Food Chemistry

TXRF spectral information enhanced by multivariate analysis: a new strategy for food fingerprint

--Manuscript Draft--

Manuscript Number:	FOODCHEM-D-22-02736R1
Article Type:	Research Article (max 7,500 words)
Keywords:	Total-reflection X-ray fluorescence, Seed traceability, Principal component analysis, Standard normal variate, generalized least squares weighting, partial least square discriminant analysis, beans, Phaseolus vulgaris L.
Corresponding Author:	Ignazio Allegretta University of Bari Department of Soil Plant and Food Science: Universita degli Studi di Bari Aldo Moro Dipartimento di Scienze del Suolo della Pianta e degli Alimenti Bari, ITALY
First Author:	Ignazio Allegretta
Order of Authors:	Ignazio Allegretta
	Giacomo Squeo
	Concetta E. Gattullo
	Carlo Porfido
	Antonio Cicchetti
	Francesco Caponio
	Stefano Cesco
	Carlo Nicoletto
	Roberto Terzano
Abstract:	The increased costumers' request of safe and high-quality food products makes food traceability a priority for frauds identification and quality certification. Elemental profiling is one of the strategies used for food traceability, and TXRF spectroscopy is widely used in food analysis even if its potentialities have not been fully investigated. In this work, a new method for food traceability using directly TXRF spectra coupled with multivariate analyses, was tested. Twenty-four different beans' genotypes (Phaseolus vulgaris L.) grown onto two different sites have been studied. After the development of the method for beans' analysis, TXRF spectra were collected and processed with PCA combined with SNV and GLSW filter obtaining a perfect clustering of the seeds according to their geographical origin. Finally, using PLS-DA, beans were correctly classified demonstrating that TXRF spectra can be successfully used as fingerprint for food/seed traceability and that elemental quantification procedure is not necessary to this aim.

UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

DIPARTIMENTO DI SCIENZE DEL SUOLO, DELLA PIANTA E DEGLI ALIMENTI – DI.S.S.P.A.

Dear Editor,

Please find enclosed the revised version of the manuscript titled *"**TXRF spectral information enhanced by multivariate analysis: a new strategy for food fingerprint**"* by Ignazio Allegretta, Giacomo Squeo, Concetta Eliana Gattullo, Carlo Porfido, Antonio Cicchetti, Francesco Caponio, Stefano Cesco, Carlo Nicoletto and Roberto Terzano to be submitted as an original research article to ***Food Chemistry***.

We have improved our manuscript according to the reviewers' comments and suggestions. Moreover, we also answered to their questions. In this new version, two tables (Table 1 and Table 2) and four supplementary figures (Figures S1, S2, S3 and S4) were also replaced according to the reviewers' suggestions. In some cases, reviewers asked for citing further works for the improvement of the discussion. Since we reached the maximum number of cited documents (40 papers), we preferred to use the literature we had already cited to improve these points. However, in the answer to the reviewers we added further works which could be find in literature to strengthen our statements. We would like to leave to you the decision if exceeding the maximum number of references can be allowed or not.

According to the journal's guidelines, also this new version of the manuscript does not exceed 7500 words (not including references), includes three tables and three figures (additional tables and figures are submitted as supplementary materials) and cites 40 documents. Moreover, you will find a clean version of the manuscript

Hoping that you and the reviewers can consider the paper suitable for publication of Food Chemistry in the present form,

We would like to thank you for your attention.

Sincerely yours,

Ignazio Allegretta, PhD

Via Amendola, 165/A 70126 Bari (Italy)
☎ (+39) 080 5443574 • (+39) 080 5442850
E-mail: ignazio.allegretta@uniba.it
www.disspa.uniba.it
c.f. 80002170720 p. iva 01086760723

Response to Reviewers

Reviewer #2: The present contribution is focused on the application of TXRF in combination with multivariate analysis as a method for food fingerprint. Despite of the fact that the topic of research could be of general interest for readers of Food Chemistry, the manuscript cannot be published in its present form and major revision (including additional experimental work) is needed.
We would like to thank the reviewer for his/her comments about the suitability of the present manuscript for Food Chemistry. We also thank him/her for his/her suggestions which allowed us to improve our work and clarify some parts of the manuscript. We have answered to each of the Reviewer's questions as detailed in the following lines.

My major concerns are:
* Authors compare TXRF results with the ones obtained by digestion and ICP-OES. However, authors did not use the proper sample digestion conditions to obtain quantitative results (see Table 1). Authors should repeat the digestion (adding HF to the initial acid mixture) and also they should use additional standards for calibration purposes (in the manuscript they used only two calibration standards…).
Since the Reviewer has remarked these two points (ICP-OES calibration and digestion procedure) in other parts of the revision, in order not to be redundant, we have answered to his/her comments in other parts of the revision, which are specifically dedicated to each one of the two points.

* The authors only used two locations points. Therefore, in order to demonstrate the real capability of the TXRF + multivariate analysis approach for food fingerprint other locations should be tested…
We think that this aspect has been already clearly specified in the manuscript. The number of samples is really high (48) and they belong to 24 different genotypes. The genetic variability strongly affects the uptake of elements by the plant, and this is well-known from the literature. However, in spite of this important genetic diversity, all the 48 samples can be divided into two groups after multivariate analysis and their provenance is univocally identified. We can agree on the point that the number of sites can be increased, however we should take into account two aspects: (i) this is the first paper in which TXRF spectra (without performing any quantitative analysis) is used to trace the food origin; (ii) other works (also cited in this manuscript and published in high ranking journals) do not deal with such a high number of samples per site and use just 2-3 sites. In this last case, they use quantitative data obtained after TXRF spectra elaboration. So, the fact that TXRF can be used to identify the food origin is well known, while the possibility to obtain this result directly from the spectra without performing a quantification procedure is something new for TXRF (in particular applied to food). However, in the conclusion we added a final sentence underlining the fact that this is the first study and in the future further sites and other types of food should be taken into account to widen the applicability of the developed method (see lines 492-495 of the new version).

* I am not sure if it is a good strategy to delete the spectral region from 9 to 10.5 keV to carry out the multivariate analysis. In the case of TXRF analysis, the internal standard is not only used for quantification purposes but also to compensate, for instance, inaccuracies arising from sample deposition...perhaps a better alternative could be the normalization of the spectra considering the Ga peak…at least a discussion about this point should be included in the manuscript text.
We agree with the Reviewer's observations. In fact, the elaboration was also carried out including the region 9-10.5 keV and on the normalized spectra. The results were very similar to those already reported in the manuscript as you can see in the following figures:

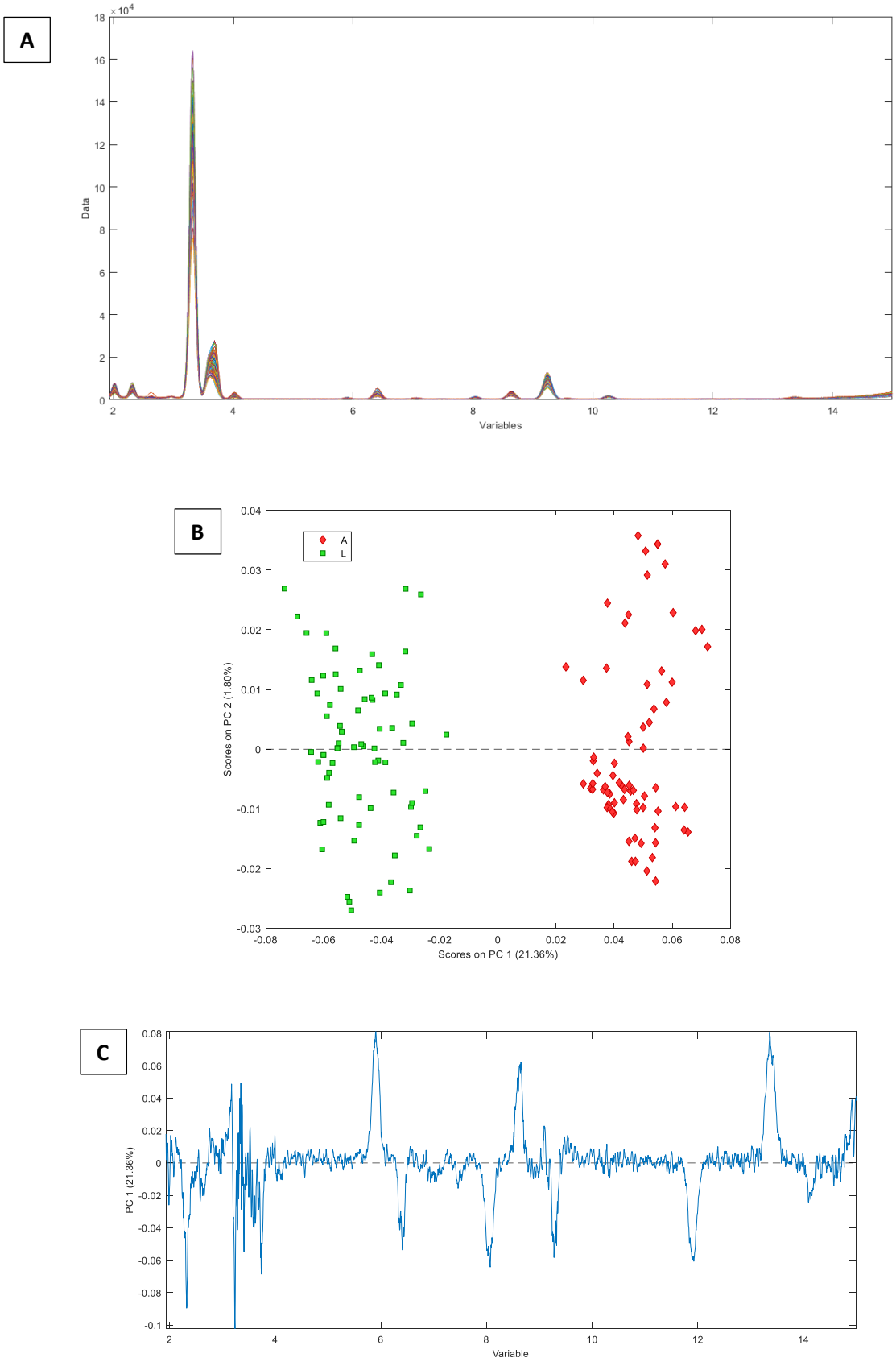

Figure 1. Elaboration carried out on the non-normalized data including the region 9-10.5 keV. A) TXRF spectra; B) PCA score plot; C) PCA loading plot.

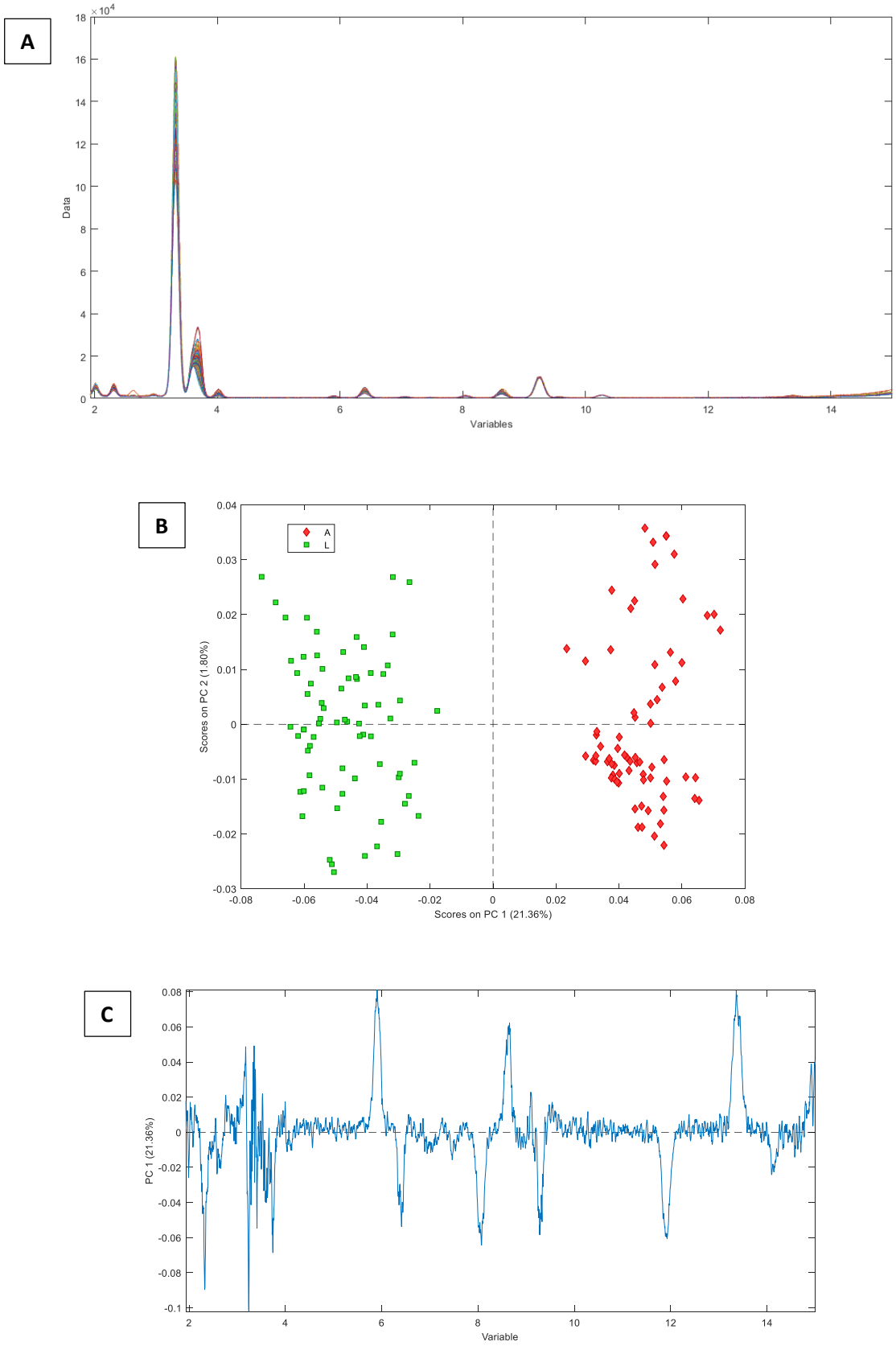

Figure 2. Elaboration carried out on the normalized data including the region 9-10.5 keV. A) TXRF spectra; B) PCA score plot; C) PCA loading plot.

From the figures, it emerges that the results are not affected by the normalization and/or by the inclusion of the Ga region. This is a worthwhile result because allow us to exclude any influence of the IS on the

discrimination results. The choice of working with the raw spectra (not normalized) and without the 9-10.5 keV region was to test the real feasibility of the TXRF signal for food authentication. From the practical point of view, this may represent a strength of the technique. Of course, this depends on the context and on the scope. In our work, we specifically aimed at skipping the quantification step and test the hypothesis that the TXRF signal could be used as such for discrimination. The fact that it works without any normalization is, in our opinion, one of the most important results presented. It's clear that, if the attention has to be put on the elemental concentration, IS addition and normalization are fundamental steps that cannot be avoided. Comments on this point are now reported in the text (lines 421-434 of the new version).

In addition, some more specific comments are described below.
COMMENTS:

o Highlights: It would be useful to include that TXRF analysis was performed after suspension preparation.

The sample preparation as suspension is not very new in the field of TXRF analysis. Considering that the maximum number of highlights is 5 and that the other points are more relevant than the sample preparation, we decided not to add this highlight to the list.

o Abstract: Please add the full name of PCA, SNV and GLSW.

Since the abstract should not exceed 150 words we think that it is not convenient to report the full name of these methods. For this reason, we decided to write the short name in the abstract and the full name in the keywords.

o Introduction:

- Line 55: AAS is not a multielemental technique. So, I guess that it is not widely used for fingerprinting…

Like ICP-OES/MS, AAS allows the analysis of a series of elements but not contemporary. For this reason, AAS is a multielemental technique but one element per time can be analysed. For elemental profiling you don't need to analyse contemporarily all the elements but you just need to quantify a series of elements and then process the quantitative data.

- Line 61: Limits of detection for XRF can be lower than 10ppm! It depends on the instrument configuration and power…

We have now adjusted the concentration range reporting the one suggested by Reviewer 4.

- Line 63: TXRF is better in terms of limits of detection than AAS and ICP-OES.

We modified the sentence according to the reviewer's suggestion.

- Lines 64-70: When authors describe the fundamentals and basic principles of TXRF they should also discuss about the requirement of sample preparation as a thin layer.

We added this information in the manuscript.

- At the end of the introduction section, authors should also highlight that in the manuscript samples are prepared as a suspension.

We added this information at the end of the introduction.

o Materials and Methods:

- Page 107: it is not necessary to include the day/month of the transplantation.

This information has been cancelled.

- Line 114: Add information about the approximate particle size. This parameter it is critical for suspension preparation.

This information is now reported in the manuscript.

- Line 114: did you find any contamination arising from the ZrO2 balls??

In the preparation of the samples for the present study we didn't observe any contamination. The use of ZrO2 balls usually can release some Hf whose L lines were not detected and Zr which cannot be detected with the present TXRF instrument

- Line 118: Why authors prepare the samples suspending 100 mg of sample in 5 mL of 1% Triton X-100? Additional information or a suitable reference should be included.

We have added our previous work as reference for this issue since it is already reported in the reference list and we reached the maximum number of references for the journal. Usually, for the analysis of organic materials, samples can be prepared as slurries dispersing 50-100 mg of sample into 2.5-5.0 mL of a dispersant agent (Triton X-100 solution, PVA solution, etc). Other types of slurry preparation can be found in the literature. This kind of sample preparation is reported also in other works. It is worth mentioning other two papers which deal with pulse flours and legume-based pasta which are similar matrixes: De Angelis et al. (2021) Antinutritional factors, mineral composition and functional properties of dry fractionated flours as influenced by the type of pulse, Heliyon 7, e06177; De Angelis et al. (2022) Dry fractionation as a promising technology to reuse the physically defected legume-based gluten-free pasta, International Journal of Food Science and Technology, Vol. 57, pag. 4816-4824.

- Line 124: authors stated "all these operations were carried out under a laminar flow hood". Also the weighting of the samples?

Only drop deposition and drying were carried out under a laminar flow hood to avoid any contamination of the reflector surface from the laboratory environment (usually, Ca, Zn, Fe and few other elements). We replaced the sentence with this new one: "Sample deposition and drying were carried out under a laminar flow hood to avoid the specimen contamination."

- Lines 132: "analysis were done in triplicate": the same sample deposited three times in different reflectors? Three different sample preparations deposited once?

We have now clarified this point (see lines 138-139).

- Line 141: Why authors did perform a filtration procedure after the digestion? It was some remaining solid particles?

Filtration (or, alternatively, centrifugation) of digested samples before ICP-OES analysis is a common procedure, as also indicated in EPA methods 3052 and 6010D. The presence of any type of particulate (also from the laboratory environment) can interfere with the sample injection or even block the injector, thus filtration is always recommended.

- Line 147: Authors stated "A three points calibration curve". In fact, they used only two calibration standards since the first point of the calibration curve is the blank…Usually a minimum of 5 calibration standards are used…

We thank the Reviewer for noticing us the issues related to the calibration procedure, which was done using two points (and not three, as erroneously indicated in the original MS). We apologize for this mistake. The text has been now modified as follows: A two points calibration curve was employed, using i) the blank acidic solution as zero point and ii) a calibration solution prepared at the concentrations of 2 mg/L (for trace elements) or 10 mg/L (for major elements).

As reported in Boss and Fredeen (1997), linear calibration curves built on two points (the blank and the high standard at a known concentration) are widely used in ICP-OES, as linearity is generally high over four to six orders of magnitude of concentration. Use of nonlinear curves in ICP-OES is negligible compared to other techniques (i.e. AAS), where the linearity is generally limited to two or three orders. Additionally, the EPA Method 6010D (Inductively coupled plasma—optical emission spectrometry) literally states that "The instrument may be calibrated using a single point standard and a calibration blank (ICB) or a multipoint calibration curve". In our experience with similar samples from the digestion

of vegetables, the use of additional calibration points does not improve the analytical performances and is time consuming. What is important is that the element concentration in the sample does not exceed that of the higher standard. Some examples of papers where the same calibration procedure was adopted are listed below.

C.B. Boss & K.J. Fredeen (1997). Concepts, Instrumentation and Techniques in Inductively Coupled Plasma Optical Emission Spectrometry. Perkin Elemer, 2nd edition. pp. 125.
I. Allegretta, C.E. Gattullo, M. Renna et al. (2019). Rapid multi-element characterization of microgreens via total-reflection X-ray fluorescence (TXRF) spectrometry. Food Chemistry 296: 86–93
R. Fourati, A. Scopa, C.B. Ahmed, et al. (2017). Leaf biochemical responses and fruit oil quality parameters in olive plants subjected to airborne metal pollution. Chemosphere 168: 514-522.
C.E. Gattullo, C. Mininni, A. Parente, et al. (2017). Effects of municipal solid waste- and sewage sludge-compost-based growing media on the yield and heavy metal content of four lettuce cultivars. Environmental Science and Pollution Research 24: 25406–25415.

- Line 153: to avoid matrix effects matrix-matched calibration standards should be used to calibrate the ICP-OES system.

We thank the Reviewer for this request of clarification. For what concerns the major elements, dilution with water was applied at the same ratio both for samples and for the solutions used for calibration (i.e. the "zero point" solution, the multistandard and the phosphorus solutions at 10 mg/L). In this way, the matrix effects are negligible since the solutions have a similar composition. This aspect has been clarified in the revised MS (lines 160-162)

Calibration solutions were prepared by diluting a multi-element calibration reference solution (Certipur® ICP Multi-element standard solution IV, 1000 mg/L, Merck GaA, Darmstadt, Germany) and an P standard solution (1000 mg/L, Carlo Erba, Milano, Italy) with the blank acidic solution, taking into account the possible dilution factor of the samples. Analysis of major elements was carried out after diluting the samples with deionized water (1:10 v/v for P, K and Ca) and, accordingly, the same water-dilution was applied for the standard solutions used for calibration (including the "zero point").

- Line 162: add information about the number of replicated (between brackets).

We added this piece of information.

- Line 170: Add more information about the first data set (number of elements x numer of samples (including 3 replicates)).

We added these pieces of information according to the Reviewer's suggestion.

- Line 172: What is the meaning of "column autoscaling"? Add more information in the manuscript text.

We added these pieces of information according to the Reviewer's suggestion.

- Line 180: "A veneitan blinds scheme was used for cross-validation". Add more information in the manuscript text.

We added these pieces of information according to the Reviewer's suggestion.

- Line 185: (144 x 4096). Add more information (i.e, 144 x 4096, samples x channels in the TXRF spectrum).

We added these pieces of information according to the Reviewer's suggestion.

o Results and Discussion:

- Line 215: why precision for Fe is so bad?
- Line 221: why recovery for Fe is so bad?

We are very sorry, an error occurred when copying the Fe concentration of the reference material measured by ICP-OES. The right value is 18.0 ± 1.3 mg kg^{-1} instead of 20.5 ± 2.4 mg kg^{-1}. As a consequence, the recovery

is lower (128% instead of 145%) and also the RSD, which is now 7% instead of 12%. Looking at the certificate of the material, Fe was determined with WDXRF and INAA using an amount of sample of 2.5 g and 250 mg, respectively. In a previous version of the certified reference material (1567a) it was suggested to use 500 mg of sample, this can be imputed to a not homogeneous composition of the reference material when a lower amount of sample is used. In TXRF, it is not possible to use such an amount of sample because in that case we should need 25 mL of Triton solution. According the Stokes' law, this results in a very high water column which causes several problems in performing the sample preparation (shaking of the suspension, collection of the sample from the column, handling of the lab equipment, contamination of the pipette because the suspension drop should be taken from the central part of the column, etc) . Usually, Fe is not uniformly distributed in wheat grains since it is more concentrated in the pericarp (Astolfi et al. 2018 J. Cereal Sci 83, 74-82). Then, it is possible that in each drop of 10 µL of suspension deposited on the reflector, a different amount of Fe is deposited. This led to a high RSD and hence bad precision for this element. We added further comments in lines 229-233 and 240-242.

- Line 232-234: if authors wanted to obtain quantitative results by ICP-OES analysis, they need to obtain a complete digestion of the samples. So the use of HF is mandatory for this type of samples. When using this acid an additional treatment with H3BO3 is required before the analysis but this is a common practice in analytical laboratories. I recommend to use a suitable digestion method to obtain ICP-OES results with enough quality for comparison purposes. They used ICP as a reference technique but they did not obtain results of enough quality to do so…

We agree with the Reviewer that the total digestion of a sample usually requires the use of HF. Nevertheless, this is not always true for organic samples, including plant samples, and in general for samples with less than 10% of SiO2. The EPA method 3052, developed for the microwave assisted acid digestion of siliceous and organically based matrices, reports that HF is not mandatory for plant samples. Indeed, in many scientific papers the acidic digestion of plant samples is performed using HNO3 and H2O2 (Arceusz et al. 2010 Food Chemistry 120: 52–58; Allegretta et al. 2019 Food Chemistry 296: 86–93), and some of these papers are specifically related to legumes (Momen et al. 2006 Analytica Chimica Acta 565: 81-88; Herreros-Chavez et al. 2019 J Food Compos Anal 82: 103240). Some Authors also state that the use of HF is not recommended due to the longer, dangerous and burdensome procedures (including additional steps to sample preparation that can affect the analytical results), also considering that the acid leaching of plant samples with HNO3 gives similar results compared with the total digestion using HF (Sastre et al. 2002 Analytica Chimica Acta 462:59–72; Tam and Yao 1999 Bull. Environ. Contam. Toxicol. 62: 708). As for calcium quantification, 80%-recovery can be considered acceptable since it falls within the recovery range required for any method validation (80-120%).

- Line 240: please, add also information regarding the Ni concentration in the sample digest.

Ni was not detected with ICP-OES. This is why we reported in the table "<LOD". It is not possible to determine its concentration with ICP-OES if we cannot even detect it.

- To compare TXRF and ICP-OES results a statistical test should be applied.

We can agree with the reviewer if the purpose was to certify that both techniques gave the same results. In this case, a t-Student test should be performed. However, as we reported in the manuscript, a reference material for beans is not commercially available and we had to use another reference material whose matrix is not exactly the same since it is wheat flour (NIST 1567b). On the contrary, our purpose was to demonstrate that the performances obtained with TXRF and ICP-OES on the reference material are similar to those obtained on the beans and in order to demonstrate this, the ratio between the estimated concentrations obtained with both methods (TXRF/ICP-OES) should be considered. In fact, if the ratios TXRF/ICP-OES obtained on the reference material and the beans is similar, it can be concluded that the methods behave in the same way on the reference material (wheat flour) and on the beans.

- Line 252: the reported ranges include both locations?

Yes, we have now added a short sentence to make this point more clear (line 276).

- Lines 282-283: this is not true for Fe… why?

As already stated at lines 287-292 of the original MS, the available fraction of a nutrient in soil is only one of the aspects influencing the plant uptake of the micronutrient. Indeed, many other physical-chemical properties of the soil, as well as the plant nutritional status, the plant physiology, the presence of microbial symbiosis in the rhizosphere, etc., might have modulated the Fe uptake by bean plants.

o References: Revise this section there are some typing errors (i.e., line 533 "Ramírez" instead of "Ramìrez", line 560 "Jablan" instead of "Jablean").

We corrected these mistakes.

o Figures:

- Figure 1: What about Br concentrations in Asiago samples?

Bromine is not quantified in all the samples since the quantified concentration is lower than LOQ. For this reason it wasn't reported in the figure for all the beans from Asiago. We have now written a sentence in the manuscript to underline this aspect (line 288).

- Figure 2: the first 2PCs only accounted for the 48.78% of the total variance. Why authors did not use at least 3PCs which is the improvement of the explained variance?

Please consider that, as reported in the original manuscript (lines 272-273), other PCs have been considered during the study. Nonetheless, no useful information to our aim were found. Moreover, we would like to point out that, as reported in some excellent papers (e.g.: Bro, R., Smilde, A.K. (2014). Principal component analysis. Analytical Methods, 6, 2812-2831; Kjeldahl, K., & Bro, R. (2010). Some common misunderstandings in chemometrics. Journal of Chemometrics, 24, 558-564) the stress should be put not on the amount of variability explained but on what explained variance means. Our aim was to extract useful information linked to the growing site and this information was brought by the first two PCs. Thus, to our aim, these two were enough.

- Figure 3: it would be useful to highlight the region removed from the original spectrum (9-10.5 keV).

The removed region could be seen as a flat line in the Figure. This additional information has been added in the Figure caption.

- Figure S1: Improve the quality of this figure. Is it possible to export the spectrum in an excel format, for instance? Label Si (reflector), Ga (IS) and Mo (X-ray tube).

We have replaced the old spectrum with a new one where we reported the pieces of information requested by the Reviewer. However, the quality of the figure doesn't depend on the original file but on the conversion of the figures in the pdf file.

- Figure S2 and S3: move the legend of the figure out the first figure.

We have corrected the figures

o Tables:

- Table 1: One of the reasons of the low recoveries for light elements (P-K) by TXRF can be related with absorption issues. Add a discussion in the manuscript text. The same behavior has been reported in recent manuscripts dealing with the use of suspension and TXRF analysis. Add information about the number of replicates used to estimate the SD in the footnote table. Use significant digits to report the experimental values. Why S, Cl, Se, Br and Rb values are not reported?

As for the underestimation of P, an explanation is already reported in manuscript. As for the citation, probably the reviewer refers to Maltsev et al 2021, Food Chemistry, 343, 128502. However, since the same topic was treated by Dalipi et al 2017 (even if P was not included in their discussion) we prefer to cite this paper due to the number of citations limit of the journal.

We have anyway added the information requested in the revised text (see lines 227-228). Finally, sulphur and halogens were not analysed by ICP-AES. Although quantification of these elements by ICP-AES is

possible, accuracy may be significantly affected by the loss of volatile forms of S and halogens during the sample digestion step. Losses of volatile compounds may occur even in closed-vessel digestion systems, both during digestion (from the venting valves) and at the end of digestion (when opening the vessels), especially for samples containing large amounts of organic matter, which can undergo massive and uncontrolled oxidative reactions. Rubidium is not usually quantified in plants because it's not an important element in physiological or nutritional studies. However, it is detected by TXRF and we prefer to keep it since it is important in the study of beans. We decided to delete the Se data because it wasn't found in the beans and it doesn't add any relevant information to the work.

- Table 2: Add information about the number of replicates used to estimate the SD in the footnote table. Use significant digits to report the experimental values. Instead of calculating the ratio TXRF/ICP-OES to compare both data sets, an adequate statistical test should be used.
We added the requested pieces of information and we corrected the significant digits. With regard to the statistical test, we have already addressed this point in a previous comment.

- Table S2: Add information about the experimental procedures used to estimate the exchangeable bases and available elements.
Both the determinations were done according to common analytical procedures, in particular we determined the exchangeable bases following the method described by Sumner and Miller (1996), and the available micronutrients according to Lindsay and Norvell (1978).
Sumner, M.E. and Miller, W.P. (1996) Cation Exchange Capacity. In: Sparks, D.L., Ed., Methods of Soil Analysis, Part 3: Chemical Methods, Soil Science Society of America, Madison, 1201-1230.
Lindsay, W.L., Norvell, W.A., 1978. Development of a DTPA Soil Test for Zinc, Iron, 842 Manganese, and Copper. Soil Sci. Soc. Am. J. 42, 421–428.
Also in this case, since we reached the maximum number of references we have decide to not include them in the reference list. In addition, these two methods are widely used and very well-known and since the work is not focused on the soil characteristics we think that, if we have to choose, we can avoid to report these references in the text.
Anyway, we leave to the Editor the decision if exceeding the maximum number of references can be allowed or not in this case.

Reviewer #3: Honestly, it is not clear to me whether or not this manuscript may be suitable for publication in Food Chemistry.
The manuscript is well written, the analytical methodology and the experimental design is correct. Nevertheless, the manuscript has a considerable lack of novelty. TXRF spectroscopy has been used for a long time. Chemometric analysis does not show any novelty statistic tool.
Authors should find some way to bring some novel advance to their work. Analyzing 24 genotypes of bean of two different sites does not provide any relevant merit.
In addition, introduction is quite long with many basic information, more appropriate for a book chapter than an article for a Q1 scientific journal.
We would like to thank the reviewer for her/his comment on the robustness of the method and the quality of the writing. However, we don't agree with her/his comments about the not-appropriate choice of the Journal and the novelty of the manuscript.
As for the suitability of the manuscript for Food Chemistry Journal, the "Aims and Scope" of Food Chemistry clearly state:

- ***"Food Chemistry publishes papers dealing with the advancement of the chemistry and biochemistry of foods or the analytical methods/approach used. All papers should focus on the novelty of the research carried out."***
- ***"Chemistry relating to major and minor components of food, their nutritional, physiological, sensory, flavour and microbiological aspects"***
- ***"Analytical papers related to the microbiological, sensory, nutritional, physiological, authenticity and origin aspects of food. Papers should be primarily concerned with new or novel methods (especially instrumental or rapid) provided adequate validation is described including sufficient***

data from real samples to demonstrate robustness. Papers dealing with significant improvements to existing methods, or data from application of existing methods to new foods, or commodities produced in unreported geographical areas, will also be considered."

We think that all these aspects can be found in our work. In fact, it deals with minor and major elements in beans (24 genotypes of particular regional productions). Moreover, we developed a new method using a well-established technique for the analysis of food.

As far as the novelty of the research is concerned, we didn't say that we are the first ones who applied TXRF for the analysis of food or that we are the first ones who use PCA, LDA or PLS-DA in food analysis. Literature is full of papers on these topics. But, to our knowledge, we are the first ones who have used the TXRF spectrum (not the quantified data) as a fingerprint for food traceability applying multivariate analysis and machine learning methods. This is clearly a step forward from what is already present in the literature on the use of TXRF and/or multivariate analysis. Moreover, as you stated in your comments these data were treated and presented with a correct methodology and experimental design which make this manuscript scientifically robust. The introduction that you find "more appropriate for a book chapter than an article for a Q1 scientific journal" is in our opinion useful to follow this path. In fact, we presented some of the works done with TXRF on food and some of the works where TXRF was used in combination with multivariate methods. After reading this introduction, it results clearer that we are placing a milestone after those represented by the mentioned previous works. Finally, we would like to mention that the suitability of the manuscript for Food Chemistry was pointed out by the other two reviewers and that Reviewer 4 also underlined the novelty of the approach.

Resutls are scarcely discussed. There is no reference/s of previous studies from lines 257 - 347 and 360 - 402. This would help increase the quality of the article.

We thank the reviewer for this comment which helped us to improve the quality of the manuscript. We improved the discussion comparing our results with previous studies on beans and by comparing the performances of the method we developed with previous works on food origin traceability. We couldn't include further references because we reached the maximum allowed number of cited works (40) but we used previously cited papers to improve the discussion according to the reviewer's suggestions. The new comments are at lines 289-293, 311-316, 348-351 and 421-434.

Reviewer #4: General comments

The authors have presented the newly developed method for bean analysis by using TXRF and a novel approach for determination of the geographic origin of food based on spectral analysis which avoids elemental quantification and works with spectra processed by SNV and GLSW spectra filtering. TXRF results have been verified by reference material analysis and were compared with results obtained by ICP-OES. Fingerprinting on filtered spectra data was compared with fingerprinting on quantitative data by using PCA, LDA, PCA-LDA, and PLS-DA. Both methods resulted in good clustering of samples according to the site. However, no relation was found concerning the genotype. The manuscript presents novel results in terms of method development and a completely new approach for food fingerprinting and as such is of interest to readers of Food Chemistry. I have only a few minor suggestions for authors, although the manuscript is already well written and clear.

We want to thank the reviewer for his/her suggestions and corrections. We want also to thank him/her for his/her appreciation of the work and for stressing the novel aspects of the work in terms of both results and method.

Specific comments

Line 61: You can state „(1-100 ppm …)" instead of „(10-100 ppm …).

We reported the correction suggested by the reviewer.

Line 81: I would avoid the statement „(generally in the range from 0-20 kV)" as this depends on the excitation source.

We deleted this part.

Line 104: Please add a short description of the soil analysis after mentioning Table S2. This is necessary as a considerable discussion is dedicated to the soil properties (lines 282-292).
We followed the reviewer's suggestion and we added a short description of the soil.

Line 257: Please correct „benn".
We corrected this word.

Line 261: Something is missing to finish the sentence „… Zn and Rb".
We replaced the sentence with the following one "Independently from the genotype, the elements K, Mn, Ni, Zn and Rb are more concentrated in beans grown in Asiago."

Line 273-281: Please add a table with basic statistics on analysis results for beans grown in Legnaro and Asiago. It would be easier to follow the discussion on elemental composition considering different growing sites.
We added this table in the supplementary information due to the tables/figures limits of the journal

Line 439: Please delete „can be used".
We corrected this mistake

TXRF spectral information enhanced by multivariate analysis: a new strategy for food fingerprint

Ignazio Allegretta[a]*, Giacomo Squeo[a], Concetta E. Gattullo[a], Carlo Porfido[a], Antonio Cicchetti[a], Francesco Caponio[a], Stefano Cesco[b], Carlo Nicoletto[c], Roberto Terzano[a]

[a] Department of Soil, Plant and Food Sciences, University of Bari Aldo Moro, Via G. Amendola 165/A, 70126, Bari, Italy.

[b] Faculty of Science and Technology, Free University of Bozen-Bolzano, Piazza Università 5, 39100 Bolzano, Italy.

[c] Department of Agronomy, Food, Natural resources, Animals, and Environment (DAFNAE), University of Padova, Viale dell'Università 16, 35020 Legnaro, Italy

*ignazio.allegretta@uniba.it

Abstract

The increased costumers' request of safe and high-quality food products makes food traceability a priority for frauds identification and quality certification. Elemental profiling is one of the strategies used for food traceability, and TXRF spectroscopy is widely used in food analysis even if its potentialities have not been fully investigated. In this work, a new method for food traceability using directly TXRF spectra coupled with multivariate analyses, was tested. Twenty-four different beans' genotypes (*Phaseolus vulgaris* L.) grown onto two different sites have been studied. After the development of the method for beans' analysis, TXRF spectra were collected and processed with PCA combined with SNV and GLSW filter obtaining a perfect clustering of the seeds according to their geographical origin. Finally, using PLS-DA, beans were correctly classified demonstrating that TXRF spectra can be successfully used as fingerprint for food/seed traceability and that elemental quantification procedure is not necessary to this aim.

Keywords: Total-reflection X-ray fluorescence, Seed traceability, Principal component analysis, Standard normal variate, generalized least squares weighting, partial least square discriminant analysis, beans, *Phaseolus vulgaris* L.

1 Introduction

The rediscovering of regional and local biodiversity in a globalized world together with the improved consumers' awareness about the food quality, safety, provenance and production process, pose food traceability at the center of international laws, standards and agreements (EC Regulation 178/2002; US Department of Agriculture, 2011). According to the ISO 22000:2018(en), *traceability* is the "ability to follow the history, application, movement and location of an object through specific stage(s) of production, processing and distribution" which, in other words, is "the ability to identify at any specified stage of the food chain (from the production to distribution) where the food came from (one step back) and where the food went to (one step forward)" (Espiñeira and Santaclara, 2016). This means that the control institution must be able to answer questions about product authenticity and adulterations, identify the geographical origin of the raw material(s), the transformation cycle and the production site(s), certify the quality of the food in terms of content of high-value metabolites, micro and macronutrients and the absence of potentially toxic compounds and elements. All these questions cannot be answered by using one single analytical method, but they should be addressed by several genetic, biological, chemical and biochemical techniques (Danezis et al., 2016). The most used strategies for food traceability are DNA profiling (Sica et al., 2021), molecular profiling (Binetti et al., 2017), isotope analysis (Camin et al., 2017) and elemental profiling (Benincasa et al., 2008). In all the profiling strategies, genes, molecules, stable/instable isotopes and elements are identified and quantified and then data are processed using multivariate data analysis procedures (Roberts and Cozzolino, 2016), among which principal component analysis (PCA), partial least square regression (PLSR), linear discriminant analysis (LDA), and partial least square discriminant analysis (PLS-DA) are some of the most

applied methods. In the case of elemental profiling, the quantification is performed mainly by using atomic spectroscopy, and the most used techniques are inductively coupled plasma-optical emission spectroscopy (ICP-OES) or -mass spectroscopy (ICP-MS), atomic absorption spectroscopy (AAS) and X-ray fluorescence spectroscopy (XRF). The first three methods allow to achieve very low limits of detection (reaching also sub ppb levels in the case of ICP-MS), but the dissolution of the sample matrix is mandatory to perform the analysis (with the exception of a particular configuration of AAS called graphite furnace atomic absorption spectroscopy - GFAAS). On the contrary, with XRF the analysis can be performed directly on the matrix even if detection limits are higher than in the abovementioned spectroscopies (1-100 ppm depending on the element and on the instrument configuration). However, with a particular configuration of XRF, named total-reflection X-ray fluorescence spectroscopy (TXRF), better performances can be achieved in terms of detection limits respect to AAS and ICP-OES (Allegretta et al., 2019). In TXRF, the sample is deposited onto a reflector as a thin film and a primary X-ray beam hits the reflector at an angle lower than the critical one (<1°). In these conditions, the primary beam does not refract into the reflector but it is totally reflected, causing the double excitation of the sample. This phenomenon produces a reduction in the matrix absorption effects and an increase in the signal-to-noise ratio, resulting in an improvement of the detection limits (in the range of ppm for solids and ppb for liquids) (Klockenkämper and von Bohlen, 2015). For this reason, TXRF was widely used in the analysis of food (Allegretta et al., 2019; Borgese et al., 2015; Dalipi et al., 2017). Moreover, it is considered a valuable tool for food traceability of coffee (Haswell and Walmasley, 1998), wine (Haswell and Walmasly, 1998; Obhođaš et al., 2021; Vitali Čepo et al., 2022), whisky (Shand et al., 2017), honey (Nečemer et al., 2009) tea and herbal infusions (Winkler et al., 2020). In all these studies, TXRF was used just for the element quantification and, afterwards, the results were processed with multivariate data analysis methods. By using this approach, TXRF has more or less the same applicability of the other atomic spectroscopies. Furthermore, PCA and dendrograms are usually the only applied multivariate methods on these data, and just recently a clustering method with linear discriminant

analysis (LDA) (Vitali Čepo et al., 2022) and regularized discriminant analysis (RDA) (Nečemer et al., 2009) were applied. However, respect to other atomic spectroscopies, TXRF provides a continuum spectrum as an output, which could be potentially used, in combination with chemometric approaches, like a fingerprint of the sample, as it happens in molecular analysis such as nuclear magnetic resonance (NMR) (Binetti et al., 2017), chromatography (Bevilacqua et al., 2017) and infrared (IR) spectroscopies (Abbas et al., 2020; Squeo et al., 2019). The use of XRF spectra like fingerprint is a well-established method in other research fields (Panchuk et al., 2018). On the basis of the authors' knowledge, XRF spectra and, in particular, TXRF ones, have never been used as food fingerprint and/or for traceability purposes.

The present work aims at developing, testing and validating a new method of food fingerprinting and traceability using TXRF spectra. To fulfill this goal, a group of 24 different genotypes of beans (*Phaseolus vulgaris* L.) coming from two cultivation areas were analyzed. Since there is no available procedure for the analysis of legumes (and in particular beans) with TXRF, the first step consisted in the development of a TXRF analytical method where samples were prepared as suspension., The method was validated comparing both results and performances with an official analytical technique (ICP-OES). Then, the elemental quantitative data as well as the TXRF signals were processed by unsupervised (PCA) and supervised multivariate techniques (LDA, PLS-DA), with the aim at developing a classification method for geographical origin of beans according to the growing site.

2 Materials and Methods

2.1 Beans and growing sites

Twenty-four genotypes of beans (Table S1) were sown into two different sites: one was the Experimental Farm "L. Toniolo" of the University of Padova in Legnaro (Veneto region, Italy, 7 m a.s.l.), situated at 30 km-distance from the coast; the other was a local farm in Asiago (Veneto

104 region, Italy, 994 m a.s.l.), located in a mountain region at 90 km-distance from the coast. The
105 physical-chemical characteristics of the two soils are reported in Table S2. The Asiago soil is a silty
106 clay-silty clay loam soil with a neutral pH with a ratio C/N of 15.11. The soil of Legnaro is a slight
107 alkaline silt loam soil having a lower C/N ratio than the Asiago soil. Both soils have a similar
108 elemental composition with the exception of Zn and Cl which are more concentrated in Asiago and
109 Legnaro, respectively. Moreover, in Asiago, Mn, Fe, Ni and Zn are more available than in Legnaro.
110 Seeds were preliminary sown in seedling starter trays (one seed per hole) using a peat-based geo
111 substrate (PBE Substrates, perlite grain size = 2-6 mm, pH = 5-6, EC = 0.4 mS/cm), keeping them
112 in a protect environment, providing a regular and optimal watering. After 10 days, the plantlets
113 were transplanted into the two soils and cultivated using the same agronomical practices. Pods were
114 harvested at commercial ripening. Overall, the crop cycle ended 136 and 137 days after the
115 transplanting for Legnaro and Asiago, respectively. Further specifications about the agricultural
116 practices adopted are reported in Sica et al. (2021).
117 Before the analysis, all the 48 bean samples were left drying in the oven for 48 h at 105 °C. Then,
118 the samples were powdered with a vibro-milling system (MM 400, Retsch GmbH, Haan, Germany)
119 equipped with ZrO_2 jars and balls, for 2 min at 30 Hz (mean particle size ≈ 60 μm). The powdered
120 samples were stored in a desiccator until their preparation for TXRF and ICP-OES analyses.

121

122 *2.2 TXRF analysis*

123 For TXRF analysis, 100.0 ± 5.0 mg of powder were suspended in 5 mL of 1% Triton X-100
124 solution (Sigma Aldrich CHEMIE GmbH, Steinheim, Germany) (Allegretta et al., 2019).
125 Furthermore, 10 μL of Ga (1000 mg/L, Sigma Aldrich CHEMIE GmbH, Steinheim, Germany) were
126 added as internal standard. The slurry was shaken using a vortex, then it was placed in an ultrasonic
127 bath for 15 min. After mixing again with vortex, 10 μL of the slurry were pipetted onto a siliconized
128 (SERVA Electrophoresis GmbH, Heidelberg, Germany) quartz reflector, and the sample was dried

for 10 min on a hot plate at 50°C. Sample deposition and drying were carried out under a laminar flow hood to avoid the specimen contamination..

To check the accuracy of the method, a reference material must be analyzed. However, there is no reference material of trace elements in beans or at least in legumes which can be used for this purpose. The reference material "Wheat flour" (NIST 1567b) has a matrix which is similar to the beans, and for this reason it was chosen and prepared using the same procedure applied for the bean samples. Quartz reflectors were then analyzed with an S2 Picofox TXRF Spectrometer (Bruker Nano GmbH, Berlin, Germany) equipped with a Mo source (50 kV, 0.6 mA, 30W), a multilayer monochromator and an XFlash® silicon drift detector of 20 cm^2 (FWHM < 140 eV on Mn-Kα). Three suspensions for each sample were prepared and a live time of 1000 s was selected for the TXRF analyses.

2.3 ICP-OES analysis

For ICP-OES analyses, 150 mg of powdered sample or reference material (NIST 1567b) were pre-digested overnight with 7 mL of ultrapure HNO_3 (≥69.0%, TraceSELECT®) and 1 mL of H_2O_2 (30%, TraceSELECT®), in Teflon tubes, at room temperature. Tubes were then placed in a MultiwaveGO microwave (Anton Paar, Graz, Austria) and a single step heating program (20 min to 180 °C, held for 15 min, and final cooling) was used for digestion. A blank solution (7 mL HNO_3 and 1 mL H_2O_2) underwent the same process. Finally, the mineralized samples were diluted to 25 mL with deionized water, filtered with Whatman® 42 filter paper and stored at 4 °C until analysis. Both the samples and the reference material were run in triplicate, while more than ten replications were run for the blank. Analyses were led with a Thermo iCAP 6000 series ICP-OES (Thermo Fisher Scientific Inc., Waltham, MA, USA) using the following settings: 1.15 kW RF power; 0.50 L/min auxiliary Ar flow; 0.90 L/min nebulizer Ar flow (0.5 L/min for P, K, Ca and Sr determination); 12 L/min coolant Ar flow; axial plasma configuration.

A two points calibration curve was employed, using i) the blank acidic solution as zero point, ii) a calibration solution prepared at the concentrations of 2 mg/L, or iii) a calibration solution prepared at the concentrations of 10 mg/L. Calibration solutions were prepared by appropriately diluting a multi-element calibration reference solution (Certipur® ICP Multi-element standard solution IV, 1000 mg/L, Merck GaA, Darmstadt, Germany) and an P standard solution (1000 mg/L, Carlo Erba, Milano, Italy) with the blank acidic solution. Analysis of major elements was carried out after diluting the samples with deionized water (1:10 v/v for P, K and Ca). The same dilution with deionized water was applied for the standard solutions used for calibration (including the "zero point"). The following emission wavelengths (nm) were selected for the elements' detection and quantification: 213.62 (P), 766.49 (K), 422.67 (Ca), 257.61 (Mn), 259.94 (Fe), 231.60 (Ni), 324.75 (Cu), and 213.85 (Zn). Each sample was analyzed in triplicate.

2.4 Parameters of validation

Validation of the analytical method was performed determining accuracy and precision of the analysis of the certified reference material NIST 1567b. Accuracy was assessed as the recovery with respect to the certified element's concentration, and precision was estimated as the standard deviation of the replicates (n=3). The limits of detection (LOD) and quantification (LOQ) were estimated as three and ten times, respectively, the parameter σ. The latter is calculated as the standard deviation of 10 replicates of the blank for ICP-OES, while $\sigma = C_i \frac{\sqrt{N_{BG}}}{N_i}$ for TXRF (where C_i = concentration of the element, N_i = net area of the element peak expressed in counts, N_{BG} = area of the background under the element peak expressed in counts) (Klockenkämper and von Bohlen, 2015).

177 *2.5 Multivariate data analysis*

178 Multivariate data analyses were carried out on two distinct datasets. The first dataset (144 × 12,
179 samples – including replicates – × elements) consisted of the elementary quantitative data of the
180 samples. These data were explored by Principal Component Analysis (PCA). Before, column
181 autoscaling (i.e., the mean value of the *i*th variable is subtracted from each entry of the variable and
182 then is divided by its standard deviation) was applied (Bro and Smilde, 2014). Then, Linear
183 Discriminant Analysis (LDA), PCA-LDA, and Partial Least Square Discriminant Analysis (PLS-DA)
184 were used as supervised classification methods for beans discrimination according to the growing
185 site. Before, the dataset was split in a calibration set (33 independent samples, 99 considering the
186 three replicates per each independent sample) and a validation set (15 independent samples, 45
187 considering the three replicates per each independent sample). Three different calibration and
188 validation sets were randomly created and used to evaluate the robustness of the models. The samples
189 list per each set is reported in Table S3. Each classification model was built on the calibration set,
190 evaluated by means of a cross-validation step and then tested on the corresponding real external
191 validation set. A venetian blinds scheme was used for cross-validation, with 10 data splits and a blind
192 thickness equal to 3. The optimal number of principal components (PCs) in PCA-LDA and the
193 optimal number of latent variables (LVs) for PLS-DA classification models were defined according
194 to the lowest classification error in cross-validation.

195 The second dataset consisted of the TXRF spectral dataset. First, from the original matrix (144 ×
196 4096, samples – including replicates – × channels in the TXRF spectrum) three uninformative spectral
197 regions were removed since the fluorescence signal of these regions was not connected to the sample
198 itself: i) the spectral region with E≤1.9 keV (which is characterized by a strong noise and the Si K-
199 alpha peak coming from the reflector); ii) the region with 9≤E≤10.5 keV (characterized by the
200 presence of Ga fluorescence lines); iii) the part with E≥15 keV (Rayleigh and Compton scattering
201 peaks). This procedure led to a final matrix of 144 × 2312 that was explored by PCA. Before PCA,
202 the data matrix was subjected to different signal preprocessing. At the beginning, only smoothing

(polynomial order 2, points 11) and mean centering were applied. Subsequently, also Standard Normal Variate (SNV), first derivative (1D; Savitzky-Golay, polynomial order 2, points 11) and Generalized Least Squares Weighting (GLSW; clutter source: x-block classes, $\alpha = 0.02$) were tested separately or in combination. After exploration, PLS-DA was used as supervised classification method for beans discrimination according to the growing site. The same sample subsets (Table S3) were used for the training and the testing of the models' performance. The optimal number of LVs for PLS-DA classification model was defined according to the lowest classification error in cross-validation. A venetian blinds scheme was used for cross-validation, with 10 data splits and a blind thickness equal to 3. For all the supervised classification analyses carried out on both the datasets, sensitivity, specificity, and classification error were used as figures of merit to evaluate the performance of the models (Ballabio et al., 2018). Per each class, sensitivity indicates the true positive rate, *i.e.*, the number of samples belonging to P-class that are correctly classified as P; specificity indicates the true negative rate, *i.e.*, the number of samples from class N correctly rejected from P-class; the overall classification error is calculated as 1 – mean of the sensitivities for the two classes. Multivariate data analyses were carried out in Matlab environment (The MathWorks, Inc. MA, USA) using original codes, the PLS_toolbox (Eigenvector Research Inc., USA) and the classification toolbox (Ballabio et al., 2013).

3 Results and Discussion

3.1 TXRF analysis of beans: method development, validation and comparison with ICP-OES

To assess the performance of TXRF and validate the method, analyses were led on the NIST 1567b reference material. The TXRF spectrum revealed the presence of P, S, Cl, K, Ca, Mn, Fe, Cu, Zn, Se, Br and Rb whose concentrations are reported in Table 1. For TXRF results, the recovery on all the elements ranged between 80% (P) and 116% (Ca) demonstrating a good accuracy of the method. The lower accuracy for P can be imputed to absorption effects which occur in particular

when samples are prepared as slurries (Dalipi et al., 2017). The precision was also good for all the elements (RSD<10%) except for Fe that reached the 29%. This could be related to a non-perfect homogeneity of the standard material a therefore to the different amount of Fe which is deposited onto the reflector. In fact, according to the reference material's certificate, a mass of 500 mg should be used to obtain a uniform sample for Fe determinations. However, this sample mass is too high for the preparation of suspensions for TXRF analysis. Moreover, the mass deposited onto each reflector is much lower than the suggested one and this can result in a low precision for Fe. The LODs and LOQs decreased with the increase of the atomic number ranging from 24 and 80 mg/kg for P to 0.1 and 0.5 for Br, respectively. Rubidium was detected but its concentration in the reference material was too close to the LOQ and, therefore, its quantification was not accurate nor precise. Nickel was neither certified nor detected and its LOD and LOQ were 0.2 and 0.6 mg/kg, respectively.

In the case of ICP-OES results, the recovery ranged between 80% (Ca) to 118% (Cu) demonstrating a good accuracy, similar to TXRF, with the exception of Fe. For this latter element the recovery was 128%. In this case a better precision is achieved for Fe. LOD and LOQ were always higher than those obtained with TXRF, except for P whose LOD and LOQ were an order of magnitude lower than with TXRF.

Finally, the ratio between TXRF and ICP-OES data demonstrated a good agreement between results obtained by the two techniques. The only two exceptions were recorded in the case of Fe and Ca since their recovery was at the minimum acceptable level in the case of ICP-OES analysis (80%). The difference in Ca quantification was also observed in the analysis of other food and plant materials (Allegretta et al., 2019) and it is likely due to the not complete digestion of the sample caused by the presence of Si in the cell wall as occurs in wheat (Robberecht et al., 2008). Complete digestion of the sample using hydrofluoric acid would require a special sample nebulizer or elimination of HF by additional sample manipulation. In addition, HF is extremely dangerous to handle and therefore its use is discouraged.

Since the certified reference material used is not exactly the same material object of this study, a comparison between TXRF and ICP-OES results on the analysis of beans was carried out in order to see if the same analytical performances obtained on the NIST 1567b could be achieved when analysising beans. To study such performance, the ratio between TXRF and ICP-OES results was used and it was compared with the one obtained for the NIST 1567b. For this purpose, 5 different bean samples were analyzed with both TXRF and ICP-OES, and elements concentrations are reported in Table 2. In general, in bean samples the investigated elements were more concentrated than in NIST 1567b. This was particularly evident in the case of K whose concentration was an order of magnitude higher than in the reference material. Nickel was quantified only with TXRF, since its concentration was lower than the LOD (3.5 mg/kg) for ICP-OES. The ratio between the results obtained with TXRF and ICP-OES obtained for the beans is very similar to that obtained for the analysis of the certified reference material (Table 1). With the exception of Ni, for all the elements, TXRF/ICP-OES ranged between 0.85 and 1.03, which underlines a good agreement between the two sets of data. In the case of Ca, this ratio was 1.25 and confirmed the results obtained analyzing NIST 1567b (1.45 ratio), confirming a discrepancy between the two methods for the quantification of this element, probably deriving from the uncomplete sample dissolution, as discussed before.

3.2 Traceability of beans on the base of their elemental composition

A total of 48 bean samples harvested from the site of Asiago or Legnaro were analyzed with TXRF in triplicate. A total of 12 elements were detected in the bean samples (P, S, Cl, K, Ca, Mn, Fe, Ni, Cu, Zn, Br and Rb as shown in Figure S1) and the results are reported in Figure 1. Independently from the origin and genotype, K was the most concentrated element (12783-19586 mg kg^{-1}) followed by P (3394-6194 mg kg^{-1}), S (1607-3067 mg kg^{-1}) and Ca (447-2692 mg kg^{-1}). Manganese (9.9-25.7 mg kg^{-1}), Fe (41.8-102.7 mg kg^{-1}), Ni (0.5-4.2 mg kg^{-1}), Cu (7.5-15.2 mg kg^{-1}) and Zn (19.5-45.6 mg kg^{-1}) were found in trace levels, in agreement with the literature (Blair et al.,

2016; Cabrera et al., 2003; Ramírez-Ojeda et al., 2018). The concentrations of Br (0.4-3.1 mg kg^{-1}) and Rb (1.1-7.5 mg kg^{-1}) were also in trace levels but no data on these two elements are available in the literature for a comparison. Results did not show any capability of specific bean genotypes to accumulate a particular element(s) (Figure 1). On the contrary, an influence of the growing site on the concentration of some elements in beans was observed. Independently from the genotype, the elements K, Mn, Ni, Zn and Rb are more concentrated in the beans grown in Asiago. Halides, S, Ca and Cu seemed to be more concentrated in the beans harvested in Legnaro, but also in this case some exceptions were found (Figure 1). Only the concentration of P was not influenced by the growing site (Figure 1). Finally, Br is lower than LOD in some beans from Asiago.

This strong dependance of certain elements concentrations with the growing site is pointed out in previous works on beans and food in general. In fact, even if Blair et al. (2016) demonstrated that the accumulation of certain elements is correlated to specific DNA loci, the variability of the concentration of macro and microelements is affected by the fact that the two studied genotypes were grown in different kind of soils and conditions.

However, such univariate data analysis (Figure 1) could not provide a clear discrimination of the samples and of the effect of the considered variables (genotype and growing site) on the mineral content of beans. For this reason, data were also treated using a multivariate data analysis approach. Figure 2 reports the results of the PCA carried out on the elemental concentrations data. The first two PCs explained roughly 50% of the data variability. In the case of totally independent variables, this would mean that PC1 and PC2 wrapped up the information of around 6 of the original variables (Bro and Smilde, 2014). However, the loading plot in Figure 2A shows that several variables were well correlated, indicating that the great part of the information could be already observed in the first two PCs. In fact, the exploration of other PCs (data not shown) did not highlight useful information to our aim (*i.e.*, clusters according to the growing site and/or genotype of the beans). The loading plot shows roughly two/three variables clusters that could be separated along PC1 and PC2. Along PC1, Rb is one cluster, followed by the second made up of Mn, Ni, K, Zn, Br, Cl, P, Ca, and the last one

represented by Fe, S and Cu. Along PC2, the first cluster was made up of Ca, P, Cl, Br, Cu, Fe, S while the other one of K, Zn, Mn, Ni and Rb. In the relative score plot (Figure 2B), the samples were well clustered according to the site. Beans from Asiago were located roughly along the positive half of PC2, due to the higher content of K, Zn, Mn, Ni and Rb, while the opposite could be stated for samples from Legnaro, which were mostly characterized by higher contents of Ca, P, Cl, Br, Cu, Fe, S. Similarly to other studies (Nečemer et al., 2009; Obhođaš et al., 2021; Shand et al., 2017; Vitali Čepo et al., 2022), Cl and Br demonstrated to be very important for the identification of the food origin. These two elements cannot be quantified with analytical methods which require the mineralization of the matrix (i.e. ICP-OES) because it causes the loss of both halogens (Allegretta et al., 2019), and for this reason the use of TXRF or other XRF spectroscopies (Obhođaš et al., 2021) is mandatory for their quantification.

The concentrations of Mn, Ni and Zn measured in beans were in accordance with the values of the plant-available fraction measured in the two soils for the three elements (Table S2). In particular, a higher content of Mn, Zn and Ni was found in beans cultivated on the Asiago soil, which was characterized by higher values of available fractions of Mn, Zn and Ni compared to the Legnaro one. The higher accumulation of Cl in beans from Legnaro also seemed to be related to the higher content of this halide in the soil. Nevertheless, for other elements such as P, K, Ca, Fe and Cu, no clear relation with the soil chemical properties was observed. However, it should be considered that the process which led to the accumulation of elements in plants (and in seeds) depends on several variables such as genetic traits, soil characteristics, environmental conditions, physiological processes, plant symbiosis with soil microorganisms, etc. (Astolfi et al., 2018; Ghanbari et al., 2013; Westermann et al., 2011), and that the soil chemical properties alone cannot fully explain this clustering. Although the samples were well clustered, a high dispersion within each class was observed, likely due to the genetic variability of the beans, and no clusters were observed according to the genotype. Finally, it should be noticed that the "Fagiolo nano Valdarno" grown in Legnaro (sample 17L) clustered together with the Asiago samples. This can be related to its higher concentration of Mn and Ni and

lower concentration of S. Overall, the PCA proved that the beans' elemental composition was affected by the growing site, and therefore it was an encouraging result toward a geographic fingerprinting. To test this hypothesis, two supervised multivariate classification techniques were used, namely LDA and PLS-DA. The results of the classification are reported in Table 3. Both the models allowed to reach a very good discrimination according to the growing sites. The results were slightly influenced by the specific random subset used, which affected in some cases the figures of merit of the classification in cross-validation or prediction. Focusing on the prediction results, a prediction error of 0.16 was observed only in one case using LDA for a total of 5 erroneous class assignments. In accordance with the PCA results (Figure 2), the most discriminant variables were Mn, Zn, Rb, Ni, Ca, Fe, Cu and S. Additionally, LDA could be even coupled with PCA and, in such a case, after having calculated a PCA model, the scores were used as input in LDA. PCA-LDA modelling allowed to achieve similar or even better results than LDA with lower error in prediction, which were observed only in the case of set 3. Finally, by using PLS-DA, in two cases some misclassifications were observed bringing to a classification error of 0.09 linked, in both the cases, to an erroneous assignment of only 2 samples.

Overall, the classification results were very promising. In fact, in most of the cases, sensitivity and specificity of 100% and no classification errors were obtained. The performances achieved in classification are better than previous works done on the application of LDA on TXRF data for tracing food origin (Nečemer et al., 2009; Vitali Čepo et al., 2022). Surely, an increase in the numerosity of the dataset would be useful for testing and confirming these performances.

Nonetheless, it should be considered that this approach still requires the quantification of the elements which consists of several steps before (i.e., calibration of the instrument) and after (i.e., spectra evaluation and fitting) analysis, which slow down the outcome of the analysis. In food traceability, and in particular for the identification of frauds, the analytical process should be as fast as possible. As an alternative and to speed up the whole analytical process, multivariate statistical tools could be directly used to process the TXRF signals, as reported in the following section.

3.3 Development and validation of a fingerprint approach using TXRF spectra

The full TXRF dataset was firstly explored by PCA. The literature about the multivariate analysis of TXRF signal is scarce or totally absent. Consequently, the raw profiles were preliminarily inspected to identify proper preprocessing methods. Signal preprocessing is a fundamental step to remove irrelevant and/or disturbing signals that might interfere during subsequent data analysis (Olivieri et al., 2019). The raw signals did not show any specific systematic and/or random unwanted variation, so that only smoothing was applied in the first step and, after mean centering, PCA was performed.

The first two PCs explained 97.5% of the data variability. However, because it was expected that the major source of variability in the dataset was linked to the genotype, 5 PCs were initially calculated for the exploration purpose, accounting for a total of 99.80% of data variability. The score plots (PC1 *vs* PC2 to PC5) together with the relative loading plot are reported in Figure S2.

PC1 explained more than 90% of the data variability but, as expected, it did not highlight information linked to the bean's classes. PC2 seemed to highlight some useful information with Asiago samples that mostly laid on the negative half of the plot, while Legnaro samples roughly on the positive side. In any case, such a clustering is not so straightforward because of a clear overlapping among the classes. From the corresponding loading plot, it was figured out that the main contribution to this displacement was due to the lower intensity of the Ca K-lines (3.69 and 4.01 keV) in Asiago beans. One sample from Asiago, namely 29A (all the three replicates) felt in the positive part of PC2. The exploration of the other PCs did not point out other useful insight according to the problem under study. Overall, although samples clustering along PC2 was not so straightforward, from data exploration it seemed that useful information about the growing site could be retrieved by a multivariate approach applied to the TXRF signal. These results were compared to those obtained by applying SNV or first derivative preprocessing to the TXRF spectra (Figure S3 and S4) but even the application of these preprocessing did not clearly improve the clustering of the samples according to the growing site.

After this starting data exploration, it was clear that the relevant information was masked by other sources of variability. To face with this issue, generalized least squares weighting (GLSW) filter was applied. GLSW is a powerful orthogonal filter successfully used both in regression and classification contexts (Martens et al., 2003; Serranti et al., 2013). In short, it removes the variation in the data not related to properties of interest. In the case of classification problems, like the one in this study, GLSW removes the within-class variance as much as possible without reducing the between-class variance (Serranti et al., 2013). More details about this filter could be found elsewhere (Martens et al., 2003; Miller et al., 2010). Figure 3A reports the results of the PCA carried out on the GLSW preprocessed data matrix. The efficacy of the GLSW filter could be clearly caught from the score plot where samples clustered perfectly according to the growing site along the first PC. However, on the other hand, after applying GLSW, the loading plot became less interpretable. Consequently, it becomes a hard task to highlight the importance of the original variables to samples clustering although it could be argued that the variables between 2 and 4 keV and those after 14 keV showed higher loadings value. The effect of SNV preprocessing before GLSW was tested and the results are reported in Figure 3B. From the score plot it could be observed that samples still clustered perfectly according to the growing site although the distribution changed and a higher dispersion within each class was observed. In this sense, it seemed that SNV reduced to some extent the ability of GLSW to reduce the within class variance or, on the other hand, that some scattering was contributing to the samples clustering. However, the main advantage of using SNV is highlighted in the loading plot where a much more meaningful profile is observed. Considering that the samples clustered along PC1, only the relative loadings have been reported, and it could be observed that the variables at around 2.0, 2.3, 3.2, 3.4, 3.7, 6.4, 8.1 and 11.9 keV showed higher positive loadings on PC1 and thus characterize samples from Legnaro. Those regions correspond to the P Kα (2.01 keV), S Kα (2.31 keV), Ca Kα (3.69 keV), Fe Kα (6.40 keV), Cu Kα (8.05 keV) and Br Kα (11.92 keV). The positive peaks at 3.2 and 3.4 keV can be ascribed to several processes such as the overlapping of fluorescence lines (atmospheric Ar Kβ at 3.2 keV, the tails of K Kα and Ca Kα) and the scattering due to the

410 interaction of the Mo L-lines (source) with the sample matrix. On the contrary, those around 3.31 (K
411 Kα), 5.90 (Mn Kα), 8.64 (Zn Kα) and 13.40 keV (Rb Kα) characterized samples from Asiago. The
412 contribution of the Cl Kα (2.62 keV) and Ni Kα (7.48 keV) was less relevant in the separation of the
413 two group of beans respect to the other fluorescence lines. For Cl, this was also observed when
414 concentration values were used (see the loading plot in Figure 2B), probably due to the high
415 variability of the data also for the same genotype. In the case of Ni, it gave a relevant contribution for
416 the separation when concentration values were used (Figure 2B), while, with the use of spectra, it
417 gave a very small contribution. This can be explained by the fact that the peak of Ni Kα was very
418 weak (low counts) and, in some cases, its concentration was close to the LOQ (0.5 mg/kg). Low
419 counts cause an error in the quantification. Using the raw spectra, an error in the discrimination
420 produced by the not precise quantification of Ni (due to the low counts) was avoided. In fact, the Ni
421 Kα peak did not significantly influence the classification. With this approach, an *a priori* choice of
422 the elements is avoided. In previous works just some elements were selected for the PCA. Nečemer
423 et al. (2009) chose four elements (Cl, K, Mn and Rb) for the discrimination of honey sources; Obhođaš
424 et al. (2021) selected 5 on 9 quantified elements for the PCA in order to identify the origin of Croatian
425 and Austrian wines; Vitali Čepo et al. (2022) considered the most significant elements (K, Mn, Ni,
426 Rb, Sr and Ba) to distinguish wines from different regions. On the contrary, in this work the
427 contribution from all elements is considered by using TXRF without performing a selection of
428 elements' signal, with the exception of the fluorescence lines of the internal standard which is not
429 part of the sample itself. As known, PCA underlines the inner relation between the variables and, for
430 this reason, when the full spectra are used it points out the link between fluorescence signals, instead
431 of taking into account the single elements absolute values. In this sense, and in the context of a
432 fingerprinting strategy, the use of the internal standard appears not useful. As a confirmation, the
433 elaboration performed both on the normalized spectra and on the signals including the region 9-10.5
434 keV, did not brought to different results (data not shown).

435 To sum up, from data exploration it has been possible to conclude that i) the multivariate analysis of
436 the TXRF spectra could serve as a tool to extract and use information about samples properties; ii)
437 GLSW filter is fundamental to highlight the differences linked to the growing site; iii) coupling SNV
438 with GLSW gave more direct interpretation of the chemical information able to cluster the bean's
439 classes.
440 As a next step, a PLS-DA supervised method for beans classification was built and validated. To this
441 aim, the samples were divided again in calibration and validation sets and the results are reported in
442 Table 3. The performance of GLSW was compared to that of the coupled preprocessing SVN-GLSW.
443 When only GLSW was used, the optimal number of LVs was equal to 2, regardless the subset
444 considered. In calibration, perfect specificity and selectivity of the models were observed without any
445 classification error. In cross-validation, the performances of the models were slightly worse than in
446 calibration. Accordingly, the classification error also slightly increased ranging from 0.04 to 0.06.
447 Nonetheless, the PLS-DA model gave very good results in prediction where, only considering the set
448 1, some erroneous misclassifications were reported. When GLSW was coupled with SNV the results
449 were even better. In fact, in this case minor performance of the classification models was observed
450 only for the set 3 in cross-validation and for the set 1 in prediction, with only one sample of Legnaro
451 that was erroneously classified as coming from Asiago.
452 The good results achieved with this classification model confirm that TXRF spectra can be used as
453 fingerprint for food traceability and that elemental quantification procedure is not necessary. With
454 TXRF, elemental quantification can be performed using different strategies (internal standard
455 addition, external standardization, etc.) and each method can give good or bad results according to
456 the element, the sample and the working conditions (Dalipi et al., 2017). The errors caused by the
457 wrong quantification of the element (due to incorrect calibration or evaluation procedure) cannot
458 occur when the raw spectrum is considered. Moreover, the elimination of a step of the sample
459 preparation (internal standard addition) also reduces further contribution to the overall analytical
460 error associated to the quantification because both systematic and random errors connected to the

internal standard addition are not introduced. Indeed, with this approach, the deletion of the internal standard spectral region did not lead to a misclassification of the beans' origin (Table 3). Of course, all the errors connected with other sample preparation procedures (sample weighting, slurry preparation, deposition on the reflector, centering of the drop, etc.) are not eliminated with this procedure, but these are related with the good practice for a correct TXRF analysis (Klockenkämper and von Bohlen, 2015; von Bohlen and Fernández-Ruiz, 2020). Finally, the elimination of the element quantification step also speeds up the analytical process since both the calibration of the spectrometer, the addition of the internal standard and the concentration calculation (which involves the spectrum evaluation and fitting) are not performed. The sample can be directly classified according to its spectrum and this is crucial for food/seeds screening, authentication, fraud identification and traceability.

4 Conclusions

In the present work, TXRF was used for the first time for the elemental characterization of legumes seeds, and in particular of beans. The developed method was validated and compared with an established technique for the elemental analysis of food (ICP-OES, after sample dissolution). TXRF was applied for the study of the elemental accumulation in 24 different genotypes of beans grown on two sites with different pedo-environmental characteristics. On the base of the solely quantitative data, no clear information about the effect of the genotype and/or growing site on the accumulation of the elements in bean seeds could be drawn. On the contrary, the exploration of the data with PCA performed on the quantitative data allowed to differentiate the two groups of samples based on the growing site, while genotype influenced only the variability of the data. Classification methods (LDA, PCA-LDA and PLS-DA) were also applied obtaining a remarkable classification performance of the samples according to the growing site with low errors in prediction.

In a second step, a model for the identification of the beans' growing site was developed applying a PLS-DA classification model coupled with GLSW filter directly on the raw TXRF spectra. The two groups of beans were clearly clustered even if the influence of the spectral variables on the clustering was not readable. This problem was overcome by preprocessing the spectra with SNV, which allowed to better identify the regions of the TXRF spectrum which characterize the beans coming from the two sites. The elimination of the quantification step allows reducing errors due to the concentration estimation and speed up the process of food analysis and clustering which is crucial for food/seeds screening, authentication, fraud identification and traceability. These results demonstrate for the first time the possibility to use TXRF spectra as a food fingerprint for the identification of the geographical origin. Further investigations on different kinds of food coming from a larger number of sites should be considered in order to confirm and generalize the applicability of the method for food traceability.

Acknowledgement

TXRF analysis were performed at the "Micro X-ray Lab" of the University of Bari (Italy). Ignazio Allegretta was supported by a research grant on the project PON R&I "Studio del sistema suolo-pianta mediante tecniche analitiche innovative che impiegano raggi X" — Progetto AIM1809249 — attività 1, linea 1. Giacomo Squeo was supported by research grant on the project PON R&I "Estrazione, caratterizzazione e valorizzazione di composti bioattivi ottenuti da olivo e legumi" — Progetto AIM1809249 — attività 2, linea 1. The agronomic research was realized withing the BIRD192128/19 project funded by the University of Padova. This work was developed in the frame of COST Action CA18130 ENFORCE TXRF supported by COST (European Cooperation in Science and Technology).

References

Abbas, O., Pissard, A., Baeten, V. (2020). Near-infrared, mid-infrared, and Raman spectroscopy. In Y. Pico, (Ed.), *Chemical Analysis of Food* (pp 77-134) Elsevier

511 Allegretta, I., Gattullo, C.E., Renna, M., Paradiso, V.M., Terzano, R. (2019). Rapid multi-element
512 characterization of microgreens via total-reflection X-ray fluorescence (TXRF) spectrometry. Food
513 Chemistry, 296, 86-93. https://doi.org/10.1016/j.foodchem.2019.05.187
514 Astolfi, S., Pii, Y., Terzano, R., Mimmo, T., Celletti, S., Allegretta, I., Lafiandra, D., Cesco, S.
515 (2018) Does Fe accumulation in durum wheat seeds benefit from improved whole-plant sulfur
516 nutrition?. Journal of Cereal Science, 83, 74-82. https://doi.org/10.1016/j.jcs.2018.07.010
517 Ballabio, D., Consonni, V. (2013) Classification tools in chemistry. Part 1: linear models. PLS-DA.
518 Analithical Methods, 5, 3790-3798. https://doi.org/10.1039/C3AY40582F
519 Ballabio, D., Grisoni, F., Todeschini, R. (2018) Multivariate comparison of classification
520 performance measueres. Chemometrics and Intelligent Laboratory Systems, 174, 33-44.
521 https://doi.org/10.1016/j.chemolab.2017.12.004
522 Benincasa, C., Lewis, J., Sindona, G., Tegarelli, A. (2008) The use of multi element profiling to
523 differentate between cow and buffalo milk, Food Chemistry, 110, 257-262.
524 https://doi.org/10.1016/j.foodchem.2008.01.049
525 Bevilacqua, M., Bro, R., Marini, F., Rinnan, Å., Rasmussen, M. A., Skov, T. (2017) Recent
526 chemometrics advances for foodomics. Trends in Analytical Chemistry, 96, 42-51.
527 https://doi.org/10.1016/j.trac.2017.08.011
528 Binetti, G., Del Coco, L., Ragone, R., Zelasco, S., Perri, E., Montemurro, C., Valentini, R., Naso,
529 D., Fanizzi, F.P., Schena, F.P. (2017) Cultivar classification of Apulian olive oils: use of artificial
530 neural networks for comparing NMR, NIR and mercerological data. Food Chemistry, 219, 131-138.
531 https://doi.org/10.1016/j.foodchem.2016.09.041
532 Blair, M.W., Wu, X., Bhandari, D., Astudillo, C. (2016) Genetic dissection of ICP-detected nutrient
533 accumulation in the whole seed of common bean (Phaseolus vulgaris L.). Frontiers in Plant Science,
534 7. Article 219. https://doi.org/10.3389/fpls.2016.00219

Borgese, L., Bilo, F., Dalipi, R., Bontempi, E., Depero, L.E. (2015). Total reflection X-ray fluorescence as a tool for food screening. Spectrochimica Acta Part B, 113, 1-15. https://doi.org/10.1016/j.sab.2015.08.001

Bro, R., Smilde, A.K. (2014). Principal component analysis. Analytical Methods, 6, 2812-2831. https://doi.org/10.1039/C3AY41907J

Cabrera, C., Lloris, F., Giménez R., Olalla, M., López, C. (2003) Mineral content in légumes and nuts : contribution to the Spanish dietary intake. Science of the Total Environment, 308, 1-14. https://doi.org/10.1016/S0048-9697(02)00611-3

Camin, F., Boner, M., Bontempo, L., Fauhl-Hassek, C., Kelly, S.D., Riedl, J., Rossmann, A. (2017). Stable isotope techniques for verifying the declared geographical origin of food in legal cases. Trends in Food Science and Technology, 61, 176-187. https://doi.org/10.1016/j.tifs.2016.12.007

Dalipi, R., Marguí, E., Borgese, L., Depero, L.E. (2017). Multi-element analysis of vegetal foodstuff by means of low power total reflection X-ray fluorescence (TXRF) spectrometry. Food Chemistry, 218, 348-355. https://doi.org/10.1016/j.foodchem.2016.09.022

Danezis, G.P., Tsagkaris, A.S., Camin, F., Brusic, V., Georgious, C.A. (2016).Food authentication: techniques, trends & emerging approaches. Trends in Analytical Chemistry, 85, 123-132. https://doi.org/10.1016/j.trac.2016.02.026

Espiñeira, M., Santaclara, F.J., (2016). What is food traceability?. In Advances in M. Espiñeira, F.J. Santaclara (Eds.), Food Traceability Techniques and Technologies (pp 3-8). Elsevier.

European Community. Regulation (EC) No 178/2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety. URL (https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32002R0178)

Ghanbari A.A., Shakiba, M.R., Toorchi, M., Choukan, R. (2013). Nitrogen changes in the leaves and accumulation of some minerals in the seeds of red, white, and Chitti beans (Phaseolus vulgaris) under water deficit conditions. Australian Journal of Crop Science, 7, 706-712.

561 Haswell, S.J., Walmasley, A.D. (1998). Multivariate data visualization methods based on multi-
562 elemental analysis of wines and coffees using total reflection X-ray fluorescence analysis. Journal
563 of Analytical Atomic Spectroscopy, 13, 131-134. https://doi.org/10.1039/A705317G
564 ISO 22000:2018. Food safety management systems – Requirements for any organization in the food
565 chain.
566 Klockenkämper, R., von Bohlen, A. (2015). Total-reflection X-ray fluorescence analysis and related
567 methods, John Wiley & Sons Inc.
568 Martens, H., Høy, M., Wise, B.M., Bro, R., Brockhoff, P.B. (2003). Pre- whitening of data by
569 covariance- weighted pre- processing. Journal of Chemometrics, 17, 153-165.
570 https://doi.org/10.1002/cem.780
571 Miller, C. E. (2010). Chemometrics in process analytical technology (PAT). In K.A. Bakeev (Ed.),
572 *Process analytical technology: spectroscopic tools and implementation strategies for the chemical*
573 *and pharmaceutical industries* (pp. 353-438) John Wiley & Sons.
574 Nečemer, M., Košir, I.J., Kump, P., Kropf, U., Jamnik, M., Beronrlj, J., Ogrinc, N., Golob, T.
575 (2009). Application of total reflection X-ray spectrometry in combination with chemometric
576 methods for determination of the botanical origin Slovenian honey, Journal of Agricultural and
577 Food Chemistry, 57, 4409-4414. https://doi.org/10.1021/jf900930b
578 Obhođaš, J., Valković, V., Vinković, A., Sudac, D., Čanađija, I., Pensa, T., Fiket, Ž., Turyanskaya,
579 A., Bretschneider, T., Wilhelmer, C., Gunchin, G., Kregsamer, P., Wobreauschek, P., Streli, C.
580 (2021). X-ray fluorescence techniques for element abundance analysis in wine. ACS Omega, 6,
581 22643-22654. https://doi.org/10.1021/acsomega.1c02731
582 Oliveri, P., Malegori, C., Simonetti, R., Casale, M. (2019). The impact of signal pre-processing on
583 the final interpretation of analytical outcomes–A tutorial. Analytica Chimica. Acta, 1058, 9-17.
584 https://doi.org/10.1016/j.aca.2018.10.055

585 Panchuk, V., Yaroshenko, I., Legin, A., Semenov, V., Kirsanov, D. (2018). Application of
586 chemometric methods to XRF-data – a tutorial revew. Analytica Chimica, 1040, 19-32.
587 https://doi.org/10.1016/j.aca.2018.05.023
588 Ramírez-Ojeda, A.M., Moreno-Rojas, R., Cámara-Martos, F. (2018). Mineral and trace element
589 content in legumes (lentis, chickpeas and beans) : bioaccesibility and probabilistic assessment of the
590 dietary intake. Journal of Food Composition and Analysis, 73, 17-28.
591 https://doi.org/10.1016/j.jfca.2018.07.007
592 Robberecht, H., Van Dyck, K., Bosscher, D., Van Cauwenbergh, R. (2008). Silicon in foods:
593 content and bioavailability. International Journal of Food Properties, 11, 638-645.
594 https://doi.org/10.1080/10942910701584252
595 Roberts, J.J., Cozzolino, D. (2013). An overview on the application of chemometrics in food
596 science and technology—an approach to quantitative data analysis. Food Analytical Methods, 9,
597 3258-3267. https://doi.org/10.1007/s12161-016-0574-7
598 Serranti, S., Cesare, D., Marini, F., Bonifazi, G. (2013). Classification of oat and groat kernels using
599 NIR hyperspectral imaging. Talanta, 103, 276-284. https://doi.org/10.1016/j.talanta.2012.10.044
600 Shand, C.A., Wendler, R., Dawson, L., Yates, K., Stephenson, H. (2017). Multivariate analysis of
601 Scotch whisky by total reflection X-ray fluorescence and chemometric methods: a potential tool in
602 the identification of the counterfeits. Analytica Chimica Acta, 976, 14-24.
603 https://doi.org/10.1016/j.aca.2017.04.041
604 Sica, P., Scariolo, F., Galvao, A., Battaglai, D., Nicoletto, C., Maucieri C., Palumbo, F., Franklin,
605 D., Cabrera, M., Borin, M., Sambo, P., Barcaccia, G. (2021). Molecular hallmarks, agronomic
606 performances and seed nutraceutical properties to exploit neglected genetic resources of common
607 beans grown by organic farming in two contrasting environments. Frontiers in Plant Science, 12,
608 Article 674985 https://doi.org/10.3389/fpls.2021.674985

609 Squeo, G., Grassi, S., Paradiso, V. M., Alamprese, C., Caponio, F. (2019). FT-IR extra virgin olive
610 oil classification based on ethyl ester content. Food Control, 102, 149-156.
611 https://doi.org/10.1016/j.foodcont.2019.03.027
612 US Department of Agriculture (2011). Proposed rule to establish general regulations for improving
613 the traceability of U.S. livestock moving interstate when animal disease events take place. URL
614 (http://www.aphis.usda.gov/traceability/)
615 Vitali Čepo, D., Karoglan, M., Borgese, L., Depero, L.E., Marguí, E., Jablan, J. (2022). Application
616 of benchtop total-reflection X-ray fluorescence spectrometry and chemometrics in classification of
617 origin and type of Croatian wines. Food Chemistry X, 13. Article 100209.
618 https://doi.org/10.1016/j.fochx.2022.100209
619 von Bohlen, A., Fernández-Ruiz, R. (2020). Experimental evidence of matrix effects in total-
620 reflection X-ray fluorescence analysis: coke case. Talanta, 209, 120562.
621 https://doi.org/10.1016/j.talanta.2019.120562
622 Westermann, D.T., Tarán, H., Muñoz-Perea, C.G., Singh, S.P. (2011). Plant and seed nutrient
623 uptake in common bean in seven organic and conventional production systems. Canadian Journal of
624 Plant Science, 91, 1089-1099. https://doi.org/10.4141/cjps10114
625 Winkler, A., Rauwolf, M., Sterba, J.H., Wobrauschek, P., Streli, C., Turyanskaya, (2020). A. Total
626 reflection X-ray fluorescence analysis of elemental composition of herbal infusions and teas.
627 Journal of the Science of Food and Agriculture, 100, 4226-4236. https://doi.org/10.1002/jsfa.10463
628
629
630
631
632
633
634

Figure Caption List

Figure 1. Concentration (mean ± SD, *n*=3) of P, S, Cl, K, Ca, Mn, Fe, Ni, Cu, Zn, Br and Rb in the 24 genotypes of beans grown in the soils of Legnaro and Asiago as determined by TXRF analyses.

Figure 2. Loading plot (A) and score plot (B) of the PCA performed on the concentrations of the elements in the studied beans. A (●), Asiago, L (●), Legnaro.

Figure 3. Score plot and loading plot of PCA after GLSW filter (A) and PCA results after SNV coupled with GLSW (B). A (●), Asiago, L (●), Legnaro. The flat line between 9-10.5 keV indicate the region removed characterized by the presence of Ga fluorescence lines.

Figure S1 TXRF spectrum of a sample of bean.

Figure S2. Score plots (PC1 *vs* PC2 to PC5) and loading plot of the PCA carried out on the smoothed spectral dataset. A (●), Asiago, L (●), Legnaro.

Figure S3. Score plots (PC1 vs PC2 to PC5) and loading plot of the PCA carried out on SNV preprocessed TXRF spectra.

Figure S4. Score plots (PC1 vs PC2 to PC5) and loading plot of the PCA carried out on 1D preprocessed TXRF spectra.

Table Caption List

660 **Table 1.** Comparison between TXRF and ICP-OES results obtained after the analysis of the
661 reference material NIST 1567b
662 **Table 2.** Comparison between TXRF and ICP-OES results obtained on five different samples of
663 bean.
664 **Table 3.** Results of beans classification.
665 **Table S1.** List and pictures of the seeds of the 24 genotypes of beans analysed in this study.
666 **Table S2.** Physical and chemical characteristics of Asiago and Legnaro soils.
667 **Table S3.** List of samples in the calibration and validation sets, respectively.

TXRF spectral information enhanced by multivariate analysis: a new strategy for food fingerprint

Ignazio Allegretta[a]*, Giacomo Squeo[a], Concetta E. Gattullo[a], Carlo Porfido[a], Antonio Cicchetti[a], Francesco Caponio[a], Stefano Cesco[b], Carlo Nicoletto[c], Roberto Terzano[a]

[a] Department of Soil, Plant and Food Sciences, University of Bari Aldo Moro, Via G. Amendola 165/A, 70126, Bari, Italy.

[b] Faculty of Science and Technology, Free University of Bozen-Bolzano, Piazza Università 5, 39100 Bolzano, Italy.

[c] Department of Agronomy, Food, Natural resources, Animals, and Environment (DAFNAE), University of Padova, Viale dell'Università 16, 35020 Legnaro, Italy

*ignazio.allegretta@uniba.it

Abstract

The increased costumers' request of safe and high-quality food products makes food traceability a priority for frauds identification and quality certification. Elemental profiling is one of the strategies used for food traceability, and TXRF spectroscopy is widely used in food analysis even if its potentialities have not been fully investigated. In this work, a new method for food traceability using directly TXRF spectra coupled with multivariate analyses, was tested. Twenty-four different beans' genotypes (*Phaseolus vulgaris* L.) grown onto two different sites have been studied. After the development of the method for beans' analysis, TXRF spectra were collected and processed with PCA combined with SNV and GLSW filter obtaining a perfect clustering of the seeds according to their geographical origin. Finally, using PLS-DA, beans were correctly classified demonstrating that TXRF spectra can be successfully used as fingerprint for food/seed traceability and that elemental quantification procedure is not necessary to this aim.

Keywords: Total-reflection X-ray fluorescence, Seed traceability, Principal component analysis, Standard normal variate, generalized least squares weighting, partial least square discriminant analysis, beans, *Phaseolus vulgaris* L.

1 Introduction

The rediscovering of regional and local biodiversity in a globalized world together with the improved consumers' awareness about the food quality, safety, provenance and production process, pose food traceability at the center of international laws, standards and agreements (EC Regulation 178/2002; US Department of Agriculture, 2011). According to the ISO 22000:2018(en), *traceability* is the "ability to follow the history, application, movement and location of an object through specific stage(s) of production, processing and distribution" which, in other words, is "the ability to identify at any specified stage of the food chain (from the production to distribution) where the food came from (one step back) and where the food went to (one step forward)" (Espiñeira and Santaclara, 2016). This means that the control institution must be able to answer questions about product authenticity and adulterations, identify the geographical origin of the raw material(s), the transformation cycle and the production site(s), certify the quality of the food in terms of content of high-value metabolites, micro and macronutrients and the absence of potentially toxic compounds and elements. All these questions cannot be answered by using one single analytical method, but they should be addressed by several genetic, biological, chemical and biochemical techniques (Danezis et al., 2016). The most used strategies for food traceability are DNA profiling (Sica et al., 2021), molecular profiling (Binetti et al., 2017), isotope analysis (Camin et al., 2017) and elemental profiling (Benincasa et al., 2008). In all the profiling strategies, genes, molecules, stable/instable isotopes and elements are identified and quantified and then data are processed using multivariate data analysis procedures (Roberts and Cozzolino, 2016), among which principal component analysis (PCA), partial least square regression (PLSR), linear discriminant analysis (LDA), and partial least square discriminant analysis (PLS-DA) are some of the most

53 applied methods. In the case of elemental profiling, the quantification is performed mainly by using
54 atomic spectroscopy, and the most used techniques are inductively coupled plasma-optical emission
55 spectroscopy (ICP-OES) or -mass spectroscopy (ICP-MS), atomic absorption spectroscopy (AAS)
56 and X-ray fluorescence spectroscopy (XRF). The first three methods allow to achieve very low
57 limits of detection (reaching also sub ppb levels in the case of ICP-MS), but the dissolution of the
58 sample matrix is mandatory to perform the analysis (with the exception of a particular configuration
59 of AAS called graphite furnace atomic absorption spectroscopy - GFAAS). On the contrary, with
60 XRF the analysis can be performed directly on the matrix even if detection limits are higher than in
61 the abovementioned spectroscopies (1-100 ppm depending on the element and on the instrument
62 configuration). However, with a particular configuration of XRF, named total-reflection X-ray
63 fluorescence spectroscopy (TXRF), better performances can be achieved in terms of detection limits
64 respect to AAS and ICP-OES (Allegretta et al., 2019). In TXRF, the sample is deposited onto a
65 reflector as a thin film and a primary X-ray beam hits the reflector at an angle lower than the critical
66 one (<1°). In these conditions, the primary beam does not refract into the reflector but it is totally
67 reflected, causing the double excitation of the sample. This phenomenon produces a reduction in the
68 matrix absorption effects and an increase in the signal-to-noise ratio, resulting in an improvement of
69 the detection limits (in the range of ppm for solids and ppb for liquids) (Klockenkämper and von
70 Bohlen, 2015). For this reason, TXRF was widely used in the analysis of food (Allegretta et al.,
71 2019; Borgese et al., 2015; Dalipi et al., 2017). Moreover, it is considered a valuable tool for food
72 traceability of coffee (Haswell and Walmasley, 1998), wine (Haswell and Walmasly, 1998;
73 Obhođaš et al., 2021; Vitali Čepo et al., 2022), whisky (Shand et al., 2017), honey (Nečemer et al.,
74 2009) tea and herbal infusions (Winkler et al., 2020). In all these studies, TXRF was used just for
75 the element quantification and, afterwards, the results were processed with multivariate data
76 analysis methods. By using this approach, TXRF has more or less the same applicability of the
77 other atomic spectroscopies. Furthermore, PCA and dendrograms are usually the only applied
78 multivariate methods on these data, and just recently a clustering method with linear discriminant

analysis (LDA) (Vitali Čepo et al., 2022) and regularized discriminant analysis (RDA) (Nečemer et al., 2009) were applied. However, respect to other atomic spectroscopies, TXRF provides a continuum spectrum as an output, which could be potentially used, in combination with chemometric approaches, like a fingerprint of the sample, as it happens in molecular analysis such as nuclear magnetic resonance (NMR) (Binetti et al., 2017), chromatography (Bevilacqua et al., 2017) and infrared (IR) spectroscopies (Abbas et al., 2020; Squeo et al., 2019). The use of XRF spectra like fingerprint is a well-established method in other research fields (Panchuk et al., 2018). On the basis of the authors' knowledge, XRF spectra and, in particular, TXRF ones, have never been used as food fingerprint and/or for traceability purposes.

The present work aims at developing, testing and validating a new method of food fingerprinting and traceability using TXRF spectra. To fulfill this goal, a group of 24 different genotypes of beans (*Phaseolus vulgaris* L.) coming from two cultivation areas were analyzed. Since there is no available procedure for the analysis of legumes (and in particular beans) with TXRF, the first step consisted in the development of a TXRF analytical method where samples were prepared as suspension. The method was validated comparing both results and performances with an official analytical technique (ICP-OES). Then, the elemental quantitative data as well as the TXRF signals were processed by unsupervised (PCA) and supervised multivariate techniques (LDA, PLS-DA), with the aim at developing a classification method for geographical origin of beans according to the growing site.

2 Materials and Methods

2.1 Beans and growing sites

Twenty-four genotypes of beans (Table S1) were sown into two different sites: one was the Experimental Farm "L. Toniolo" of the University of Padova in Legnaro (Veneto region, Italy, 7 m a.s.l.), situated at 30 km-distance from the coast; the other was a local farm in Asiago (Veneto

104 region, Italy, 994 m a.s.l.), located in a mountain region at 90 km-distance from the coast. The
105 physical-chemical characteristics of the two soils are reported in Table S2. The Asiago soil is a silty
106 clay-silty clay loam soil with a neutral pH with a ratio C/N of 15.11. The soil of Legnaro is a slight
107 alkaline silt loam soil having a lower C/N ratio than the Asiago soil. Both soils have a similar
108 elemental composition with the exception of Zn and Cl which are more concentrated in Asiago and
109 Legnaro, respectively. Moreover, in Asiago, Mn, Fe, Ni and Zn are more available than in Legnaro.
110 Seeds were preliminary sown in seedling starter trays (one seed per hole) using a peat-based geo
111 substrate (PBE Substrates, perlite grain size = 2-6 mm, pH = 5-6, EC = 0.4 mS/cm), keeping them
112 in a protect environment, providing a regular and optimal watering. After 10 days, the plantlets
113 were transplanted into the two soils and cultivated using the same agronomical practices. Pods were
114 harvested at commercial ripening. Overall, the crop cycle ended 136 and 137 days after the
115 transplanting for Legnaro and Asiago, respectively. Further specifications about the agricultural
116 practices adopted are reported in Sica et al. (2021).
117 Before the analysis, all the 48 bean samples were left drying in the oven for 48 h at 105 °C. Then,
118 the samples were powdered with a vibro-milling system (MM 400, Retsch GmbH, Haan, Germany)
119 equipped with ZrO_2 jars and balls, for 2 min at 30 Hz (mean particle size ≈ 60 µm). The powdered
120 samples were stored in a desiccator until their preparation for TXRF and ICP-OES analyses.

121

122 *2.2 TXRF analysis*

123 For TXRF analysis, 100.0 ± 5.0 mg of powder were suspended in 5 mL of 1% Triton X-100
124 solution (Sigma Aldrich CHEMIE GmbH, Steinheim, Germany) (Allegretta et al., 2019).
125 Furthermore, 10 µL of Ga (1000 mg/L, Sigma Aldrich CHEMIE GmbH, Steinheim, Germany) were
126 added as internal standard. The slurry was shaken using a vortex, then it was placed in an ultrasonic
127 bath for 15 min. After mixing again with vortex, 10 µL of the slurry were pipetted onto a siliconized
128 (SERVA Electrophoresis GmbH, Heidelberg, Germany) quartz reflector, and the sample was dried

for 10 min on a hot plate at 50°C. Sample deposition and drying were carried out under a laminar flow hood to avoid the specimen contamination.

To check the accuracy of the method, a reference material must be analyzed. However, there is no reference material of trace elements in beans or at least in legumes which can be used for this purpose. The reference material "Wheat flour" (NIST 1567b) has a matrix which is similar to the beans, and for this reason it was chosen and prepared using the same procedure applied for the bean samples. Quartz reflectors were then analyzed with an S2 Picofox TXRF Spectrometer (Bruker Nano GmbH, Berlin, Germany) equipped with a Mo source (50 kV, 0.6 mA, 30W), a multilayer monochromator and an XFlash® silicon drift detector of 20 cm^2 (FWHM < 140 eV on Mn-Kα). Three suspensions for each sample were prepared and a live time of 1000 s was selected for the TXRF analyses.

2.3 ICP-OES analysis

For ICP-OES analyses, 150 mg of powdered sample or reference material (NIST 1567b) were pre-digested overnight with 7 mL of ultrapure HNO_3 (≥69.0%, TraceSELECT®) and 1 mL of H_2O_2 (30%, TraceSELECT®), in Teflon tubes, at room temperature. Tubes were then placed in a MultiwaveGO microwave (Anton Paar, Graz, Austria) and a single step heating program (20 min to 180 °C, held for 15 min, and final cooling) was used for digestion. A blank solution (7 mL HNO_3 and 1 mL H_2O_2) underwent the same process. Finally, the mineralized samples were diluted to 25 mL with deionized water, filtered with Whatman® 42 filter paper and stored at 4 °C until analysis. Both the samples and the reference material were run in triplicate, while more than ten replications were run for the blank. Analyses were led with a Thermo iCAP 6000 series ICP-OES (Thermo Fisher Scientific Inc., Waltham, MA, USA) using the following settings: 1.15 kW RF power; 0.50 L/min auxiliary Ar flow; 0.90 L/min nebulizer Ar flow (0.5 L/min for P, K, Ca and Sr determination); 12 L/min coolant Ar flow; axial plasma configuration.

A two points calibration curve was employed, using i) the blank acidic solution as zero point, ii) a calibration solution prepared at the concentrations of 2 mg/L, or iii) a calibration solution prepared at the concentrations of 10 mg/L. Calibration solutions were prepared by appropriately diluting a multi-element calibration reference solution (Certipur® ICP Multi-element standard solution IV, 1000 mg/L, Merck GaA, Darmstadt, Germany) and an P standard solution (1000 mg/L, Carlo Erba, Milano, Italy) with the blank acidic solution. Analysis of major elements was carried out after diluting the samples with deionized water (1:10 v/v for P, K and Ca). The same dilution with deionized water was applied for the standard solutions used for calibration (including the "zero point"). The following emission wavelengths (nm) were selected for the elements' detection and quantification: 213.62 (P), 766.49 (K), 422.67 (Ca), 257.61 (Mn), 259.94 (Fe), 231.60 (Ni), 324.75 (Cu), and 213.85 (Zn). Each sample was analyzed in triplicate.

2.4 Parameters of validation

Validation of the analytical method was performed determining accuracy and precision of the analysis of the certified reference material NIST 1567b. Accuracy was assessed as the recovery with respect to the certified element's concentration, and precision was estimated as the standard deviation of the replicates (n=3). The limits of detection (LOD) and quantification (LOQ) were estimated as three and ten times, respectively, the parameter σ. The latter is calculated as the standard deviation of 10 replicates of the blank for ICP-OES, while $\sigma = C_i \frac{\sqrt{N_{BG}}}{N_i}$ for TXRF (where C_i = concentration of the element, N_i = net area of the element peak expressed in counts, N_{BG} = area of the background under the element peak expressed in counts) (Klockenkämper and von Bohlen, 2015).

177 *2.5 Multivariate data analysis*

178 Multivariate data analyses were carried out on two distinct datasets. The first dataset (144 × 12,
179 samples – including replicates – × elements) consisted of the elementary quantitative data of the
180 samples. These data were explored by Principal Component Analysis (PCA). Before, column
181 autoscaling (i.e., the mean value of the *i*th variable is subtracted from each entry of the variable and
182 then is divided by its standard deviation) was applied (Bro and Smilde, 2014). Then, Linear
183 Discriminant Analysis (LDA), PCA-LDA, and Partial Least Square Discriminant Analysis (PLS-DA)
184 were used as supervised classification methods for beans discrimination according to the growing
185 site. Before, the dataset was split in a calibration set (33 independent samples, 99 considering the
186 three replicates per each independent sample) and a validation set (15 independent samples, 45
187 considering the three replicates per each independent sample). Three different calibration and
188 validation sets were randomly created and used to evaluate the robustness of the models. The samples
189 list per each set is reported in Table S3. Each classification model was built on the calibration set,
190 evaluated by means of a cross-validation step and then tested on the corresponding real external
191 validation set. A venetian blinds scheme was used for cross-validation, with 10 data splits and a blind
192 thickness equal to 3. The optimal number of principal components (PCs) in PCA-LDA and the
193 optimal number of latent variables (LVs) for PLS-DA classification models were defined according
194 to the lowest classification error in cross-validation.

195 The second dataset consisted of the TXRF spectral dataset. First, from the original matrix (144 ×
196 4096, samples – including replicates – × channels in the TXRF spectrum) three uninformative spectral
197 regions were removed since the fluorescence signal of these regions was not connected to the sample
198 itself: i) the spectral region with E≤1.9 keV (which is characterized by a strong noise and the Si K-
199 alpha peak coming from the reflector); ii) the region with 9≤E≤10.5 keV (characterized by the
200 presence of Ga fluorescence lines); iii) the part with E≥15 keV (Rayleigh and Compton scattering
201 peaks). This procedure led to a final matrix of 144 × 2312 that was explored by PCA. Before PCA,
202 the data matrix was subjected to different signal preprocessing. At the beginning, only smoothing

(polynomial order 2, points 11) and mean centering were applied. Subsequently, also Standard Normal Variate (SNV), first derivative (1D; Savitzky-Golay, polynomial order 2, points 11) and Generalized Least Squares Weighting (GLSW; clutter source: x-block classes, $\alpha = 0.02$) were tested separately or in combination. After exploration, PLS-DA was used as supervised classification method for beans discrimination according to the growing site. The same sample subsets (Table S3) were used for the training and the testing of the models' performance. The optimal number of LVs for PLS-DA classification model was defined according to the lowest classification error in cross-validation. A venetian blinds scheme was used for cross-validation, with 10 data splits and a blind thickness equal to 3. For all the supervised classification analyses carried out on both the datasets, sensitivity, specificity, and classification error were used as figures of merit to evaluate the performance of the models (Ballabio et al., 2018). Per each class, sensitivity indicates the true positive rate, *i.e.*, the number of samples belonging to P-class that are correctly classified as P; specificity indicates the true negative rate, *i.e.*, the number of samples from class N correctly rejected from P-class; the overall classification error is calculated as 1 – mean of the sensitivities for the two classes. Multivariate data analyses were carried out in Matlab environment (The MathWorks, Inc. MA, USA) using original codes, the PLS_toolbox (Eigenvector Research Inc., USA) and the classification toolbox (Ballabio et al., 2013).

3 Results and Discussion

3.1 TXRF analysis of beans: method development, validation and comparison with ICP-OES

To assess the performance of TXRF and validate the method, analyses were led on the NIST 1567b reference material. The TXRF spectrum revealed the presence of P, S, Cl, K, Ca, Mn, Fe, Cu, Zn, Se, Br and Rb whose concentrations are reported in Table 1. For TXRF results, the recovery on all the elements ranged between 80% (P) and 116% (Ca) demonstrating a good accuracy of the method. The lower accuracy for P can be imputed to absorption effects which occur in particular

when samples are prepared as slurries (Dalipi et al., 2017). The precision was also good for all the elements (RSD<10%) except for Fe that reached the 29%. This could be related to a non-perfect homogeneity of the standard material a therefore to the different amount of Fe which is deposited onto the reflector. In fact, according to the reference material's certificate, a mass of 500 mg should be used to obtain a uniform sample for Fe determinations. However, this sample mass is too high for the preparation of suspensions for TXRF analysis. Moreover, the mass deposited onto each reflector is much lower than the suggested one and this can result in a low precision for Fe. The LODs and LOQs decreased with the increase of the atomic number ranging from 24 and 80 mg/kg for P to 0.1 and 0.5 for Br, respectively. Rubidium was detected but its concentration in the reference material was too close to the LOQ and, therefore, its quantification was not accurate nor precise. Nickel was neither certified nor detected and its LOD and LOQ were 0.2 and 0.6 mg/kg, respectively.

In the case of ICP-OES results, the recovery ranged between 80% (Ca) to 118% (Cu) demonstrating a good accuracy, similar to TXRF, with the exception of Fe. For this latter element the recovery was 128%. In this case a better precision is achieved for Fe. LOD and LOQ were always higher than those obtained with TXRF, except for P whose LOD and LOQ were an order of magnitude lower than with TXRF.

Finally, the ratio between TXRF and ICP-OES data demonstrated a good agreement between results obtained by the two techniques. The only two exceptions were recorded in the case of Fe and Ca since their recovery was at the minimum acceptable level in the case of ICP-OES analysis (80%). The difference in Ca quantification was also observed in the analysis of other food and plant materials (Allegretta et al., 2019) and it is likely due to the not complete digestion of the sample caused by the presence of Si in the cell wall as occurs in wheat (Robberecht et al., 2008). Complete digestion of the sample using hydrofluoric acid would require a special sample nebulizer or elimination of HF by additional sample manipulation. In addition, HF is extremely dangerous to handle and therefore its use is discouraged.

Since the certified reference material used is not exactly the same material object of this study, a comparison between TXRF and ICP-OES results on the analysis of beans was carried out in order to see if the same analytical performances obtained on the NIST 1567b could be achieved when analysising beans. To study such performance, the ratio between TXRF and ICP-OES results was used and it was compared with the one obtained for the NIST 1567b. For this purpose, 5 different bean samples were analyzed with both TXRF and ICP-OES, and elements concentrations are reported in Table 2. In general, in bean samples the investigated elements were more concentrated than in NIST 1567b. This was particularly evident in the case of K whose concentration was an order of magnitude higher than in the reference material. Nickel was quantified only with TXRF, since its concentration was lower than the LOD (3.5 mg/kg) for ICP-OES. The ratio between the results obtained with TXRF and ICP-OES obtained for the beans is very similar to that obtained for the analysis of the certified reference material (Table 1). With the exception of Ni, for all the elements, TXRF/ICP-OES ranged between 0.85 and 1.03, which underlines a good agreement between the two sets of data. In the case of Ca, this ratio was 1.25 and confirmed the results obtained analyzing NIST 1567b (1.45 ratio), confirming a discrepancy between the two methods for the quantification of this element, probably deriving from the uncomplete sample dissolution, as discussed before.

3.2 Traceability of beans on the base of their elemental composition

A total of 48 bean samples harvested from the site of Asiago or Legnaro were analyzed with TXRF in triplicate. A total of 12 elements were detected in the bean samples (P, S, Cl, K, Ca, Mn, Fe, Ni, Cu, Zn, Br and Rb as shown in Figure S1) and the results are reported in Figure 1. Independently from the origin and genotype, K was the most concentrated element (12783-19586 mg kg^{-1}) followed by P (3394-6194 mg kg^{-1}), S (1607-3067 mg kg^{-1}) and Ca (447-2692 mg kg^{-1}). Manganese (9.9-25.7 mg kg^{-1}), Fe (41.8-102.7 mg kg^{-1}), Ni (0.5-4.2 mg kg^{-1}), Cu (7.5-15.2 mg kg^{-1}) and Zn (19.5-45.6 mg kg^{-1}) were found in trace levels, in agreement with the literature (Blair et al.,

2016; Cabrera et al., 2003; Ramírez-Ojeda et al., 2018). The concentrations of Br (0.4-3.1 mg kg^{-1}) and Rb (1.1-7.5 mg kg^{-1}) were also in trace levels but no data on these two elements are available in the literature for a comparison. Results did not show any capability of specific bean genotypes to accumulate a particular element(s) (Figure 1). On the contrary, an influence of the growing site on the concentration of some elements in beans was observed. Independently from the genotype, the elements K, Mn, Ni, Zn and Rb are more concentrated in the beans grown in Asiago. Halides, S, Ca and Cu seemed to be more concentrated in the beans harvested in Legnaro, but also in this case some exceptions were found (Figure 1). Only the concentration of P was not influenced by the growing site (Figure 1). Finally, Br is lower than LOD in some beans from Asiago.

This strong dependance of certain elements concentrations with the growing site is pointed out in previous works on beans and food in general. In fact, even if Blair et al. (2016) demonstrated that the accumulation of certain elements is correlated to specific DNA loci, the variability of the concentration of macro and microelements is affected by the fact that the two studied genotypes were grown in different kind of soils and conditions.

However, such univariate data analysis (Figure 1) could not provide a clear discrimination of the samples and of the effect of the considered variables (genotype and growing site) on the mineral content of beans. For this reason, data were also treated using a multivariate data analysis approach. Figure 2 reports the results of the PCA carried out on the elemental concentrations data. The first two PCs explained roughly 50% of the data variability. In the case of totally independent variables, this would mean that PC1 and PC2 wrapped up the information of around 6 of the original variables (Bro and Smilde, 2014). However, the loading plot in Figure 2A shows that several variables were well correlated, indicating that the great part of the information could be already observed in the first two PCs. In fact, the exploration of other PCs (data not shown) did not highlight useful information to our aim (*i.e.*, clusters according to the growing site and/or genotype of the beans). The loading plot shows roughly two/three variables clusters that could be separated along PC1 and PC2. Along PC1, Rb is one cluster, followed by the second made up of Mn, Ni, K, Zn, Br, Cl, P, Ca, and the last one

represented by Fe, S and Cu. Along PC2, the first cluster was made up of Ca, P, Cl, Br, Cu, Fe, S while the other one of K, Zn, Mn, Ni and Rb. In the relative score plot (Figure 2B), the samples were well clustered according to the site. Beans from Asiago were located roughly along the positive half of PC2, due to the higher content of K, Zn, Mn, Ni and Rb, while the opposite could be stated for samples from Legnaro, which were mostly characterized by higher contents of Ca, P, Cl, Br, Cu, Fe, S. Similarly to other studies (Nečemer et al., 2009; Obhođaš et al., 2021; Shand et al., 2017; Vitali Čepo et al., 2022), Cl and Br demonstrated to be very important for the identification of the food origin. These two elements cannot be quantified with analytical methods which require the mineralization of the matrix (i.e. ICP-OES) because it causes the loss of both halogens (Allegretta et al., 2019), and for this reason the use of TXRF or other XRF spectroscopies (Obhođaš et al., 2021) is mandatory for their quantification.

The concentrations of Mn, Ni and Zn measured in beans were in accordance with the values of the plant-available fraction measured in the two soils for the three elements (Table S2). In particular, a higher content of Mn, Zn and Ni was found in beans cultivated on the Asiago soil, which was characterized by higher values of available fractions of Mn, Zn and Ni compared to the Legnaro one. The higher accumulation of Cl in beans from Legnaro also seemed to be related to the higher content of this halide in the soil. Nevertheless, for other elements such as P, K, Ca, Fe and Cu, no clear relation with the soil chemical properties was observed. However, it should be considered that the process which led to the accumulation of elements in plants (and in seeds) depends on several variables such as genetic traits, soil characteristics, environmental conditions, physiological processes, plant symbiosis with soil microorganisms, etc. (Astolfi et al., 2018; Ghanbari et al., 2013; Westermann et al., 2011), and that the soil chemical properties alone cannot fully explain this clustering. Although the samples were well clustered, a high dispersion within each class was observed, likely due to the genetic variability of the beans, and no clusters were observed according to the genotype. Finally, it should be noticed that the "Fagiolo nano Valdarno" grown in Legnaro (sample 17L) clustered together with the Asiago samples. This can be related to its higher concentration of Mn and Ni and

lower concentration of S. Overall, the PCA proved that the beans' elemental composition was affected by the growing site, and therefore it was an encouraging result toward a geographic fingerprinting. To test this hypothesis, two supervised multivariate classification techniques were used, namely LDA and PLS-DA. The results of the classification are reported in Table 3. Both the models allowed to reach a very good discrimination according to the growing sites. The results were slightly influenced by the specific random subset used, which affected in some cases the figures of merit of the classification in cross-validation or prediction. Focusing on the prediction results, a prediction error of 0.16 was observed only in one case using LDA for a total of 5 erroneous class assignments. In accordance with the PCA results (Figure 2), the most discriminant variables were Mn, Zn, Rb, Ni, Ca, Fe, Cu and S. Additionally, LDA could be even coupled with PCA and, in such a case, after having calculated a PCA model, the scores were used as input in LDA. PCA-LDA modelling allowed to achieve similar or even better results than LDA with lower error in prediction, which were observed only in the case of set 3. Finally, by using PLS-DA, in two cases some misclassifications were observed bringing to a classification error of 0.09 linked, in both the cases, to an erroneous assignment of only 2 samples.

Overall, the classification results were very promising. In fact, in most of the cases, sensitivity and specificity of 100% and no classification errors were obtained. The performances achieved in classification are better than previous works done on the application of LDA on TXRF data for tracing food origin (Nečemer et al., 2009; Vitali Čepo et al., 2022). Surely, an increase in the numerosity of the dataset would be useful for testing and confirming these performances.

Nonetheless, it should be considered that this approach still requires the quantification of the elements which consists of several steps before (i.e., calibration of the instrument) and after (i.e., spectra evaluation and fitting) analysis, which slow down the outcome of the analysis. In food traceability, and in particular for the identification of frauds, the analytical process should be as fast as possible. As an alternative and to speed up the whole analytical process, multivariate statistical tools could be directly used to process the TXRF signals, as reported in the following section.

3.3 Development and validation of a fingerprint approach using TXRF spectra

The full TXRF dataset was firstly explored by PCA. The literature about the multivariate analysis of TXRF signal is scarce or totally absent. Consequently, the raw profiles were preliminarily inspected to identify proper preprocessing methods. Signal preprocessing is a fundamental step to remove irrelevant and/or disturbing signals that might interfere during subsequent data analysis (Olivieri et al., 2019). The raw signals did not show any specific systematic and/or random unwanted variation, so that only smoothing was applied in the first step and, after mean centering, PCA was performed.

The first two PCs explained 97.5% of the data variability. However, because it was expected that the major source of variability in the dataset was linked to the genotype, 5 PCs were initially calculated for the exploration purpose, accounting for a total of 99.80% of data variability. The score plots (PC1 *vs* PC2 to PC5) together with the relative loading plot are reported in Figure S2.

PC1 explained more than 90% of the data variability but, as expected, it did not highlight information linked to the bean's classes. PC2 seemed to highlight some useful information with Asiago samples that mostly laid on the negative half of the plot, while Legnaro samples roughly on the positive side. In any case, such a clustering is not so straightforward because of a clear overlapping among the classes. From the corresponding loading plot, it was figured out that the main contribution to this displacement was due to the lower intensity of the Ca K-lines (3.69 and 4.01 keV) in Asiago beans. One sample from Asiago, namely 29A (all the three replicates) felt in the positive part of PC2. The exploration of the other PCs did not point out other useful insight according to the problem under study. Overall, although samples clustering along PC2 was not so straightforward, from data exploration it seemed that useful information about the growing site could be retrieved by a multivariate approach applied to the TXRF signal. These results were compared to those obtained by applying SNV or first derivative preprocessing to the TXRF spectra (Figure S3 and S4) but even the application of these preprocessing did not clearly improve the clustering of the samples according to the growing site.

After this starting data exploration, it was clear that the relevant information was masked by other sources of variability. To face with this issue, generalized least squares weighting (GLSW) filter was applied. GLSW is a powerful orthogonal filter successfully used both in regression and classification contexts (Martens et al., 2003; Serranti et al., 2013). In short, it removes the variation in the data not related to properties of interest. In the case of classification problems, like the one in this study, GLSW removes the within-class variance as much as possible without reducing the between-class variance (Serranti et al., 2013). More details about this filter could be found elsewhere (Martens et al., 2003; Miller et al., 2010). Figure 3A reports the results of the PCA carried out on the GLSW preprocessed data matrix. The efficacy of the GLSW filter could be clearly caught from the score plot where samples clustered perfectly according to the growing site along the first PC. However, on the other hand, after applying GLSW, the loading plot became less interpretable. Consequently, it becomes a hard task to highlight the importance of the original variables to samples clustering although it could be argued that the variables between 2 and 4 keV and those after 14 keV showed higher loadings value. The effect of SNV preprocessing before GLSW was tested and the results are reported in Figure 3B. From the score plot it could be observed that samples still clustered perfectly according to the growing site although the distribution changed and a higher dispersion within each class was observed. In this sense, it seemed that SNV reduced to some extent the ability of GLSW to reduce the within class variance or, on the other hand, that some scattering was contributing to the samples clustering. However, the main advantage of using SNV is highlighted in the loading plot where a much more meaningful profile is observed. Considering that the samples clustered along PC1, only the relative loadings have been reported, and it could be observed that the variables at around 2.0, 2.3, 3.2, 3.4, 3.7, 6.4, 8.1 and 11.9 keV showed higher positive loadings on PC1 and thus characterize samples from Legnaro. Those regions correspond to the P Kα (2.01 keV), S Kα (2.31 keV), Ca Kα (3.69 keV), Fe Kα (6.40 keV), Cu Kα (8.05 keV) and Br Kα (11.92 keV). The positive peaks at 3.2 and 3.4 keV can be ascribed to several processes such as the overlapping of fluorescence lines (atmospheric Ar Kβ at 3.2 keV, the tails of K Kα and Ca Kα) and the scattering due to the

410 interaction of the Mo L-lines (source) with the sample matrix. On the contrary, those around 3.31 (K
411 Kα), 5.90 (Mn Kα), 8.64 (Zn Kα) and 13.40 keV (Rb Kα) characterized samples from Asiago. The
412 contribution of the Cl Kα (2.62 keV) and Ni Kα (7.48 keV) was less relevant in the separation of the
413 two group of beans respect to the other fluorescence lines. For Cl, this was also observed when
414 concentration values were used (see the loading plot in Figure 2B), probably due to the high
415 variability of the data also for the same genotype. In the case of Ni, it gave a relevant contribution for
416 the separation when concentration values were used (Figure 2B), while, with the use of spectra, it
417 gave a very small contribution. This can be explained by the fact that the peak of Ni Kα was very
418 weak (low counts) and, in some cases, its concentration was close to the LOQ (0.5 mg/kg). Low
419 counts cause an error in the quantification. Using the raw spectra, an error in the discrimination
420 produced by the not precise quantification of Ni (due to the low counts) was avoided. In fact, the Ni
421 Kα peak did not significantly influence the classification. With this approach, an *a priori* choice of
422 the elements is avoided. In previous works just some elements were selected for the PCA. Nečemer
423 et al. (2009) chose four elements (Cl, K, Mn and Rb) for the discrimination of honey sources; Obhođaš
424 et al. (2021) selected 5 on 9 quantified elements for the PCA in order to identify the origin of Croatian
425 and Austrian wines; Vitali Čepo et al. (2022) considered the most significant elements (K, Mn, Ni,
426 Rb, Sr and Ba) to distinguish wines from different regions. On the contrary, in this work the
427 contribution from all elements is considered by using TXRF without performing a selection of
428 elements' signal, with the exception of the fluorescence lines of the internal standard which is not
429 part of the sample itself. As known, PCA underlines the inner relation between the variables and, for
430 this reason, when the full spectra are used it points out the link between fluorescence signals, instead
431 of taking into account the single elements absolute values. In this sense, and in the context of a
432 fingerprinting strategy, the use of the internal standard appears not useful. As a confirmation, the
433 elaboration performed both on the normalized spectra and on the signals including the region 9-10.5
434 keV, did not brought to different results (data not shown).

To sum up, from data exploration it has been possible to conclude that i) the multivariate analysis of the TXRF spectra could serve as a tool to extract and use information about samples properties; ii) GLSW filter is fundamental to highlight the differences linked to the growing site; iii) coupling SNV with GLSW gave more direct interpretation of the chemical information able to cluster the bean's classes.

As a next step, a PLS-DA supervised method for beans classification was built and validated. To this aim, the samples were divided again in calibration and validation sets and the results are reported in Table 3. The performance of GLSW was compared to that of the coupled preprocessing SVN-GLSW. When only GLSW was used, the optimal number of LVs was equal to 2, regardless the subset considered. In calibration, perfect specificity and selectivity of the models were observed without any classification error. In cross-validation, the performances of the models were slightly worse than in calibration. Accordingly, the classification error also slightly increased ranging from 0.04 to 0.06. Nonetheless, the PLS-DA model gave very good results in prediction where, only considering the set 1, some erroneous misclassifications were reported. When GLSW was coupled with SNV the results were even better. In fact, in this case minor performance of the classification models was observed only for the set 3 in cross-validation and for the set 1 in prediction, with only one sample of Legnaro that was erroneously classified as coming from Asiago.

The good results achieved with this classification model confirm that TXRF spectra can be used as fingerprint for food traceability and that elemental quantification procedure is not necessary. With TXRF, elemental quantification can be performed using different strategies (internal standard addition, external standardization, etc.) and each method can give good or bad results according to the element, the sample and the working conditions (Dalipi et al., 2017). The errors caused by the wrong quantification of the element (due to incorrect calibration or evaluation procedure) cannot occur when the raw spectrum is considered. Moreover, the elimination of a step of the sample preparation (internal standard addition) also reduces further contribution to the overall analytical error associated to the quantification because both systematic and random errors connected to the

internal standard addition are not introduced. Indeed, with this approach, the deletion of the internal standard spectral region did not lead to a misclassification of the beans' origin (Table 3). Of course, all the errors connected with other sample preparation procedures (sample weighting, slurry preparation, deposition on the reflector, centering of the drop, etc.) are not eliminated with this procedure, but these are related with the good practice for a correct TXRF analysis (Klockenkämper and von Bohlen, 2015; von Bohlen and Fernández-Ruiz, 2020). Finally, the elimination of the element quantification step also speeds up the analytical process since both the calibration of the spectrometer, the addition of the internal standard and the concentration calculation (which involves the spectrum evaluation and fitting) are not performed. The sample can be directly classified according to its spectrum and this is crucial for food/seeds screening, authentication, fraud identification and traceability.

4 Conclusions

In the present work, TXRF was used for the first time for the elemental characterization of legumes seeds, and in particular of beans. The developed method was validated and compared with an established technique for the elemental analysis of food (ICP-OES, after sample dissolution). TXRF was applied for the study of the elemental accumulation in 24 different genotypes of beans grown on two sites with different pedo-environmental characteristics. On the base of the solely quantitative data, no clear information about the effect of the genotype and/or growing site on the accumulation of the elements in bean seeds could be drawn. On the contrary, the exploration of the data with PCA performed on the quantitative data allowed to differentiate the two groups of samples based on the growing site, while genotype influenced only the variability of the data. Classification methods (LDA, PCA-LDA and PLS-DA) were also applied obtaining a remarkable classification performance of the samples according to the growing site with low errors in prediction.

In a second step, a model for the identification of the beans' growing site was developed applying a PLS-DA classification model coupled with GLSW filter directly on the raw TXRF spectra. The two groups of beans were clearly clustered even if the influence of the spectral variables on the clustering was not readable. This problem was overcome by preprocessing the spectra with SNV, which allowed to better identify the regions of the TXRF spectrum which characterize the beans coming from the two sites. The elimination of the quantification step allows reducing errors due to the concentration estimation and speed up the process of food analysis and clustering which is crucial for food/seeds screening, authentication, fraud identification and traceability. These results demonstrate for the first time the possibility to use TXRF spectra as a food fingerprint for the identification of the geographical origin. Further investigations on different kinds of food coming from a larger number of sites should be considered in order to confirm and generalize the applicability of the method for food traceability.

Acknowledgement

TXRF analysis were performed at the "Micro X-ray Lab" of the University of Bari (Italy). Ignazio Allegretta was supported by a research grant on the project PON R&I "Studio del sistema suolo-pianta mediante tecniche analitiche innovative che impiegano raggi X" — Progetto AIM1809249 — attività 1, linea 1. Giacomo Squeo was supported by research grant on the project PON R&I "Estrazione, caratterizzazione e valorizzazione di composti bioattivi ottenuti da olivo e legumi" — Progetto AIM1809249 — attività 2, linea 1. The agronomic research was realized withing the BIRD192128/19 project funded by the University of Padova. This work was developed in the frame of COST Action CA18130 ENFORCE TXRF supported by COST (European Cooperation in Science and Technology).

References

Abbas, O., Pissard, A., Baeten, V. (2020). Near-infrared, mid-infrared, and Raman spectroscopy. In Y. Pico, (Ed.), *Chemical Analysis of Food* (pp 77-134) Elsevier

511 Allegretta, I., Gattullo, C.E., Renna, M., Paradiso, V.M., Terzano, R. (2019). Rapid multi-element
512 characterization of microgreens via total-reflection X-ray fluorescence (TXRF) spectrometry. Food
513 Chemistry, 296, 86-93. https://doi.org/10.1016/j.foodchem.2019.05.187
514 Astolfi, S., Pii, Y., Terzano, R., Mimmo, T., Celletti, S., Allegretta, I., Lafiandra, D., Cesco, S.
515 (2018) Does Fe accumulation in durum wheat seeds benefit from improved whole-plant sulfur
516 nutrition?. Journal of Cereal Science, 83, 74-82. https://doi.org/10.1016/j.jcs.2018.07.010
517 Ballabio, D., Consonni, V. (2013) Classification tools in chemistry. Part 1: linear models. PLS-DA.
518 Analithical Methods, 5, 3790-3798. https://doi.org/10.1039/C3AY40582F
519 Ballabio, D., Grisoni, F., Todeschini, R. (2018) Multivariate comparison of classification
520 performance measueres. Chemometrics and Intelligent Laboratory Systems, 174, 33-44.
521 https://doi.org/10.1016/j.chemolab.2017.12.004
522 Benincasa, C., Lewis, J., Sindona, G., Tegarelli, A. (2008) The use of multi element profiling to
523 differentate between cow and buffalo milk, Food Chemistry, 110, 257-262.
524 https://doi.org/10.1016/j.foodchem.2008.01.049
525 Bevilacqua, M., Bro, R., Marini, F., Rinnan, Å., Rasmussen, M. A., Skov, T. (2017) Recent
526 chemometrics advances for foodomics. Trends in Analytical Chemistry, 96, 42-51.
527 https://doi.org/10.1016/j.trac.2017.08.011
528 Binetti, G., Del Coco, L., Ragone, R., Zelasco, S., Perri, E., Montemurro, C., Valentini, R., Naso,
529 D., Fanizzi, F.P., Schena, F.P. (2017) Cultivar classification of Apulian olive oils: use of artificial
530 neural networks for comparing NMR, NIR and mercerological data. Food Chemistry, 219, 131-138.
531 https://doi.org/10.1016/j.foodchem.2016.09.041
532 Blair, M.W., Wu, X., Bhandari, D., Astudillo, C. (2016) Genetic dissection of ICP-detected nutrient
533 accumulation in the whole seed of common bean (Phaseolus vulgaris L.). Frontiers in Plant Science,
534 7. Article 219. https://doi.org/10.3389/fpls.2016.00219

535 Borgese, L., Bilo, F., Dalipi, R., Bontempi, E., Depero, L.E. (2015). Total reflection X-ray
536 fluorescence as a tool for food screening. Spectrochimica Acta Part B, 113, 1-15.
537 https://doi.org/10.1016/j.sab.2015.08.001
538 Bro, R., Smilde, A.K. (2014). Principal component analysis. Analytical Methods, 6, 2812-2831.
539 https://doi.org/10.1039/C3AY41907J
540 Cabrera, C., Lloris, F., Giménez R., Olalla, M., López, C. (2003) Mineral content in légumes and
541 nuts : contribution to the Spanish dietary intake. Science of the Total Environment, 308, 1-14.
542 https://doi.org/10.1016/S0048-9697(02)00611-3
543 Camin, F., Boner, M., Bontempo, L., Fauhl-Hassek, C., Kelly, S.D., Riedl, J., Rossmann, A. (2017).
544 Stable isotope techniques for verifying the declared geographical origin of food in legal cases.
545 Trends in Food Science and Technology, 61, 176-187. https://doi.org/10.1016/j.tifs.2016.12.007
546 Dalipi, R., Marguí, E., Borgese, L., Depero, L.E. (2017). Multi-element analysis of vegetal
547 foodstuff by means of low power total reflection X-ray fluorescence (TXRF) spectrometry. Food
548 Chemistry, 218, 348-355. https://doi.org/10.1016/j.foodchem.2016.09.022
549 Danezis, G.P., Tsagkaris, A.S., Camin, F., Brusic, V., Georgious, C.A. (2016).Food authentication:
550 techniques, trends & emerging approaches. Trends in Analytical Chemistry, 85, 123-132.
551 https://doi.org/10.1016/j.trac.2016.02.026
552 Espiñeira, M., Santaclara, F.J., (2016). What is food traceability?. In Advances in M. Espiñeira, F.J.
553 Santaclara (Eds.), Food Traceability Techniques and Technologies (pp 3-8). Elsevier.
554 European Community. Regulation (EC) No 178/2002 laying down the general principles and
555 requirements of food law, establishing the European Food Safety Authority and laying down
556 procedures in matters of food safety. URL (https://eur-lex.europa.eu/legal-
557 content/EN/ALL/?uri=celex%3A32002R0178)
558 Ghanbari A.A., Shakiba, M.R., Toorchi, M., Choukan, R. (2013). Nitrogen changes in the leaves
559 and accumulation of some minerals in the seeds of red, white, and Chitti beans (Phaseolus vulgaris)
560 under water deficit conditions. Australian Journal of Crop Science, 7, 706-712.

561 Haswell, S.J., Walmasley, A.D. (1998). Multivariate data visualization methods based on multi-
562 elemental analysis of wines and coffees using total reflection X-ray fluorescence analysis. Journal
563 of Analytical Atomic Spectroscopy, 13, 131-134. https://doi.org/10.1039/A705317G
564 ISO 22000:2018. Food safety management systems – Requirements for any organization in the food
565 chain.
566 Klockenkämper, R., von Bohlen, A. (2015). Total-reflection X-ray fluorescence analysis and related
567 methods, John Wiley & Sons Inc.
568 Martens, H., Høy, M., Wise, B.M., Bro, R., Brockhoff, P.B. (2003). Pre- whitening of data by
569 covariance- weighted pre- processing. Journal of Chemometrics, 17, 153-165.
570 https://doi.org/10.1002/cem.780
571 Miller, C. E. (2010). Chemometrics in process analytical technology (PAT). In K.A. Bakeev (Ed.),
572 *Process analytical technology: spectroscopic tools and implementation strategies for the chemical*
573 *and pharmaceutical industries* (pp. 353-438) John Wiley & Sons.
574 Nečemer, M., Košir, I.J., Kump, P., Kropf, U., Jamnik, M., Beronrlj, J., Ogrinc, N., Golob, T.
575 (2009). Application of total reflection X-ray spectrometry in combination with chemometric
576 methods for determination of the botanical origin Slovenian honey, Journal of Agricultural and
577 Food Chemistry, 57, 4409-4414. https://doi.org/10.1021/jf900930b
578 Obhođaš, J., Valković, V., Vinković, A., Sudac, D., Čanađija, I., Pensa, T., Fiket, Ž., Turyanskaya,
579 A., Bretschneider, T., Wilhelmer, C., Gunchin, G., Kregsamer, P., Wobreauschek, P., Streli, C.
580 (2021). X-ray fluorescence techniques for element abundance analysis in wine. ACS Omega, 6,
581 22643-22654. https://doi.org/10.1021/acsomega.1c02731
582 Oliveri, P., Malegori, C., Simonetti, R., Casale, M. (2019). The impact of signal pre-processing on
583 the final interpretation of analytical outcomes–A tutorial. Analytica Chimica. Acta, 1058, 9-17.
584 https://doi.org/10.1016/j.aca.2018.10.055

585 Panchuk, V., Yaroshenko, I., Legin, A., Semenov, V., Kirsanov, D. (2018). Application of
586 chemometric methods to XRF-data – a tutorial revew. Analytica Chimica, 1040, 19-32.
587 https://doi.org/10.1016/j.aca.2018.05.023
588 Ramírez-Ojeda, A.M., Moreno-Rojas, R., Cámara-Martos, F. (2018). Mineral and trace element
589 content in legumes (lentis, chickpeas and beans) : bioaccesibility and probabilistic assessment of the
590 dietary intake. Journal of Food Composition and Analysis, 73, 17-28.
591 https://doi.org/10.1016/j.jfca.2018.07.007
592 Robberecht, H., Van Dyck, K., Bosscher, D., Van Cauwenbergh, R. (2008). Silicon in foods:
593 content and bioavailability. International Journal of Food Properties, 11, 638-645.
594 https://doi.org/10.1080/10942910701584252
595 Roberts, J.J., Cozzolino, D. (2013). An overview on the application of chemometrics in food
596 science and technology—an approach to quantitative data analysis. Food Analytical Methods, 9,
597 3258-3267. https://doi.org/10.1007/s12161-016-0574-7
598 Serranti, S., Cesare, D., Marini, F., Bonifazi, G. (2013). Classification of oat and groat kernels using
599 NIR hyperspectral imaging. Talanta, 103, 276-284. https://doi.org/10.1016/j.talanta.2012.10.044
600 Shand, C.A., Wendler, R., Dawson, L., Yates, K., Stephenson, H. (2017). Multivariate analysis of
601 Scotch whisky by total reflection X-ray fluorescence and chemometric methods: a potential tool in
602 the identification of the counterfeits. Analytica Chimica Acta, 976, 14-24.
603 https://doi.org/10.1016/j.aca.2017.04.041
604 Sica, P., Scariolo, F., Galvao, A., Battaglai, D., Nicoletto, C., Maucieri C., Palumbo, F., Franklin,
605 D., Cabrera, M., Borin, M., Sambo, P., Barcaccia, G. (2021). Molecular hallmarks, agronomic
606 performances and seed nutraceutical properties to exploit neglected genetic resources of common
607 beans grown by organic farming in two contrasting environments. Frontiers in Plant Science, 12,
608 Article 674985 https://doi.org/10.3389/fpls.2021.674985

609 Squeo, G., Grassi, S., Paradiso, V. M., Alamprese, C., Caponio, F. (2019). FT-IR extra virgin olive
610 oil classification based on ethyl ester content. Food Control, 102, 149-156.
611 https://doi.org/10.1016/j.foodcont.2019.03.027
612 US Department of Agriculture (2011). Proposed rule to establish general regulations for improving
613 the traceability of U.S. livestock moving interstate when animal disease events take place. URL
614 (http://www.aphis.usda.gov/traceability/)
615 Vitali Čepo, D., Karoglan, M., Borgese, L., Depero, L.E., Marguí, E., Jablan, J. (2022). Application
616 of benchtop total-reflection X-ray fluorescence spectrometry and chemometrics in classification of
617 origin and type of Croatian wines. Food Chemistry X, 13. Article 100209.
618 https://doi.org/10.1016/j.fochx.2022.100209
619 von Bohlen, A., Fernández-Ruiz, R. (2020). Experimental evidence of matrix effects in total-
620 reflection X-ray fluorescence analysis: coke case. Talanta, 209, 120562.
621 https://doi.org/10.1016/j.talanta.2019.120562
622 Westermann, D.T., Tarán, H., Muñoz-Perea, C.G., Singh, S.P. (2011). Plant and seed nutrient
623 uptake in common bean in seven organic and conventional production systems. Canadian Journal of
624 Plant Science, 91, 1089-1099. https://doi.org/10.4141/cjps10114
625 Winkler, A., Rauwolf, M., Sterba, J.H., Wobrauschek, P., Streli, C., Turyanskaya, (2020). A. Total
626 reflection X-ray fluorescence analysis of elemental composition of herbal infusions and teas.
627 Journal of the Science of Food and Agriculture, 100, 4226-4236. https://doi.org/10.1002/jsfa.10463
628
629
630
631
632
633
634

Figure Caption List

Figure 1. Concentration (mean ± SD, *n*=3) of P, S, Cl, K, Ca, Mn, Fe, Ni, Cu, Zn, Br and Rb in the 24 genotypes of beans grown in the soils of Legnaro and Asiago as determined by TXRF analyses.

Figure 2. Loading plot (A) and score plot (B) of the PCA performed on the concentrations of the elements in the studied beans. A (●), Asiago, L (●), Legnaro.

Figure 3. Score plot and loading plot of PCA after GLSW filter (A) and PCA results after SNV coupled with GLSW (B). A (●), Asiago, L (●), Legnaro. The flat line between 9-10.5 keV indicate the region removed characterized by the presence of Ga fluorescence lines.

Figure S1 TXRF spectrum of a sample of bean.

Figure S2. Score plots (PC1 *vs* PC2 to PC5) and loading plot of the PCA carried out on the smoothed spectral dataset. A (●), Asiago, L (●), Legnaro.

Figure S3. Score plots (PC1 vs PC2 to PC5) and loading plot of the PCA carried out on SNV preprocessed TXRF spectra.

Figure S4. Score plots (PC1 vs PC2 to PC5) and loading plot of the PCA carried out on 1D preprocessed TXRF spectra.

Table Caption List

660 **Table 1.** Comparison between TXRF and ICP-OES results obtained after the analysis of the
661 reference material NIST 1567b
662 **Table 2.** Comparison between TXRF and ICP-OES results obtained on five different samples of
663 bean.
664 **Table 3.** Results of beans classification.
665 **Table S1.** List and pictures of the seeds of the 24 genotypes of beans analysed in this study.
666 **Table S2.** Physical and chemical characteristics of Asiago and Legnaro soils.
667 **Table S3.** List of samples in the calibration and validation sets, respectively.

Highlights

- 24 genotypes of bean grown on 2 different sites were studied to trace their origin
- A method for the elemental analysis of legumes with TXRF was developed
- Both quantitative and spectral data were analysed with multivariate methods
- The use of PLS-DA with SNV-GLSW on TXRF spectra allow to classify the beans origin
- TXRF spectra were used as food fingerprint to trace its geographical origin

Declaration of interests

☒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.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Table 1

Table 1. Comparison between TXRF and ICP-OES results obtained after the analysis of the reference material NIST 1567b

Element	Certified values		TXRF						ICP-OES						TXRF/ICP-OES
	Mean	SD	Mean	SD	LOD	LOQ	Recovery	RSD	Mean	SD	LOD	LOQ	Recovery	RSD	
	mg kg^{-1}						%		mg kg^{-1}				%		
P	1333	36	1060	20	24	80	80	2	1247	80	2	7	94	6	0.85
S	1645	25	1480	30	13	44	90	2	-	-	-	-	-	-	-
Cl	565	8.6	494	39	7	21	87	8	-	-	-	-	-	-	-
K	1325	20	1185	78	2	8	89	7	1325	141	9	30	100	11	0.89
Ca	191	3.3	222	4	2	6	116	2	153	11	15	51	80	7	1.45
Mn	9	0.78	10.2	0.3	0.4	1.4	113	3	9.5	0.6	0.6	2.1	106	6	1.07
Fe	14.1	0.33	16.1	4.7	0.3	1.0	114	29	18.0	1.3	2.0	6.0	128	7	0.89
Cu	2.03	0.14	2.2	0.1	0.2	0.6	106	7	2.4	0.1	0.7	2.2	118	4	0.90
Zn	11.61	0.26	11.0	0.2	0.2	0.4	95	2	11.0	1.1	1.6	5.4	95	10	1.00
Br	6.45	0.97	5.9	0.2	0.1	0.4	91	3	-	-	-	-	-	-	-
Rb	0.671	0.01	0.4	0.1	0.1	0.5	52	15	-	-	-	-	-	-	-

SD: standard deviation (n=3)

RSD : relative standard deviation

Table 2

Table 2. Comparison between TXRF and ICP-OES results obtained on five different samples of bean.

Element	Bean 1				Bean 2				Bean 3				Bean 4				Bean 5				TXRF/ICP-OES					
	TXRF		ICP-OES		TXRF		ICP-OES		TXRF		ICP-OES		TXRF		ICP-OES		TXRF		ICP-OES		Bean	Bean	Bean	Bean	Bean	Mean
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	1	2	3	4	5	
	mg/kg																									
P	4557	270	5315	33	5401	300	6368	329	5041	114	5342	210	5059	116	6918	288	4042	147	4575	218	0,86	0,85	0,94	0,73	0,88	0,85
K	17717	209	19790	2699	18816	259	22131	840	14106	378	16248	835	17615	527	23419	572	15714	97	17091	2077	0,90	0,85	0,87	0,75	0,92	0,86
Ca	674	45	526	49	1308	60	997	32	834	24	728	64	2517	347	2067	43	720	7	550	100	1,28	1,31	1,15	1,22	1,31	1,25
Mn	14,9	0,3	14,5	0,2	16,6	0,3	15,9	0,8	12,9	0,2	12,4	0,5	24,2	0,7	23,6	1,3	13,7	0,2	13,2	0,6	1,03	1,04	1,04	1,02	1,04	1,03
Ni	2,2	0,1	<LOD	-	2,3	0,2	<LOD	-	0,8	0,1	<LOD	-	2,9	0,1	<LOD	-	2,2	0,2	<LOD	-	-	-	-	-	-	-
Fe	62,7	0,8	63,8	2,4	66,7	0,8	66,1	4,5	56,9	1,1	60,4	2,1	74,6	2,2	75,1	4,7	49,7	0,6	54,7	3,9	0,98	1,01	0,94	0,99	0,91	0,97
Cu	10,0	0,2	10,2	0,3	12,9	0,9	13,4	0,8	9,0	0,3	9,5	0,8	9,4	0,2	10,4	0,5	8,3	0,1	8,5	0,5	0,98	0,97	0,95	0,90	0,97	0,95
Zn	34,9	0,3	40,8	0,9	45,6	0,2	51,3	3,1	30,0	0,4	33,1	1,8	41,4	0,9	47,1	0,6	29,1	0,2	31,2	1,4	0,86	0,89	0,91	0,88	0,93	0,89

SD: standard deviation (n=3)

RSD : relative standard deviation

Table3

Table 3. Results of beans classification.

Classification based on the elemental composition.													
			Calibration			Cross-validation			Prediction				
LDA	CVs	Class	Sn	Sp	Error	Sn	Sp	Error	Sn	Sp	Error	N	Misclassified in prediction
Set 1	1	A	1	1	0	1	1	0	0.75	0.94	0.16	12	3
		L	1	1	0	1	1	0	0.94	0.75	0.16	33	2
Set 2	1	A	1	1	0	0.96	0.96	0.04	1	1	0	24	0
		L	1	1	0	0.96	0.96	0.04	1	1	0	21	0
Set 3	1	A	1	1	0	0.95	0.92	0.06	1	1	0	33	0
		L	1	1	0	0.92	0.95	0.06	1	1	0	12	0
PCA-LDA	PCs-CVs												
Set 1	6-1	A	1	1	0	1	1	0	1	1	0	12	0
		L	1	1	0	1	1	0	1	1	0	33	0
Set 2	5-1	A	1	1	0	1	1	0	1	1	0	24	0
		L	1	1	0	1	1	0	1	1	0	21	0
Set 3	2-1	A	1	1	0	1	1	0	1	0.75	0.13	33	0
		L	1	1	0	1	1	0	0.75	1	0.13	12	3
PLS-DA	LVs												
Set 1	5	A	1	1	0	1	1	0	0.83	1	0.09	12	2
		L	1	1	0	1	1	0	1	0.83	0.09	33	0
Set 2	2	A	1	1	0	1	1	0	1	1	0	24	0
		L	1	1	0	1	1	0	1	1	0	21	0
Set 3	1	A	1	1	0	1	0.98	0.01	1	0.83	0.09	33	0
		L	1	1	0	0.98	1	0.01	0.83	1	0.09	12	2
Classification based on the TXRF spectra.													
GLSW	LVs												
Set 1	2	A	1	1	0	1	0.87	0.06	1	0.91	0.05	12	0
		L	1	1	0	0.87	1	0.06	0.91	1	0.05	33	3
Set 2	2	A	1	1	0	0.94	0.98	0.04	1	1	0	24	0
		L	1	1	0	0.98	0.94	0.04	1	1	0	21	0
Set 3	2	A	1	1	0	0.92	0.97	0.06	1	1	0	33	0
		L	1	1	0	0.97	0.92	0.06	1	1	0	12	0
SNV-GLSW	LVs												
Set 1	4	A	1	1	0	1	1	0	1	0.97	0.02	12	0
		L	1	1	0	1	1	0	0.97	1	0.02	33	1
Set 2	1	A	1	1	0	1	1	0	1	1	0	24	0
		L	1	1	0	1	1	0	1	1	0	21	0
Set 3	5	A	1	1	0	0.97	0.95	0.04	1	1	0	33	0
		L	1	1	0	0.95	0.97	0.04	1	1	0	12	0

Sn, sensitivity; Sp, specificity; Error, classification error; N, number of samples in the test sets, including replicates, per each class; CVs, number of canonical variables; PCs, number of principal components; LVs, number of latent variables; A, Asiago; L, Legnaro.

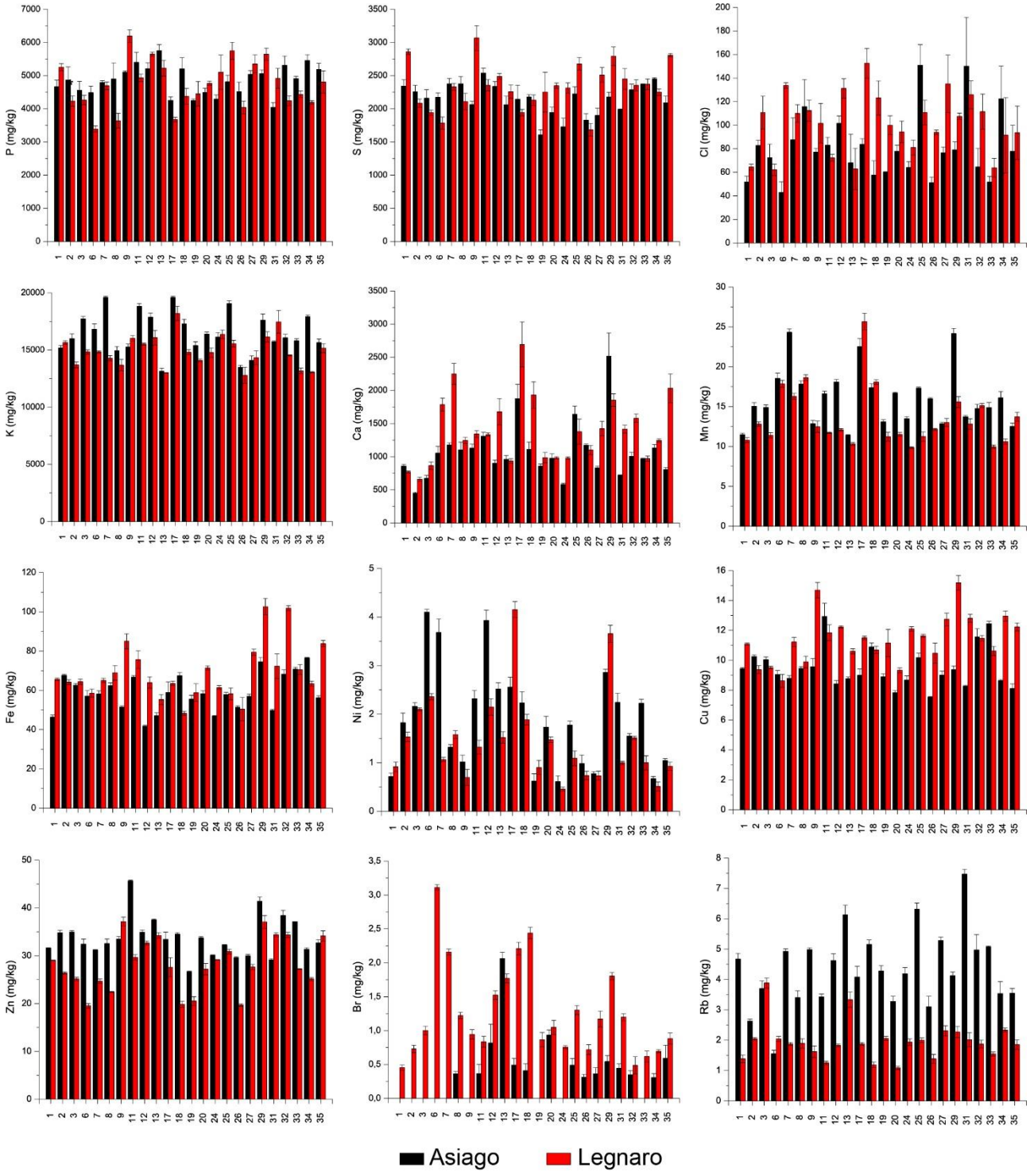

Figure 1. Concentration (mean ± SD, *n*=3) of P, S, Cl, K, Ca, Mn, Fe, Ni, Cu, Zn, Br and Rb in the 24 genotypes of beans grown in the soils of Legnaro and Asiago as determined by TXRF analyses.

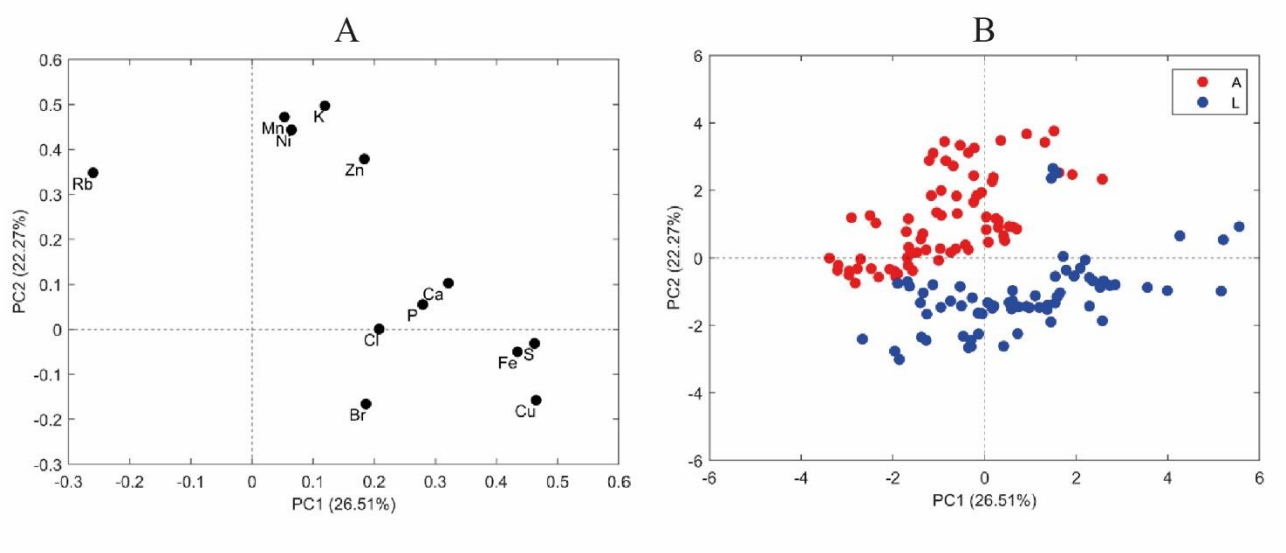

Figure 2. Loading plot (A) and score plot (B) of the PCA performed on the concentrations of the elements in the studied beans. A (●), Asiago, L (●), Legnaro.

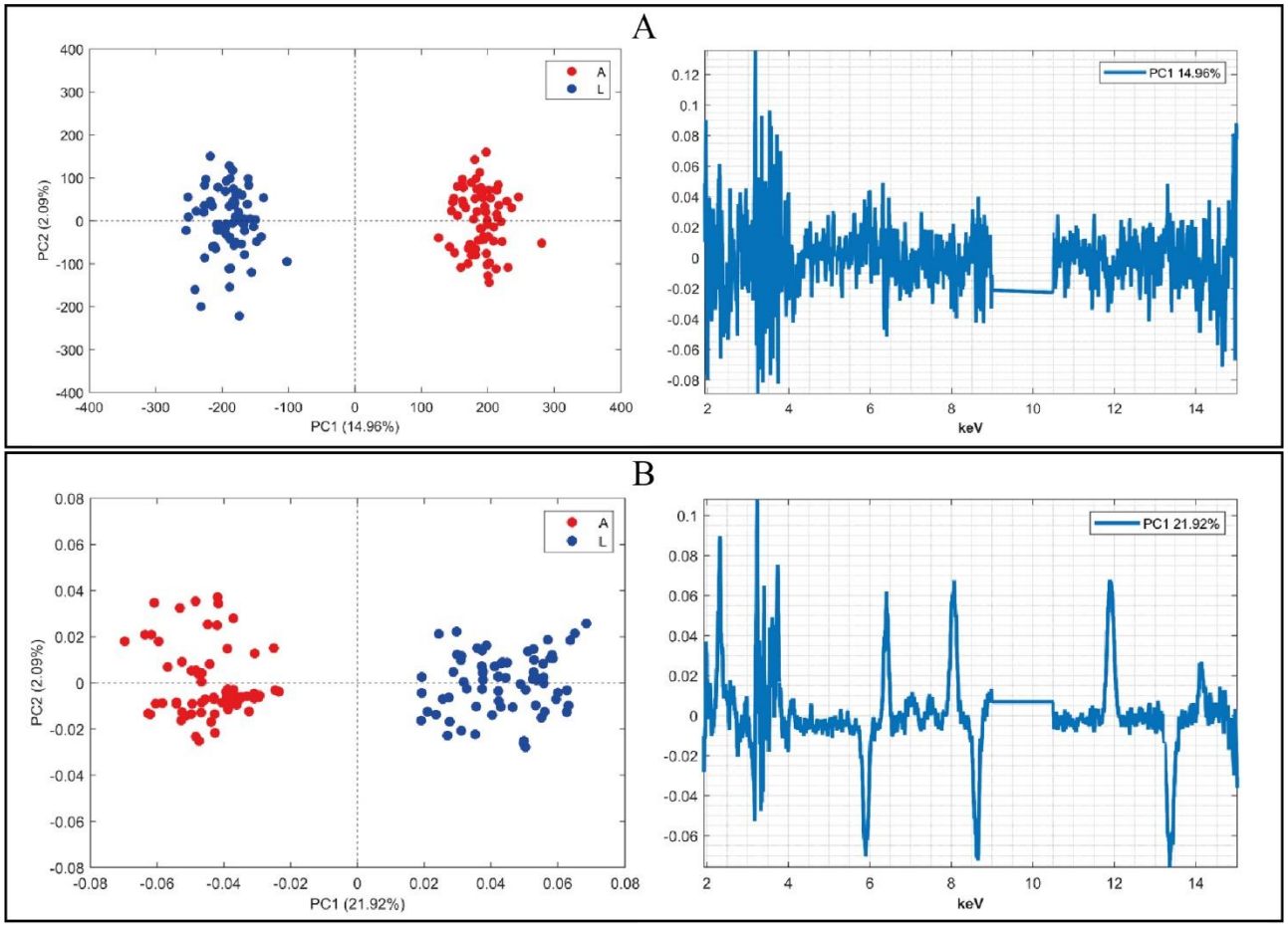

Figure 3. Score plot and loading plot of PCA after GLSW filter (A) and PCA results after SNV coupled with GLSW (B). A (●), Asiago, L (●), Legnaro.

Credit Author Statement

Ignazio Allegretta: Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Data Curation, Writing - Original Draft, Writing – Review and Editing, Visualization, Supervision, Project Administration, Funding acquisition. **Giacomo Squeo:** Methodology, Software, Validation, Formal Analysis, Data Curation, Writing - Original Draft, Writing – Review and Editing, Visualization, Supervision, Funding acquisition. **Concetta Eliana Gattullo:** Investigation, Data Curation, Writing - Original Draft, Writing – Review and Editing, Visualization, Supervision. **Carlo Porfido:** Investigation, Writing – Review and Editing. **Antonio Cicchetti:** Investigation, Data Curation. **Francesco Caponio:** Resources, Writing – Review and Editing, Supervision, Funding acquisition. **Stefano Cesco:** Conceptualization, Resources, Writing – Review and Editing, Supervision. **Carlo Nicoletto:** Conceptualization, Resources, Writing – Review and Editing, Supervision, Funding acquisition. **Roberto Terzano:** Conceptualization, Resources, Writing – Review and Editing, Supervision, Project administration, Funding acquisition.

Table S1. List and pictures of the seeds of the 24 genotypes of beans analysed in this study.

Id	Name	Type	Growth		Id	Name	Type	Growth	
1	Mangiatutto rampicante	Italian pre-commercial	Indeterminate (climbing)		19	Tondino abruzzese	Italian pre-commercial	Indeterminate (climbing)	
2	Borlotto nano A	Italian elite line	Determined (Dwarf)		20	Verdone del piave	Italian elite line	Determined (Dwarf)	
3	Borlotto nano B	Italian elite line	Determined (Dwarf)		24	Posenati	Venetian landrace	Indeterminate (climbing)	
6	Fagiolo nano creso	Italian elite line	Determined (Dwarf)		25	Semi-rampicante abruzzese	Venetian landrace	Indeterminate (climbing)	
7	Bblue lake sel. Gia	Italian pre-commercial	Indeterminate (climbing)		26	Fasol dela nonna	Venetian landrace	Indeterminate (climbing)	
8	Anellino di Trento	Italian elite line	Determined (Dwarf)		27	Maseleta rossa	Venetian landrace	Indeterminate (climbing)	
9	Anellino giallo	Italian pre-commercial	Indeterminate (climbing)		29	Meraviglia di Venezia	Venetian landrace	Indeterminate (climbing)	
11	Bortollo lingua di fuoco 3	Italian pre-commercial	Indeterminate (climbing)		31	Della Clorinda	Venetian landrace	Indeterminate (climbing)	
12	Blue lake a grano nero	Italian pre-commercial	Indeterminate (climbing)		32	Pegaso	Venetian landrace	Indeterminate (climbing)	
13	Dolico	Italian elite line	Determined (Dwarf)		33	SC-iosela	Venetian landrace	Indeterminate (climbing)	
17	Fagiolo nano valdarno	Italian elite line	Determined (Dwarf)		34	D'oro (val di fiemme)	Venetian landrace	Indeterminate (climbing)	
18	Coco nain blanc precoce	Italian elite line	Determined (Dwarf)		35	Monachelle	Venetian landrace	Indeterminate (climbing)	

Table S2. Physical and chemical characteristics of Asiago and Legnaro soils.

	Asiago	Legnaro
Sand (%)	15	15
Silt (%)	45	65
Clay (%)	40	20
$pH_{(H2O)}$	6.77	8.15
OM (%)	2.6	1.8
N_{tot} (%)	0.1	0.1
C/N	15.11	9.72
P_2O_5 (%)	0.1	0.1
P_{Olsen} (mg P_2O_5 kg^{-1})	9.2	8.5
Na_2O (%)	0.3	0.8
K_2O (%)	1.4	2.5
MgO (%)	1.6	5.3
CaO (%)	1.4	1.0
SO_3 (%)	0.2	0.1
MnO (%)	0.3	0.1
Fe_2O_3 (%)	5.5	3.9
Ni (mg kg^{-1})	49	48
Cu (mg kg^{-1})	50	39
Zn (mg kg^{-1})	105	62
Rb (mg kg^{-1})	66	86
Sr (mg kg^{-1})	53	88
Cl (mg kg^{-1})	76	128
Exchangeable bases		
Na^+ (mg kg^{-1})	18	26
K^+ (mg kg^{-1})	67	60
Mg^{2+} (mg kg^{-1})	201	247
Ca^{2+} (mg kg^{-1})	1.9	2.6
Available elements		
Mn (mg kg^{-1})	185	1.7
Fe (mg kg^{-1})	66	6.7
Ni (mg kg^{-1})	1.6	-
Cu (mg kg^{-1})	6.1	4.2
Zn (mg kg^{-1})	5.9	0.7

Table S3. List of samples in the calibration and validation sets, respectively.

Set 1		Set 2		Set 3	
Calibration	Validation	Calibration	Validation	Calibration	Validation
1A	6A	1L	1A	1L	1A
1L	7L	2A	3A	2A	3L
2A	8A	2L	7A	2L	8A
2L	9L	3L	8L	3A	9A
3A	11L	6A	9L	6A	9L
3L	13L	6L	11A	6L	13A
6L	17L	7L	18L	7A	17A
7A	18L	8A	19L	7L	18A
8L	24L	9A	24L	8L	25A
9A	27A	11L	25L	11A	25L
11A	27L	12A	32A	11L	26A
12A	29A	12L	33A	12A	27A
12L	29L	13A	34A	12L	29A
13A	34L	13L	34L	13L	29L
17A	35L	17A	35A	17L	35A
18A		17L		18L	
19A		18A		19A	
19L		19A		19L	
20A		20A		20A	
20L		20L		20L	
24A		24A		24A	
25A		25A		24L	
25L		26A		26L	
26A		26L		27L	
26L		27A		31A	
31A		27L		31L	
31L		29A		32A	
32A		29L		32L	
32L		31A		33A	
33A		31L		33L	
33L		32L		34A	
34A		33L		34L	
35A		35L		35L	

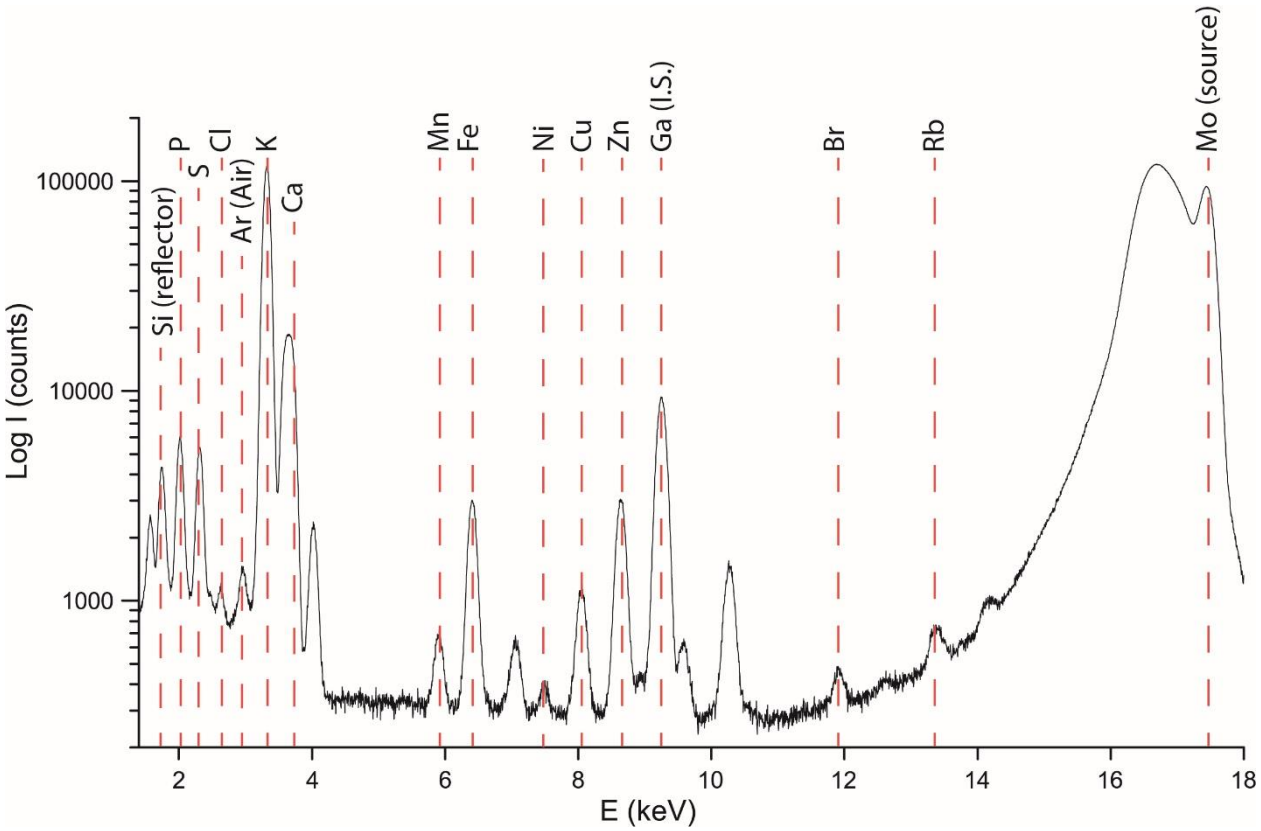

Figure S1 TXRF spectrum of a sample of bean

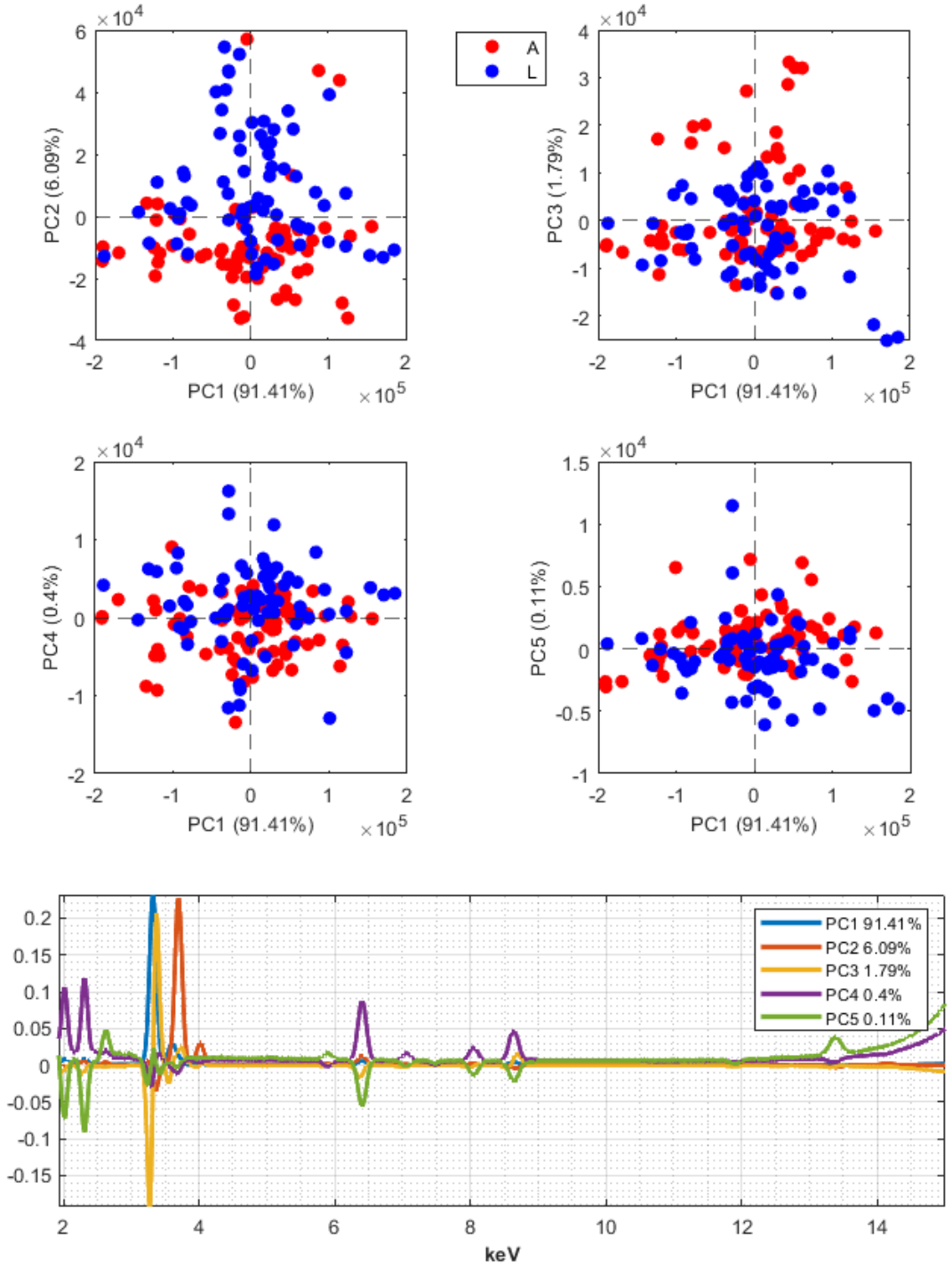

Figure S2. Score plots (PC1 *vs* PC2 to PC5) and loading plot of the PCA carried out on the smoothed spectral dataset. A (●), Asiago, L (●), Legnaro.

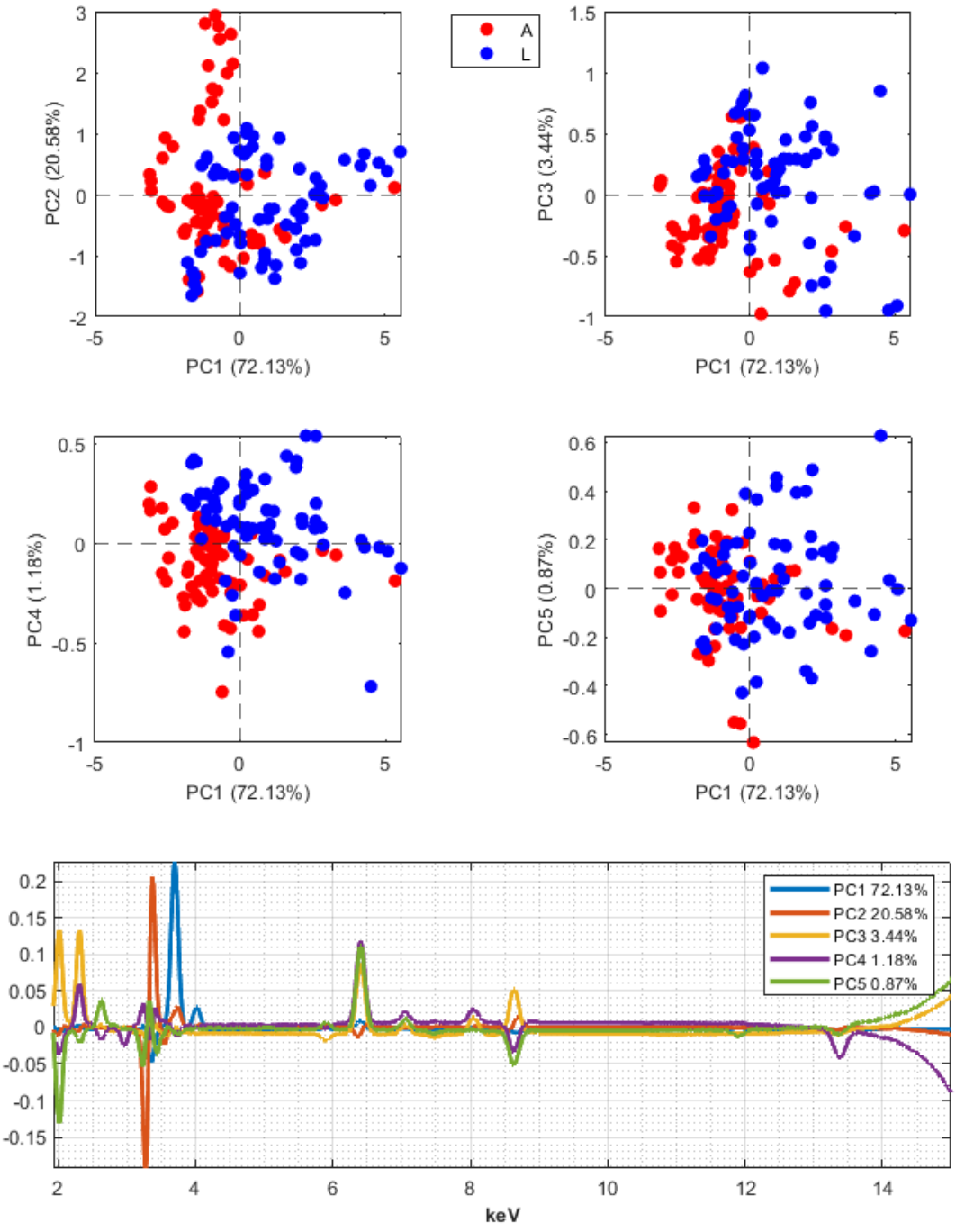

Figure S3. Score plots (PC1 *vs* PC2 to PC5) and loading plot of the PCA carried out on SNV preprocessed TXRF spectra.

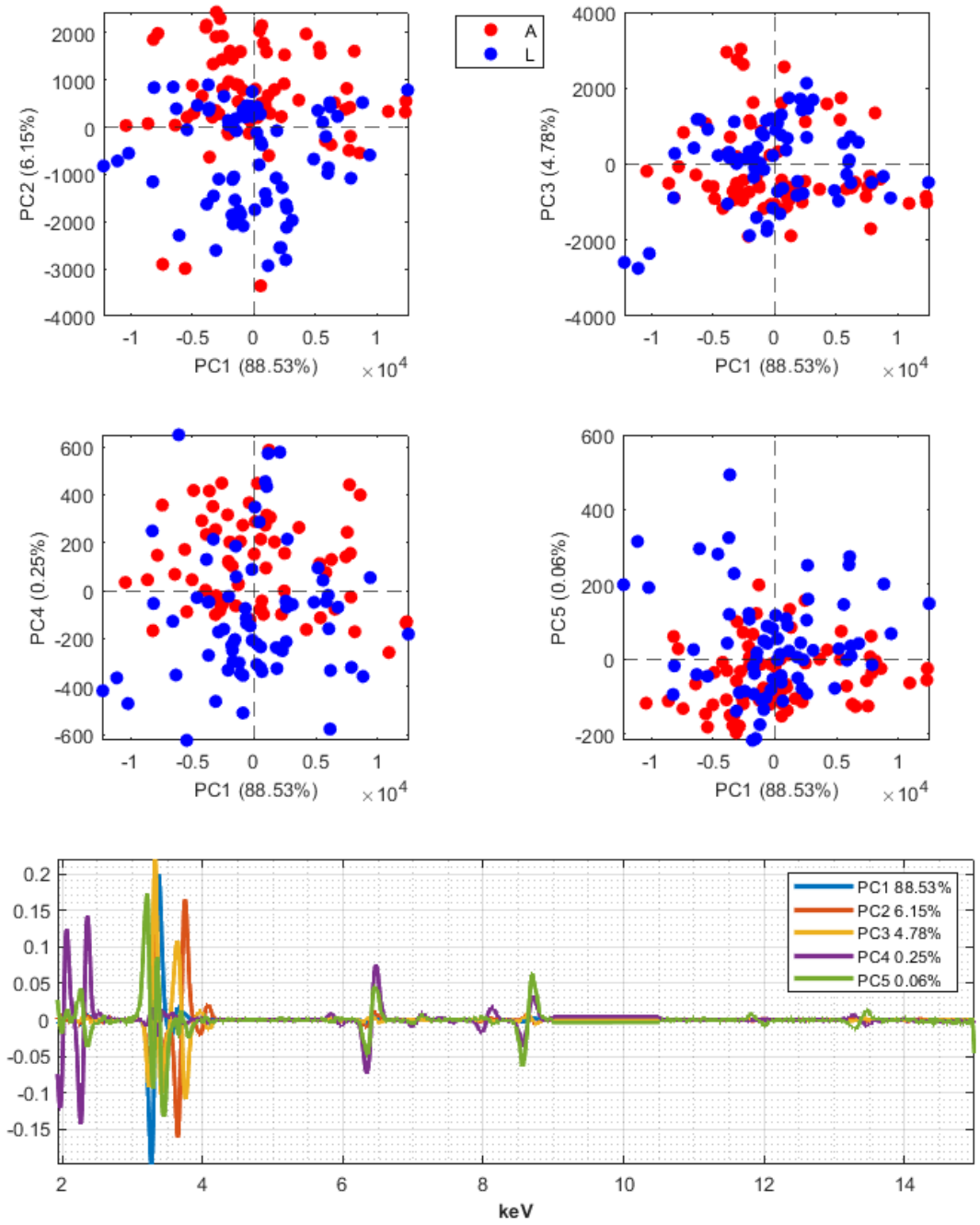

Figure S4. Score plots (PC1 *vs* PC2 to PC5) and loading plot of the PCA carried out on 1D preprocessed TXRF spectra.

Graphical Abstract

Click here to access/download;Supplementary Material;Graphical Abstract.docx