# Assessment of residual plant DNA in bulk milk for Grana Padano PDO production by a metabarcoding approach

**Plant residues identification in cow milk**

#### 

#### Pozzi1¶\* A., Nazzicari2 N., Capoferri1¶ R., Radovic3 S. and Bongioni1 G.

1 Istituto Sperimentale Lazzaro Spallanzani, Localita’ La Quercia, 26027 Rivolta d’Adda (CR), Italy

2 CREA - Council for Agricultural Research and Analysis of Agricultural Economics. Research Centre for Animal Production and Aquaculture, Viale Piacenza, 29 26900 Lodi, Italy.

3 IGA Technology Service s.r.l., Via Linussio, 51 33100 Udine, Italy.

\* Corresponding author

E-mail: [anna.pozzi@istitutospallanzani.it](mailto:anna.pozzi@istitutospallanzani.it)

¶These authors contributed equally to this work.

# Abstract

The aim of this study was to evaluate the ability of DNA metabarcoding, by *rbcl* as barcode marker, to identify and classify the small traces of plant DNA isolated from raw milk used to produce one of the most popular Italian PDO (Protected Designation of Origin): Grana Padano (GP) cheese. The label PDO is assigned to food produced, processed and prepared in a defined geographical area by using a dedicated know-how, and more specifically in this case, in accordance with the GP PDO Specification rules. A total of 42 GP bulk tank milk samples were collected from 14 dairies belonging to Grana Padano Protection Consortium and 4 commercial milk samples retrieved outside the GP production site. For the taxonomic classification, a local database with the *rbcL* sequences available in NCBI on September 2020/March 2021 for the Italian flora was generated. A total of 10,547,798 reads were produced with an average of 234,395 per sample (range 37,002-1,464,302) resulting in 16, 31 and 28 dominant OTUs at family, genus and species level, respectively.

Despite the samples came from different provinces on the edge of the GP production area, the taxonomic analysis of plant species in the GP milk samples identified 7 families, 14 genera and 14 species, the statistical analysis conducted using alpha and beta diversity approaches, did not discriminate the samples neither by type nor by province of origin. However, the milk samples are featured by a high plant variability and the lack of differences at multiple taxonomic levels could be due to the standardisation of the feed rationing, as requested by the GP rules, in order to guarantee the final product quality. The results suggest that DNA metabarcoding is a valuable resource to explore plant DNA traces in a complex matrix such as milk.

# Introduction

Current market trends show a growing interest among consumers in the origin of food products [1-3]. In this regard, PDO (Protected Designation of Origin) and PGI (Protected Geographical Indication) products, above others, are able to guarantee authenticity, tradition, taste, typicality, link with the land of origin, safety and traceability [4-5]. In details, the PDO certification ensures that the product is legally guaranteed by the European Union as authentic, or manufactured in a specific geographical area with the expected ingredients [6-7]. An accurate control system is included in the certification itself (public supervision, third-party control and self-control). However, despite this safeguard it is still possible to encounter copycats of PDO products. This phenomenon has led to the increase of sophisticated technologies, which are able to characterise the products in defence of their authenticity.

Grana Padano (GP) is one of the most popular Italian PDOs, both in Italy and internationally. In 2020, the GP production reached the quota record of 5,255,443 cheese wheels. With a total production value of 4.5 billion Euros, GP represents 59% of the entire Italian PDO-PGI food production, and 52.6% of PDO-PGI food export [8]. GP is typically produced from raw cow’s milk collected from animals bred in 32 provinces of Northern Italy associated with the Grana Padano Protection Consortium which operates in compliance with the rules defined by the product regulations. As such, the quality of the GP cheese is closely related to the typical climatic and environmental conditions, and forage quality. In fact, at least 75% of the fodder’s dry matter in the daily ration must be grown in the cheese production area [9-10].

In support of certified productions, the EU recognizes the importance of food traceability, *i.e.* the capacity to track food in all stages of the production chain [11]. Food traceability becomes even more important when the area of production influences the quality of the final product. In this context, recent advances in molecular biology make DNA markers potentially valuable tools to track the raw material in the food chain. In fact, DNA has higher stability compared to other biological markers, such as proteins, and can be isolated even in matrices with extremely low contents [12-14]. In the specific case of GP, different approaches have been described for discriminating PDO and non-PDO cheese. For example, Rocchetti et al. (2018) [15] outlined a metabolomic approach to ensure the authenticity of GP, while Faustini et al. (2019) [16] reported on a volatilome study on milk destined to the production of both GP and Parmigiano Reggiano cheeses. In addition, metabarcoding of bacterial communities has been used for the preservation and valorisation of similar food products, [17-19]. However, given the continuous pressure exerted by increasingly shrewd counterfeiting professionals, the development of efficient and less invasive traceability methods is necessary. In this, several studies have shown that feed-derived plant nuclear and/or chloroplast DNA fragments can also be detected in milk [20]. In fact, small fragments of plant DNA are able to cross the intestinal barrier and enter the bovine bloodstream. However, this DNA is highly fragmented, although it is not possible to establish at what level, as the feed is subjected to the production processes of milling, extrusion, ensiling, grinding, steam heating and pelletizing and, after ingestion, to its elaboration/degradation in the gastrointestinal tract [21-22]. In these cases, where the DNA is subjected to severe treatments and is present in small quantities, its isolation is particularly difficult and, in the absence of commercial kits optimised for this purpose, it is necessary to implement appropriate molecular approaches based on the analysis of very short DNA fragments. Although some studies were conducted for plant DNA detection in milk using classical approaches [19-23], to the best of our knowledge there are no published data about the detection of residual plant DNA in cow’s milk by High-Throughput DNA Sequencing (HTS) metabarcoding.

Therefore, the aim of this work was to investigate whether chloroplast DNA metabarcoding and dedicated data analysis could be a suitable tool to identify dietary DNA in raw milk for GP production in order to ensure the traceability of this high-quality product and meet consumer demand.

# Materials and methods

## Samples collection

A number of 42 GP bulk tank milk samples of 50 ml volumes were collected from cows fed according to PDO Specification rules (9) in 14 dairies belonging to GP Consortium in the provinces of Verona, Piacenza, Trento, Vicenza, Brescia, Cremona and Cuneo, and stored at -20°C until analysis. Additionally, four commercial milk samples were purchased from retail outside the GP production site in Sicily, Apulia, Naples and Vilminore di Scalve (BG).

One sample of Soy flour (*Glycine max*) and two feedstuffs (Feed\_1 and Feed\_2) allowed by GP PDO production specification rules were also collected as technical controls.

## DNA isolation

Milk samples were thawed at 4°C for 16 hours and total DNA was extracted from 20 ml in phenol/chloroform according to Ponzoni et al. (2009)[20]. As a minor modification of the original method, the first incubation step was conducted at 65°C for 6 hours under constant agitation.

Plant DNA was extracted from technical controls using the DNeasy Plant mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol.

Purity and quantity were determined with Nanodrop Spectrophotometer 2000 (Thermofisher Scientific, Wilmington, DE, USA).

## RuBisCO gene amplicon molecular identification

To confirm the presence of plant nucleic acid in the total DNA extracted from milk, a 351bp RuBisCO gene fragment was amplified with the following primers: RUB F2 TTGGCAGCATTCCGAGTAAC and RUB R2 GTGAGGCGGACCTTGGAAAG [24] by using the GeneAmp PCR System 9700 (Applied Biosystems FosterCity, CA). The PCR condition included a denaturation step at 95 °C for 15’, followed by 35 cycles of 94 °C for 30”, 57 °C for 90’ and 72 °C for 90’ followed by a final elongation at 72 °C for 10’. Fragments were visualised on the Bioanalyzer (Agilent Technologies, Waldbronn, Germany) and only samples with positive amplification were submitted to DNA metabarcoding analysis.

## Library preparation and sequencing

DNA samples were amplified in two steps, as follows: an initial PCR amplification of the variable *rbcL* region of the RuBisCO gene was conducted using locus specific primers with the following Illumina adapter overhangs:

**rbcL F** 5’- TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATGTCACCACAAACAGAGACTAAAGC-3’ and **rbcL R** 5’- GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTAAAATCAAGTCCACCRCG -3’

and a subsequent amplification that integrates relevant flow-cell binding domains and unique indices (NexteraXT Index Kit, FC‐131‐1001/FC‐131‐1002). Libraries were sequenced on MiSeq instrument (Illumina, San Diego, CA) using the 300 bp paired-end mode.

## Assembly of a local database and metabarcoding analysis

For the taxonomic classification, a local-database of the potential dietary species was created by using 1426 *rbcl* sequences downloaded from NCBI Nucleotide section. Species admitted by the GP PDO Production Specification Rules (clause 4) were also included.

The analysis pipeline was performed as briefly described, as follows: raw reads were trimmed in order to eliminate primer sequences by *cutadapt* (Ref1) with the following parameters: *--anywhere* (on both adapter sequences) *--overlap 5 --times 2 --minimum-length 35 --mask-adapter*. Low-quality bases were removed from 3’ by erne-filter (Ref2) by applying default parameters, excluding reads <60 bp from the further analysis, and reads with an error rate >1% were removed. Chimeric sequences were then removed with an uchime\_denovo (Ref3) implemented by usearch (https://www.drive5.com/usearch/), using default parameters.

Reads were clustered to a minimum identity of 97% generating representative sequences (cluster\_fast implemented in usearch) and blasted against the local-database without a minimum identity filter. At this point, all the possible blast hits were analysed and only the lowest unambiguous taxonomy was reported, *i.e.* if there are best hits with the same score indicating different lineage, the most common part is indicated. The information was then compacted in several Operational Taxonomic Unit (OTU) tables, depending on the considered taxonomic level, either family, genus or species.

## Data filtering

A further filter was applied to each resulting OTU table to remove entries having less than 5 reads and binning them to unassigned (UN). Out of the assigned reads for each sample, entries representing less than 1% of the reads of a sample were labelled as “subdominant” and were considered as less important in the analysis [17] on metabarcoding of GP cheese.

## Statistical analysis

Species richness was investigated through the analysis of rarefaction curves to verify that the reading depth was adequate. The function *rarecurve* from R package *vegan* [25] with parameter *step = 100* was used. A portion of obtained sequences did not align on the *rbcL* database of interest. Clusters not aligning on the *rbcL* database were aligned via blast on the *Bos Taurus* 3.1.1 genome via the NCBI online nucleotide Blast tool (<https://blast.ncbi.nlm.nih.gov/>). Alignment results were filtered for *percent\_identity* > 95% and *e-value* < 1e-5. In case of multiple alignments, the one with the smallest *e-value* was retained.

In this study each sample was considered as an isolated ecosystem, thus allowing the use of statistical tools developed in the ecological field such as alpha (Shannon and Simpson) and beta diversity analysis [26,27,28,29]. These two methodological approaches aim to provide objective and comparable measures of biological diversity present in each community. All the considered diversity statistics were computed using the R package *vegan* [25].

# Results

## DNA metabarcoding analysis

Regarding the three technical controls Soy flour, Feed\_1 and Feed\_2, a total of 700,346 paired-end-mode reads (range 217,739-261,868) were obtained with 99% of reads assigned to plant Kingdom. Rarefaction curves showed that samples reached stability in the species richness given the amount of available reads (S1 Fig.).

Results relatively to milk samples are shown in Table 1. Overall, a total of 10,547,798 reads were produced with an average of 234,395 per sample (range 37,002-1,464,302). Only one sample showed the extremely low number of reads equal to 31 and was, therefore, discarded from subsequent analyses. More specifically, for GP milk samples 1-36.6% of reads were assigned to plant Kingdom. Concerning the four commercial milk samples, percentages of assignment to plant Kingdom of 6-10% and 1.8% were reported for 3 (Apulia, Naples and Val di Scalve) and Sicily samples, respectively.

Reads not assigned to plant Kingdom were aligned to the *Bos taurus* 3.1.1 genome. On average, 63% of them were mapped, confirming the bovine origin.

**Table 1. Total number of reads and assigned reads and relative percentages produced in technical controls, GP milk and commercial milk.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sample Type** | **N\_sample** | **Tot\_Reads** | **Tot\_Reads\_ASS** | **%Eukaryota** | **%UN\_Eukaryota** |
| Techn\_ctr | 3 | 700,346 | 696,264 | 99.2 | 0.73 |
| GP milk | 42 | 8,399,591 | 260,784 | 7.21 | 92.78 |
| Comm\_Milk | 4 | 2,148,207 | 28,027 | 7.00 | 93.00 |

The total sequences obtained were clustered around 4811 centroids and subsequently aligned on the local-database of 1426 *rbcl* sequences. Samples resulted in 16, 31 and 28 dominant OTUs at family, genus and species level, respectively (Table 2).

**Table 2. Number of raw OTUs, OTUs with reads > 5 and dominant OTUs at family, genus and species level in technical controls, GP milk and commercial samples.**

|  | **Technical controls (n=3)** | | | **GP milk samples (n=41)** | | | **Commercial samples (n=4)** | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Taxonomical**  **level** | **Raw OTUs** | **Otus reads >5** | **Dominant OTU** | **Raw OTUs** | **Otus reads >5** | **Dominant**  **OTU** | **Raw OTUs** | **Otus reads >5** | **Dominant OTU** |
| Phylum | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Family | 11 | 11 | 4 | 10 | 8 | 7 | 9 | 5 | 5 |
| Genus | 30 | 27 | 9 | 25 | 16 | 14 | 16 | 9 | 8 |
| Species | 44 | 34 | 9 | 36 | 16 | 14 | 16 | 8 | 5 |

## Taxonomic analysis

The taxonomic analysis for the three technical control samples Soy flour, Feed\_1 and Feed\_2 showed that the most abundant family was the Poaceae and Fabaceae in Feed\_1 (69.3% and 24.4%) and Feed\_2 (74.8% and 7.3%), respectively, while for the Soy flour sample the almost only representative family was confirmed to be the Fabaceae (99.5%), as expected. The different levels of species composition for Feed\_1, Feed\_2 and  Soy flour were shown in  Figs 1, 2 and 3.

Fig 1. Taxonomic contents of the technical control Feed\_1.

Fig 2. Taxonomic contents of the technical control Feed\_2.

Fig 3. Taxonomic contents of the technical control Soy flour.

In the milk samples, the metabarcoding analysis associated all the assigned reads to the Streptophyta Phylum. Analysis of deeper taxa identified 7 families (*Fabaceae, Convolvulaceae, Poaceae, Rubiaceae, Malvaceae, Asteraceae and Ranunculaceae*), 14genera (*Ceratonia, Cuscuta, Glycine, Triticum, Gallium, Gossypium, Medicago, Panicum, Lolium, Helianthus, Pisum, Zea, Delphinium and Vicia*) and 14 plant species (*Ceratonia siliqua, Triticum aestivum, Medicago sativa, Gossypium mustelinum, Glycine max, Panicum pygmaeum, Panicum wiehei, Lolium perenne, Pisum sativum, Helianthus annuus, Vicia faba, Zea mays, Delphinium grandiflorum and Secale cereale*). Among the sevenfamilies identified, the most abundant were: Fabaceae (55.9%), Convolvulaceae (26.6%) and Poaceae (9.4%), followed by Rubiaceae (4.3%), Malvaceae (3.4%), Asteraceae (0.2%) and Ranunculaceae (0.1%). At a lower taxonomic level, the most representative genera were Ceratonia (33.9%), Cuscuta (26.9%), Glycine (21.2%) and Triticum (5.8%)**.**

Concerning the commercial milk samples, at the Phylum level all the assigned reads were identified as Streptophyta. With regard to analysys at deeper taxa, a minor number of families and genera were found compared to GP milk samples. In details, the five families were: *Fabaceae* (70.4%),*Convolvulaceae* (20.6%),*Poaceae* (5.0%)*, Rubiaceae* (2.7%) *and Malvaceae* (1.3%)*,* whereas the 8 genera were *Ceratonia* (40.3%)*, Cuscuta* (20.6%), *Glycine* (28.3%), *Triticum* (1.6%), *Gallium* (2.7%), *Gossypium* (1.3%), *Medicago* (1.7%) *and Panicum* (3.4%)and 5 plant species S1, S2 and S3 Tables.

In S2, S3 and S4 Figs., with regard to the abundance, are shown the heatmaps for dominant (>1%) centroids at different taxonomic levels: species, genus and family, respectively.

## Statistical analysis

Alpha diversity analysis was performed using the “Shannon-Wiener” and “Simpson” diversity indexes; in Table 3 are reported the p-values from the Wilcoxon rank sum test.

**Table 3. Average values for Shannon and Simpson richness indexes for all samples, computed at three different taxa levels. Significance test p-values computed via Wilcoxon rank sum test.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | **Taxa level** | | |
| **Richness index** | **Area** | **Family** | **Genera** | **Species** |
| Shannon | GP | 0.684 | 1.038 | 0.46 |
| outside | 0.326 | 0.867 | 0.311 |
| p-value | 0.108 | 0.128 | 0.322 |
| Simpson | GP | 0.35 | 0.557 | 0.245 |
| outside | 0.16 | 0.494 | 0.155 |
| p-value | 0.083 | 0.267 | 0.285 |

Results revealed no statistically significant differences for both indexes at three taxonomic levels (family, genus and species), with all p-values ≥ 0.05. The Shannon index shows a clearly non-normal distribution (Fig. 4). In fact, it appears to be bimodal, *i.e.* as if it was the union of two distributions, one with lower and one with higher average diversity. This bimodal trend is exacerbated by one single sample of commercial milk, that at species level has a very high diversity (Shannon alpha value ≥ 0.875). On average, we found that in the GP milk samples the richness indexes are higher than the commercial milk samples (Table 3). We then tested the differences among groups for statistical significance. The data distributions did not pass the Shapiro normality test (data not reported), so we used Wilcoxon rank sum test to compute the p-values associated with differences in significance.

An analysis of variance (ANOVA) on both Shannon and Simpson scores was performed, grouping samples by province and considering progressively deeper taxonomic levels (Fig 5). No statistically significant differences in richness were observed between the provinces. The Table 4 reports the resulting p-values.

**Fig 5. Distribution of alpha diversity as computed using Shannon’s and Simpson’s indexes.** Colours indicate different provinces of origin. The three boxes report indexes computed at different taxonomic levels.

**Table 4. p-values from ANOVA.**

|  |  |  |
| --- | --- | --- |
| **Level** | **Shannon\_pvalue** | **Simpson\_pvalue** |
| Family | 0.219 | 0.205 |
| Genus | 0.924 | 0.955 |
| Species | 0.857 | 0.792 |

Beta diversity also considers information about biological richness, but measures the variation between pairs of different communities using the Bray-Curtis index. The results are visualised in Fig 6 after the application of Multi Dimensional Scaling (MDS). Samples were highly divided into two clusters as already seen in alpha diversity analysis but, in contrast to what appeared from the alpha diversity analysis, the sample from Naples is separated from the other three extra area samples.

**Fig 6. Diversity analysis - beta diversity.** Beta diversity computed using Bray–Curtis index and then transformed via Multi-Dimensional Scaling. On the left: diversity computed by area. Commercial samples are explicitly labelled, while GP samples are reported as dots. On the right: diversity computed by province, commercial samples omitted.

# Discussion

In this work, bulk milk samples collected within the GP production area were investigated by a metabarcoding approach in order to verify if residual plant DNA could give a “snapshot” of the animal diet in the context of traceability. As far as we know, there are no published studies on DNA metabarcoding for the detection of residual plant DNA in bulk cow's milk in support of the authenticity of dairy products linked to specific territorial areas. In this regard, only a few studies have been conducted with the same goal using classic PCR approaches [20], while others have used different biological matrices such as faeces, intestinal contents etc.. [30-31]. One of the difficulties in characterising plant DNA in milk compared to other biological matrices derives from the fact that plant components resulting from the diet are strongly degraded and fragmented due to their passage through the digestive system. At the same time, it is important that the chloroplast DNA (cpDNA) target can be extracted with the least possible amount of nuclear and mitochondrial DNA and other contaminants, such as polysaccharides and resins, that can interfere and reduce the efficiency of the subsequent applications. The contamination, in fact, could reduce the efficiency of the subsequent applications. [32].

With particular reference to these aspects, the extraction method adopted in the present work proved to be effective for recovering chloroplast DNA from plants in milk. Furthermore, of considerable importance for the subsequent development of the method, the Consortium for the Barcode Of Life (CBOL) Plant Working group identified plastid coding region *rbcl and matK as* core-barcode for plant identification [33]. In the present work the *rbcl* marker was chosen because it is the most characterised plastid coding region in the GenBank database and for its higher universality compared to other DNA barcode markers described in the literature [34]. Although *matk* is considered a good marker of DNA due to its high-resolution rate, it was excluded since it is difficult to amplify with a single primer pair, especially when compared with *rbcl* which is less problematic not only in amplification [13], but also in sequencing and bioinformatics analysis.

Since GP is produced with raw milk from cows that must be fed exclusively with products permitted by the GP DOP specification, it is likely that the vegetable composition of their forages is closely linked to the production area. Therefore, the taxonomic analysis of the plant identified through the DNA isolated in their milk was performed with a local database created specifically on the basis of the Italian plant taxa. In fact, as demonstrated by several authors [31,33,35,36], for the taxonomic assignment the use of local dedicated databases reduces the possibility of wrong identifications compared to existing databases that contain millions of reference DNA sequences [37].

In the technical controls (Feed\_1, Feed\_2 and Soy flour) metabarcoding correctly identified the expected declared taxa, to which the raw materials used as fodder belong. With regard to milk samples, at the family and genus level the major number of taxa were identified in milk samples of cows belonging to GP consortium. Two unexpected plants were detected: *Ceratonia siliqua* and *Cuscuta,* for which no deeper classification was possible. The causes could be due to several factors as reported from other authors: DNA metabarcoding can produce different proportion of reads which are not related to the real quantities of each species in the sample analysed, furthermore also the sequence abundance can be affected by the density of chloroplasts in different species. Another reason could be the bias of amplification efficiency in PCR towards different species and loci due to the high level of plant DNA fragmentation, that could bring to a sharing of sequences between relative species. In particular with a manual alignment in Clustal, the *Ceratonia Siliqua (*Fabaceae family) showed an important sequence similarity (95%) with the *Glicyne max*.

Although the milk samples belonged or were external to the production area of the GP, the statistical analysis conducted using the two alpha and beta approaches did not discriminate the samples either by type or by area of origin. In particular, the beta analysis highlighted the subdivision of the samples into two distinct clusters showing how, in some cases, the milk of a specific area could actually belong to both clusters [38]. Even though the milk samples are characterised by high plant variability and came from different provinces of the GP production area, the lack of differences at multiple taxonomic levels could be due to the standardisation of feed rationing as required by the GP standard to ensure the product quality.

In order to demonstrate a link between the cow's diet and residual DNA in raw milk used in the production of GP DOP, strict control experiments are desirable. In addition, other information on the type and composition of the diet may be required as technological processes for feed production can also degrade DNA thereby distorting the final taxonomic analysis.

Results of the present work suggest that DNA metabarcoding can serve as a valuable resource to explore diet traceability. In fact, DNA Metabarcoding was found to be effective to characterise the plant component of each milk sample and then to measure the abundance of sequences at phylum, family and species level.

The results are encouraging and represent a good starting point to develop new molecular approaches, or to improve existing ones, in order to add more information to the food traceability, such as the geographical origin of the milk for the production of GP PDO.

# Acknowledgements

The authors wish to thank the “Consorzio per la Tutela del formaggio Grana Padano” for kindly providing milk samples.

This work was financially supported by the Italian Ministry of Agricultural, Food, and Forestry Policies (MIPAAF) under “New Technologies for Cheese Production (NEWTECH)” Project.

# References

1. Charlebois S, Schwab A, Henn R, Huck CW. Food Fraud. An exploratory study for measuring consumer perception towards mislabeled food products and influence on self-authentication intentions. Trends Food Sci Technol. 2016;[50](https://www.sciencedirect.com/journal/trends-in-food-science-and-technology/vol/50/suppl/C): 211-218. doi: 10.1016/j.tifs.2016.02.003.
2. Menozzi D, Giraud G, Saïdi M, Yeh CH. Choice Drivers for Quality-Labelled Food: A Cross-Cultural Comparison on PDO Cheese. Foods. 2021;10: 1176. doi: 10.3390/foods10061176.
3. Walaszczyk A, Galińska B. Food Origin Traceability from a Consumer ’s Perspective. Sustainability. 2020;12(5): 1872. doi: 10.3390/su12051872.
4. Verbeke W, Pieniak Z, Guerrero L, Hersleth M. Consumers’ awareness and attitudinal determinants of European Union quality label use on traditional foods. Bio-Based Appl Econ. 2012;1(2): 213-229. doi:10.13128/BAE-10558.
5. Van Der Lans IA, Van Ittersum K, De Cicco A, Loseby M. The role of the region of origin and EU certificates of origin in consumer evaluation of food products. Eur Rev Agric Econ. 2001; 28(4): 451–477. doi:10.1093/erae/28.4.451.
6. Van Ittersum K, Meulenberg MTG, Van Trijp HCM, Candel MJJM. Consumers’ appreciation of regional certification labels: A pan-European study. J. Agric. Econ. 2007;58: 1–23. doi:10.1111/j.1477-9552.2007.00080.x.
7. Regulation (UE) n. 1151/2012 of the European Parliament and of the Council of 21 November 2012 on Quality Schemes for Agricultural Products and Feedstuffs. Available on line: <https://eur-lex.europa.eu/>
8. https://www.qualivita.it/wp-content/uploads/2022/02/Rapporto2021-VersioneWEB.pdf
9. <https://www.granapadano.it/en-ww/production-specification-rules.aspx>
10. Stroppa A, Balli L. Grana Padano Quale latte per il Grana Padano dop. L’ Informatore Zootecnico. 2015;21: 46-47
11. Regulation (EC) No 178/2002
12. Bottero MT, Dalmasso A. Animal species identification in food products: Evolution of biomolecular methods. Vet. J. 2011;190: 34–38 doi:10.1016/j.tvjl.2010.09.024.
13. Galimberti A, De Mattia F, Losa A, Bruni I, Federici S, Casiraghi M et al. DNA barcoding as a new tool for food traceability. Food Res. Int. 2013;50(1): 55-56 doi: [10.1016/j.foodres.2012.09.036](http://dx.doi.org/10.1016/j.foodres.2012.09.036).
14. Ponzoni E, Morello L, Gianì S, Breviario D. Traceback identification of plant components in commercial compound feed through an oligonucleotide microarray based on tubulin intron polymorphism. Food Chem. 2014;162: 72–80. doi: 10.1016/j.foodchem.2014.04.021.
15. RocchettiG, Lucini L, Gallo A, MasoeroF, Trevisan M, Giuberti G. Untargeted metabolomics reveals differences in chemical fingerprints between PDO and non-PDO Grana Padano cheeses. Food Res. Int. 2018;113: 407–413. doi: 10.1016/j.foodres.2018.07.029.
16. Faustini M, Quintavalle Pastorino G, Colombani C, Chiesa LM, Panseri S, Vigo D et al. Volatilome in Milk for Grana Padano and Parmigiano Reggiano Cheeses: A First Survey. Vet. Sci. 2019;6: 41. doi: 10.3390/vetsci6020041.
17. Zago M, Bardelli T, Rossetti L, Nazzicari N, Carminati D, Galli A. et al. Evaluation of bacterial communities of Grana Padano cheese by DNA metabarcoding and DNA fingerprinting analysis. Food Microbiol. 2021;93. doi: 10.1016/j.fm.2020.103613.
18. Raats D, Offek M, Minz D, Halpern M. Molecular analysis of bacterial communities in raw milk and the impact of refrigeration on its structure and dynamics. Food Microbiol. 2011;28: 465-471. doi: 10.1016/j.fm.2010.10.009.
19. Marino M, Dubsky de Wittenau G, Saccà E, Cattonaro F, Spadotto A, Innocente N et al. Metagenomic profiles of different types of Italian high-moisture Mozzarella cheese. Food Microbiol. 2019;79: 123-131. doi:10.1016/j.fm.2018.12.007.
20. Ponzoni E, Mastromauro F, Gianì S, Breviario D. Traceability of Plant Diet Contents in Raw Cow Milk Samples. Nutrients 2009;1: 251-262. doi: 10.3390/nu1020251.
21. Bertheau Y, Helbling JC, Fortabat MN, Makhzami S, Sotinel I, Audéon C, et al. Persistence of plant DNA sequences in the blood of dairy cows fed with genetically modified (Bt176) and conventional corn silage. J. Agric. Food Chem. 2009;28: 57(2):509-16. doi: 10.1021/jf802262c. PMID: 19123817.
22. Van Eenennaam AL and Young A. E. Detection of dietary DNA, protein, and glyphosate in meat, milk, and eggs. J. Anim. Sci. 2017;95: 3247–3269. doi: 10.2527/jas2016.1346.
23. Ponzoni E, Gianì S, Mastromauro F, Breviario D. From milk to diet: Feed recognition for milk authenticity. J. Dairy Sci. 2009;92: 5583–5594. doi: 10.3168/jds.2009-2239.
24. Phipps RH, Deaville ER, Maddison BC. Detection of Transgenic and Endogenous Plant DNA in Rumen Fluid, Duodenal Digesta, Milk, Blood, and Feces of Lactating Dairy Cows. J. Dairy Sci. 2003; 86: 4070–4078. doi 10.3168/jds.S0022-0302(03)74019-3.
25. Oksanen J, Guillaume Blanchet F, Friendly M, Kindt R, Legendre P, McGlinn D, et al. Vegan: Community Ecology Package. R package version 2.5-6. 2019 [https://CRAN.R-project.org/package=vegan](https://cran.r-project.org/package=vegan)
26. Whittaker RH. Vegetation of the Siskiyou Mountains, Oregon and California. Ecological Monographs 1960; 30 (3). Wiley Online Library: 279–338.
27. Shannon CE. A mathematical theory of communication. Bell Sys Tech J. 1948;27: 379-423, 623-59.
28. Simpson EH. Measurement of diversity. Nature. 1949;163: 688.
29. Bray JR, Curtis JT. An ordination of upland forest communities of southern Wisconsin. Ecological Monographs. 1957;27: 325-349.
30. Palumbo F, Squartini A, Barcaccia G, Macolino S, Pornaro C, Pindo M et al. A multi‑kingdom metabarcoding study on cattle grazing Alpine pastures discloses intra‑seasonal shifts in plant selection and faecal microbiota. Sci. Rep. 2021;11:889. doi: 10.1038/s41598-020-79474-w.
31. Nakahara F ,Ando H,Ito H, Murakami A , Morimoto N, Yamasaki M et al. The applicability of DNA barcoding for dietary analysis of sika deer. DNA Barcodes 2015;3(1): 200-206. doi: 10.1515/dna-2015-0021.
32. Shofiqul I, Buttelmann GL, Chekhovskiy K , Kwon T, Saha MC. Isolation of Intact Chloroplast for Sequencing Plastid Genomes of Five Festuca Species. Plants. 2019;8:606. doi:10.3390/plants8120606.
33. Bruni I, Galimberti A, Caridi L, Scaccabarozzi D, De Mattia F , Casiraghi M et al. A DNA barcoding approach to identify plant species in multiflower honey. Food Chem. 2015;170: 308–315. [doi: j.foodchem.2014.08.060](http://dx.doi.org/10.1016/j.foodchem.2014.08.060).
34. Newmaster SG, Fazekas AJ, Ragupathy S. DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach. Can. J. Bot. 2006;84: 335–341 doi: 10.1139/B06-047.
35. Hawkins J, De Vere N, Griffith A, Ford CR, Allainguillaume J, Hegarty MJ et al. Using DNA Metabarcoding to Identify the Floral Composition of Honey: A New Tool for Investigating Honey Bee Foraging Preferences. Plose One. 2015;10(8):e0134735.2015. doi: 10.1371/journal.pone.0134735.
36. Bruno A, Sandionigi A, Agostinetto G, Bernabovi L, Frigerio J, Casiraghi M et al. Food Tracking Perspective: DNA Metabarcoding to Identify Plant Composition in Complex and Processed Food Product. Genes 2019;10: 248. doi:10.3390/genes10030248.
37. De Mattia F, Gentili R, Bruni I, Galimberti A, Sgorbati S, Casiraghi M et al. A multi-marker DNA barcoding approach to save time and resources in vegetation surveys. Bot. J. Linn. 2012; 169(3): 518–529. doi:10.1111/j.1095-8339.2012.01251.x.
38. Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. Appl. Environ. Microbiol. 2007;73(5): 1576–85. doi: 10.1128/AEM.01996-06. PMC 1828774. PMID 17220268.

# Supporting information

**S1 Fig. Rarefaction curves.**

**S2 Fig.** **Species heatmap.** Heatmap showing the abundance of the fourteen dominant species (≥1% total reads on at least one sample) found in milk samples. The colour scale indicates the species’ relative abundance: more intense is the colour, more abundant are the species. The province of origin is indicated by the first two letters of each sample, with the exception of commercial milk samples which are explicitly marked as such.

**S3 Fig.** **Genus heatmap.** Heatmap showing the abundance of the fourteen dominant genuses (≥1% total reads on at least one sample) found in milk samples. The colour scale indicates the species’ relative abundance: more intense is the colour, more abundant are the species. The province of origin is indicated by the first two letters of each sample, with the exception of commercial milk samples which are explicitly marked as such.

**S4 Fig.** **Family heatmap.** Heatmap showing the abundance of the seven dominant families (≥1% total reads on at least one sample) found in milk samples. The colour scale indicates the species’ relative abundance: more intense is the colour, more abundant are the species. The province of origin is indicated by the first two letters of each sample, with the exception of commercial milk samples which are explicitly marked as such.

**S1 Table. Family distribution.**

**S2 Table. Genera distribution.**

**S3 Table. Species distribution.**

**S1 Table. Family distribution.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **All samples** | **GP milk** | **Commercial milk** |
| Fabaceae | 57.207 | 55.922 | 70.38 |
| Convolvulaceae | 26.080 | 26.618 | 20.57 |
| Poaceae | 9.028 | 9.419 | 5.03 |
| Rubiaceae | 4.178 | 4.318 | 2.74 |
| Malvaceae | 3.237 | 3.427 | 1.29 |
| Asteraceae | 0.164 | 0.181 |  |
| Ranunculaceae | 0.102 | 0.112 |  |
| Campanulaceae | 0.003 | 0.003 |  |

**S2 Table. Genera distribution.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **All samples** | **GP milk** | **Commercial milk** |
| Ceratonia | 34.480 | 33.912 | 40.299 |
| Cuscuta | 26.361 | 26.925 | 20.585 |
| Glycine | 21.792 | 21.155 | 28.320 |
| Triticum | 5.389 | 5.758 | 1.600 |
| Galium | 4.250 | 4.397 | 2.738 |
| Gossypium | 3.295 | 3.490 | 1.291 |
| Medicago | 1.877 | 1.890 | 1.740 |
| Panicum | 1.038 | 0.809 | 3.392 |
| Lolium | 0.832 | 0.913 |  |
| Helianthus | 0.250 | 0.275 |  |
| Pisum | 0.136 | 0.149 |  |
| Zea | 0.115 | 0.126 |  |
| Delphinium | 0.103 | 0.113 |  |
| Vicia | 0.055 | 0.060 |  |
| Secale | 0.021 | 0.023 |  |
| Tripidium | 0.003 |  | 0.034 |
| Lobelia | 0.003 | 0.003 |  |

**S3 Table. Species distribution.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **All samples** | **GP** | **Commercial** |
| Ceratonia siliqua | 61.835 | 61.439 | 65.894 |
| Triticum aestivum | 23.396 | 23.221 | 25.184 |
| Medicago sativa | 5.245 | 5.490 | 2.738 |
| Gossypium mustelinum | 3.228 | 3.543 |  |
| Glycine max | 2.627 | 2.808 | 0.770 |
| Panicum pygmaeum | 0.880 | 0.445 | 5.346 |
| Panicum wiehei | 0.878 | 0.964 |  |
| Lolium perenne | 0.630 | 0.691 |  |
| Pisum sativum | 0.439 | 0.482 |  |
| Helianthus annuus | 0.341 | 0.374 |  |
| Vicia faba | 0.156 | 0.172 |  |
| Zea mays | 0.150 | 0.165 |  |
| Delphinium grandiflorum | 0.133 | 0.146 |  |
| Secale cereale | 0.028 | 0.030 |  |
| Galium murale | 0.022 | 0.024 |  |
| Cuscuta reflexa | 0.007 | 0.008 |  |
| Tripidium kanashiroi | 0.003 |  | 0.029 |
| Tripidium arundinaceum | 0.002 |  | 0.027 |
| Glycine stenophita | 0.001 |  | 0.011 |