Argentieri, M.P.; Accogli, R.; Fanizzi, F.P.; Avato, P. Glucosinolates profile of "mugnolo", a variety of Brassica oleracea L. native to southern Italy (Salento). *Planta Med.* **2011**, *77*, 287–292. [CrossRef] [PubMed]
7. Favela-González, K.M.; Hernández-Almanza, A.Y.; la Fuente-Salcido, N.M. The value of bioactive compounds of cruciferous vegetables (Brassica) as antimicrobials and antioxidants: A review. *J. Food Biochem.* **2020**, *44*, e13414. [CrossRef] [PubMed]
8. Lučić, D.; Pavlović, I.; Brkljačić, L.; Bogdanović, S.; Farkaš, V.; Cedilak, A.; Nanić, L.; Rubelj, I.; Salopek-Sondi, B. Antioxidant and Antiproliferative Activities of Kale (*Brassica oleracea* L. Var. acephala DC.) and Wild Cabbage (*Brassica incana* Ten.) Polyphenolic Extracts. *Molecules* **2023**, *28*, 1840. [CrossRef]
9. Peña, M.; Guzmán, A.; Martínez, R.; Mesas, C.; Prados, J.; Porres, J.M.; Melguizo, C. Preventive effects of Brassicaceae family for colon cancer prevention: A focus on in vitro studies. *Biomed. Pharmacother.* **2022**, *151*, 113145. [CrossRef] [PubMed]
10. Montaner, C.; Mallor, C.; Laguna, S.; Zufiaurre, R. Bioactive compounds, antioxidant activity, and mineral content of bróquil: A traditional crop of Brassica oleracea var. italica. *Front. Nutr.* **2023**, *9*, 1006012. [CrossRef]
11. Zeng, W.; Tao, H.; Li, Y.; Wang, J.; Xia, C.; Li, S.; Wang, M.; Wang, Q.; Miao, H. The flavor of Chinese kale sprouts is affected by genotypic variation of glucosinolates and their breakdown products. *Food Chem.* **2021**, *359*, 129824. [CrossRef]
12. Brindisi, L.J.; Lyu, W.; Juliani, H.R.; Wu, Q.; Tepper, B.J.; Simon, J.E. Determination of glucosinolates and breakdown products in Brassicaceae baby leafy greens using UHPLC-QTOF/MS and GC/MS. *Food Chem. Adv.* **2023**, *3*, 100389. [CrossRef]
13. Ali Redha, A.; Torquati, L.; Langston, F.; Nash, G.R.; Gidley, M.J.; Cozzolino, D. Determination of glucosinolates and isothiocyanates in glucosinolate-rich vegetables and oilseeds using infrared spectroscopy: A systematic review. *Crit. Rev. Food Sci. Nutr.* **2023**, 1–17. [CrossRef]
14. Cannavacciuolo, C.; Cerulli, A.; Dirsch, V.M.; Heiss, E.H.; Masullo, M.; Piacente, S. LC-MS- and 1H NMR-Based Metabolomics to Highlight the Impact of Extraction Solvents on Chemical Profile and Antioxidant Activity of Daikon Sprouts (*Raphanus sativus* L.). *Antioxidants* **2023**, *12*, 1542. [CrossRef]
15. Baky, M.H.; Shamma, S.N.; Xiao, J.; Farag, M.A. Comparative aroma and nutrients profiling in six edible versus nonedible cruciferous vegetables using MS based metabolomics. *Food Chem.* **2022**, *383*, 132374. [CrossRef] [PubMed]
16. Lucarini, M.; Di Cocco, M.E.; Raguso, V.; Milanetti, F.; Durazzo, A.; Lombardi-Boccia, G.; Santini, A.; Delfini, M.; Sciubba, F. NMR-Based Metabolomic Comparison of Brassica oleracea (Var. italica): Organic and Conventional Farming. *Foods* **2020**, *9*, 945. [CrossRef] [PubMed]
17. Sohn, S.-I.; Pandian, S.; Zaukuu, J.-L.Z.; Oh, Y.-J.; Park, S.-Y.; Na, C.-S.; Shin, E.-K.; Kang, H.-J.; Ryu, T.-H.; Cho, W.-S.; et al. Discrimination of Transgenic Canola (*Brassica napus* L.) and their Hybrids with B. rapa using Vis-NIR Spectroscopy and Machine Learning Methods. *Int. J. Mol. Sci.* **2021**, *23*, 220. [CrossRef] [PubMed]
18. Spoor, W.; Zohary, D.; Hopf, M. Domestication of plants in the Old World. 3rd ed. 316pp. New York: Oxford University Press. £19.95 (softback). *Ann. Bot.* **2001**, *88*, 666. [CrossRef]
19. Biancolillo, A.; Ferretti, R.; Scappaticci, C.; Foschi, M.; D'Archivio, A.A.; Di Santo, M.; Di Martino, L. Development of a Non-Destructive Tool Based on E-Eye and Agro-Morphological Descriptors for the Characterization and Classification of Different Brassicaceae Landraces. *Appl. Sci.* **2023**, *13*, 6591. [CrossRef]

20. Sohn, S.-I.; Pandian, S.; Zaukuu, J.-L.Z.; Oh, Y.-J.; Lee, Y.-H.; Shin, E.-K.; Thamilarasan, S.K.; Kang, H.-J.; Ryu, T.-H.; Cho, W.-S. Rapid discrimination of Brassica napus varieties using visible and Near-infrared (Vis-NIR) spectroscopy. *J. King Saud Univ.—Sci.* **2023**, *35*, 102495. [CrossRef]
21. Di Donato, F.; Di Cecco, V.; Torricelli, R.; D'Archivio, A.A.; Di Santo, M.; Albertini, E.; Veronesi, F.; Garramone, R.; Aversano, R.; Marcantonio, G.; et al. Discrimination of potato (*Solanum tuberosum* L.) accessions collected in majella national park (abruzzo, italy) using mid-infrared spectroscopy and chemometrics combined with morphological and molecular analysis. *Appl. Sci.* **2020**, *10*, 1630. [CrossRef]
22. Oliveri, P. Class-modelling in food analytical chemistry: Development, sampling, optimisation and validation issues—A tutorial. *Anal. Chim. Acta* **2017**, *982*, 9–19. [CrossRef] [PubMed]
23. Biancolillo, A.; Marini, F.; Ruckebusch, C.; Vitale, R. Chemometric strategies for spectroscopy-based food authentication. *Appl. Sci.* **2020**, *10*, 6544. [CrossRef]
24. Cocchi, M.; Biancolillo, A.; Marini, F. Chemometric Methods for Classification and Feature Selection. In *Data Analysis for Omic Sciences: Methods and Applications, Comprehensive Analytical Chemistry*; Jaumot, J., Bedia, C., Tauler, R., Eds.; Elsevier: Amsterdam, The Netherlands, 2018; Volume 82, pp. 265–299. ISBN 9780444640444.
25. Laghetti, G.; Martignano, F.; Falco, V.; Cifarelli, S.; Gladis, T.; Hammer, K. "Mugnoli": A Neglected Race of Brassica oleracea L. from Salento (Italy). *Genet. Resour. Crop Evol.* **2005**, *52*, 635–639. [CrossRef]
26. Palmitessa, O.D.; Gadaleta, A.; Leoni, B.; Renna, M.; Signore, A.; Paradiso, V.M.; Santamaria, P. Effects of Greenhouse vs. Growth Chamber and Different Blue-Light Percentages on the Growth Performance and Quality of Broccoli Microgreens. *Agronomy* **2022**, *12*, 1161. [CrossRef]
27. Hammer, K.; Montesano, V.; Direnzo, P.; Laghetti, G. Conservation of Crop Genetic Resources in Italy with a Focus on Vegetables and a Case Study of a Neglected Race of Brassica Oleracea. *Agriculture* **2018**, *8*, 105. [CrossRef]
28. Wold, S.; Sjöström, M. SIMCA: A Method for Analyzing Chemical Data in Terms of Similarity and Analogy. In *Chemometrics: Theory and Application*; ACS publications: Washington, DC, USA, 1977; pp. 243–282.
29. Oliveri, P.; Downey, G. Multivariate class modeling for the verification of food-authenticity claims. *TrAC Trends Anal. Chem.* **2012**, *35*, 74–86. [CrossRef]
30. De Luca, S.; Bucci, R.; Magrì, A.D.; Marini, F. Class Modeling Techniques in Chemometrics: Theory and Applications. In *Encyclopedia of Analytical Chemistry*; Wiley: Hoboken, NJ, USA, 2018; pp. 1–24.
31. Wold, S. Pattern recognition by means of disjoint principal components models. *Pattern Recognit.* **1976**, *8*, 127–139. [CrossRef]
32. Yue, H.H.; Qin, S.J. Reconstruction-based fault identification using a combined index. *Ind. Eng. Chem. Res.* **2001**, *40*, 4403–4414. [CrossRef]
33. Forina, M.; Oliveri, P.; Lanteri, S.; Casale, M. Class-modeling techniques, classic and new, for old and new problems. *Chemom. Intell. Lab. Syst.* **2008**, *93*, 132–148. [CrossRef]
34. Wold, S.; Martens, H.; Wold, H. The multivariate calibration problem in chemistry solved by the PLS method. In *Matrix Pencils, Lecture Notes in Mathematics*; Kågström, B., Ruhe, A., Eds.; Springer: Berlin/Heidelberg, Germany, 1983.
35. Geladi, P.; Kowalski, B.R. Partial least-squares regression: A tutorial. *Anal. Chim. Acta* **1986**, *185*, 1–17. [CrossRef]
36. Ståhle, L.; Wold, S. Partial least squares analysis with cross-validation for the two-class problem: A Monte Carlo study. *J. Chemom.* **1987**, *1*, 185–196. [CrossRef]
37. Barker, M.; Rayens, W. Partial least squares for discrimination. *J. Chemom.* **2003**, *17*, 166–173. [CrossRef]
38. Pérez, N.F.; Ferré, J.; Boqué, R. Calculation of the reliability of classification in discriminant partial least-squares binary classification. *Chemom. Intell. Lab. Syst.* **2009**, *95*, 122–128. [CrossRef]
39. Snee, R.D. Validation of Regression Models: Methods and Examples. *Technometrics* **1977**, *19*, 415–428. [CrossRef]
40. Vo, Q.V.; Trenerry, C.; Rochfort, S.; Wadeson, J.; Leyton, C.; Hughes, A.B. Synthesis and anti-inflammatory activity of indole glucosinolates. *Bioorg. Med. Chem.* **2014**, *22*, 856–864. [CrossRef] [PubMed]
41. Reale, S.; Biancolillo, A.; Foschi, M.; Di Donato, F.; Di Censo, E.; D'Archivio, A.A. Geographical discrimination of Italian carrot (*Daucus carota* L.) varieties: A comparison between ATR FT-IR fingerprinting and HS-SPME/GC-MS volatile profiling. *Food Control* **2022**, *146*, 109508. [CrossRef]
42. Langston, F.; Redha, A.A.; Nash, G.R.; Bows, J.R.; Torquati, L.; Gidley, M.J.; Cozzolino, D. Qualitative analysis of broccoli (Brassica oleracea var. italica) glucosinolates: Investigating the use of mid-infrared spectroscopy combined with chemometrics. *J. Food Compos. Anal.* **2023**, *123*, 105532. [CrossRef]
43. Foschi, M.; Tozzi, L.; Di Donato, F.; Biancolillo, A.; D'Archivio, A.A. A Novel FTIR-Based Chemometric Solution for the Assessment of Saffron Adulteration with Non-Fresh Stigmas. *Molecules* **2022**, *28*, 33. [CrossRef]
44. Wold, S.; Johansson, E.; Cocchi, M. PLS—Partial least-squares projections to latent structures. In *3D QSAR Drug Design*; KLUWER ESCOM Science Publisher: Dordrecht, The Netherlands, 1993; pp. 523–550.