Università degli Studi di Milano-Bicocca

Dipartimento di Biotecnologie e Bioscienze

Dottorato di ricerca in Tecnologie Convergenti per i Sistemi Biomolecolari (TeCSBi) – XXXIII Ciclo

Towards food traceability: discovering biomolecular technologies for complex food products

Jessica Frigerio

Matr. 725751

Tutor interno: Prof. Massimo Labra

Tutor esterno: Dr. Fabrizio De Mattia

Anno Accademico 2019/2020

Summary

1. Abstract

In the agri-food sector, food frauds are increasing every year and there is a demand from the consumer to better understand food products. Furthermore, due to globalization, the supply chain is long and wide so is important to ensure safe and high-quality food products. A suitable bio-molecular identification and traceability system could be used to assess the quality of raw materials up to the finished food products as well as to guarantee the consumer on the identity of the products purchased. DNA testing is a methodology frequently used in the food field, especially DNA barcoding methodology. DNA barcoding is a universal tool to identify species, and it currently allows for routinely supply-chain analysis. However, DNA barcoding technique has some limits in the complex food products analysis. Industrial treatments could alter DNA quality of raw material, therefore this analysis could be challenging to apply. Furthermore, DNA barcoding cannot be applied for multispecies products because is feasible only for monospecies products. Finally, DNA barcoding is time-consuming and expensive. Starting from these limits, the main objective of this PhD project was to identify and test innovative biomolecular analyses for the identity assessment of raw materials and processed food products. Three lines of research have been developed in this PhD thesis: the first one aim to verify the DNA barcoding feasibility on processed products (that can have the DNA fragmented or degraded), the second one was focused on the evaluation of the potential of an HTS-DNA metabarcoding approach for food products analysis and the third line of research aim to the development of a rapid kit to offer to companies. The stakeholders of this project are both companies, who want to guarantee high quality and safety products and final consumers, who have been increasingly careful about the food products. For this reason, the proposed methods must be simple, effective and possibly universal to be applied on a large scale.

In the first line of research was evaluated whether DNA barcoding can be applied to trace the plant component in food supplements from the starting raw material to the finished processed commercial products. DNA barcoding technique analyses large fragments of DNA, but in case of processed products DNA is fragmented and therefore small and DNA mini-barcoding could be more successful. In this study we selected a panel of 64 samples between phytoextracts obtained through three different extraction methods (i.e., maceration, percolation and sonication) with different solvents (i.e., ethanol, deionized water and glycerol) and some intermediates of the industrial production. We sequenced and analysed the sequence variability at DNA barcoding (*psbA-trnH*, ITS2) and minibarcoding (rbcL 1-B) marker regions. Phytoextracts obtained through hydroalcoholic treatment, with the lower percentage of ethanol (<40%) and aqueous processing had a major rate of DNA amplification, sequencing and identification success at the lowest temperature. This study proves that DNA barcoding is a useful tool for some typology of food supplement traceability, which would provide consumers with safe and high-quality herbal products.

A second limit that occurs in the food market is related to multispecies products. Many products contain a mix of different species, but DNA barcoding does not allow the analysis of multiple species at the same time. For this reason, the second objective of this thesis project was to test DNA metabarcoding (High-Throughput Sequencing technique in correlation with DNA barcoding). This technique is capable of analysing several genomes at the same time and distinguishing them thus recognizing the species of the mixture of processed food products. In the second line of research we wanted to characterize plant declared ingredients or contaminants (ITS2) in five categories of commercial insect-based products (novel food) with attention to putative elements of allergenic concern belonging, for example to the

insect rearing substrate. The same approach has been used to assess its sensitivity to cases of contamination and counterfeits in insect flour with low cost (and potentially allergenic) vegetable flours like wheat and soybean. Moreover, also the bacteria profile (16S rRNA) was evaluated. The DNA metabarcoding analysis revealed a high efficacy as a screening method to identify both plant ingredients and possible adulteration events, also acting as an early warning strategy for the occurrence of allergens of human concern. HTS revealed an high sensitivity because we were also able to identify vegetal traces belonging to insect farming. Concerning the bacteria profile, our data revealed that a small number of prevalent bacteria formed a "core microbiome" characterizing the products depending on the insect, suggesting that a resident microbiota is conserved. Despite different food processing levels rearing conditions or seller companies, this microbial signature can be recognized. This approach could support the development of new risk assessment procedures for novel foods by regulatory authorities to ensure their quality, safety and acceptance which will become more required in order to face the challenge of feeding the world population in the next decades. In order to evaluate the application of DNA metabarcoding also for herbal teas traceability, we analysed fifteen commercial herbal teas to identify all the species in the products and verify the correspondence with the label. Furthermore, to verify the quantitative ability of HTS, we created six mocks mixture with different percentages of five species of different matrices (leaves, seeds, bark, flowers and aerial parts). These mocks were created both starting from raw material (gr/gr) and genomic DNA (ng/µL). Finally, for all the samples we tested two different markers (*psbA-trnH* and ITS2). Results showed that not all the species declared on the label was found and we also were able to detect some contaminants. Interestingly, the two markers were able to detect different species: some species were detected only by *psbA-trnH* couple of primers, others only by

ITS2 couple of primers. For future analyses it would be necessary to use both the markers to obtain better results in traceability. Finally, the mock mixture created in the laboratory showed that there wasn't a perfect correlation between the real percentage of a plant and the feature found in the samples but is possible to have a relative quantification of all the ingredients. We obtained better results in the mock mixture created starting by genomic DNA. In conclusion in this line of research we defined that DNA metabarcoding is a valuable method for supply chain traceability, both for food products (novel food) and for herbal teas.

The third line of research aimed to create a mock-up for truffle identification using the LAMP technique, an isothermal nucleic acid amplification technique. We collected five different species of truffle (three brown truffles and two white truffles) that were morphologically identified. Two couples of primers were specifically designed (ITS2) for *Tuber magnatum* (the most expensive truffle) and another two couples as a control for DNA extraction (β-tubulin gene). It was also selected as a rapid extraction buffer (about 2 minutes for DNA extraction). LAMP reaction could amplify all truffles with the DNA extraction control, and only *T. magnatum* with the specific couples of primers. This low-cost kit allows the identification of the white truffles *T. magnatum* in a total of 90 minutes.

In conclusion, these studies allowed us to develop innovative methodologies to analyse a wide typology of products on the food market (fresh, processed, multispecies). Furthermore, we were able to develop a rapid kit for white truffles identification. This kit has the advantage of being versatile and can be adapted to the product of interest (fish, meat, oil etc).

2. Introduction

2.1. The food fraud phenomenon

The food fraud phenomenon has existed since the food was bought and sold for cash or commodities. In ancient Rome and Athens, a law was enacted for the adulteration of wine, and in the thirteenth century Europe prohibited food adulteration (Sumar and Ismail, 1995; Barerre et al., 2020). In recent years, food fraud has become a significant concern for consumers and a growing risk for the food industry. It is estimated that up to 10% of food products are frauded (Canadian Food Inspection Agency, 2020). The cost has been estimated to be around $30 to 50 billion every year (Smith, 2020). Furthermore, the horse-meat scandal of 2013 put the spotlight on the need to increase food controls. The European Parliament published on 14 January 2014 a resolution on the food crisis, fraud in the food chain and the control thereof. The principal aim was to give food fraud the full attention it warrants and to take all necessary steps to make the prevention and combating of food fraud an integral part of EU policy (European Parliament, 2016).

Furthermore, the globalization of food supply chain has extended the distances between primary producers and consumers resulting in increased complexity in the supply chain and has probably increased the issue of food fraud (Song et al., 2020). Recently, the media has drawn consumers' attention to several international cases of food fraud with an impact on public health. Cases of food fraud have led to confusion and a lack of confidence for consumers. Some recent examples are melamine in milk, horse meat in beef processed products, and pathogenic bacteria in peanut butter (Théolier et al, 2020). Operation OPSON, a joint action of Europol and Interpol conducted in 61 countries, has seized thousands of tons of counterfeit and substandard food and drink, for a total of around € 230 million in counterfeit food and drink. OPSON operations highlight the importance of food fraud around the world and how this activity affects the food chain (Guntzburger et al., 2020).

In order to reduce food fraud and protect Italian products, the Council of Ministers approved new legislation for the offences of "agri-food piracy", such as the sale of products as "Made in Italy" when they are not, falsely declaring that the food is organic or counterfeiting. In October 2020, for example, the Italian authorities seized 7 tons of food products wrongly labelled as "Made in Italy" but imported from Africa (Food Fraud Summary October, 2020).

2.1.1. The food fraud: one concept, many definitions

Currently, there is not any accepted definition of food fraud. In the table below there are some definitions from the academic literature (Robson et al., 2021). Generally, food fraud is defined as an intentional misrepresentation of food for economic gain.

Food fraud definition	**Reference**
Food fraud is a collective term used to encompasses the deliberate and intentional substitution, addition, tampering or misrepresentation of food, food ingredients or food packaging: of false or misleading statements made about a product, for economic gain.	Spink and Moyer (2011)
The intentional adulteration of food for financial advantage.	Everstine et al., (2013)
Committed when food is deliberately placed on the market for financial gain, with the intent of deception of consumers. Referred to in the USA and occasionally elsewhere as economically motivated adulteration (EMA). Two of the main types include: trading	Ellis Muhamadali et al., (2015)

of food which is unfit for consumption or harmful, or deliberately misdescribing or mislabelling food. The latter can include false statements regarding geographical origin, ingredients, or substitution with lower value (i.e. myrtle instead of oregano), or sometimes even dangerous contents not intended for human consumption (i.e. industrial dyes). The terms food fraud and food adulteration can be used to mean the same thing, when adulteration is intentional.	
An international act with motivation for economic gain.	Spink et al., (2015)
The deliberate intent to deceive, motivated by the prospect of financial gain.	Charlebois et al., (2016)
The intentional misrepresentation of fact by one person solely, or acting on behalf of an organization, in order to encourage another individual erroneously to part with something of intrinsic value.	Manning (2016)
Food fraud (including the subcategory the US Food and Drug Administration (FDA) defined as Economically Motivated Adulteration (EMA)) is illegal deception for economic gain using food.	Moyer et al., (2017)
Illegal intentional deception for economic gain using food can occur in all stages of the supply chain and often cross international borders.	Spink et al., (2017)
Food fraud covers cases where there is a violation of EU food law, which is committed intentionally to pursue an economic or financial	Bouzembrak et al., (2018)

gain through consumer deception. Food fraud in the food supply chain can arise as a result of misrepresentation associated with: product integrity (e.g. counterfeit product, expiration date), process integrity (e.g. diversion of products outside of intended markets), people integrity (e.g. characterizations such as the cyber criminals and hacktivist) and data integrity (e.g. improper, expired, fraudulent or missing common entry documents or health certificates) of information accompanying the food item throughout the supply chain.	
An intentional change in a food product that a consumer is unaware of with their purpose to deceive consumers whether to cause harm or to economically benefit.	Cruse (2019)
Intentional modification of food products and/or associated documentation for economic gain and may lead to issues of food safety, legality and/or quality depending on the activities undertaken or the agent(s) used.	Manning and Soon (2019)
Long Definition: Illegal deception for economic gain using food encompasses deliberate and intentional substitution, addition, tampering, or misrepresentation of food, food ingredients, or food packaging; or false or misleading statements made about a product for economic gain. The types of fraud include adulteration, tampering, product overrun, theft, diversion, simulation, and counterfeiting.	Spink et al., (2019)

Short Definition: Illegal deception for economic gain using food	Spink et al., (2019)

Table 1. Food fraud definitions from academic literature (Robson et al., 2021).

Due to the vast definition of food fraud, the terms food defence, food safety and food quality are important for a deep understanding of this expression. Essentially the difference is about the motivation and the intention. Fraud can be intentional or unintentional and the goal is usually the economic gain. However, in some cases the main motivation for food fraud may be related to damaging the public health, the economy or generating terror (Spink and Moyer, 2011). Spink and Moyer have illustrated in a risk matrix the correlation between motivation and intention to the different definitions. The image below is an adaptation of the risk matrix (Figure 1).

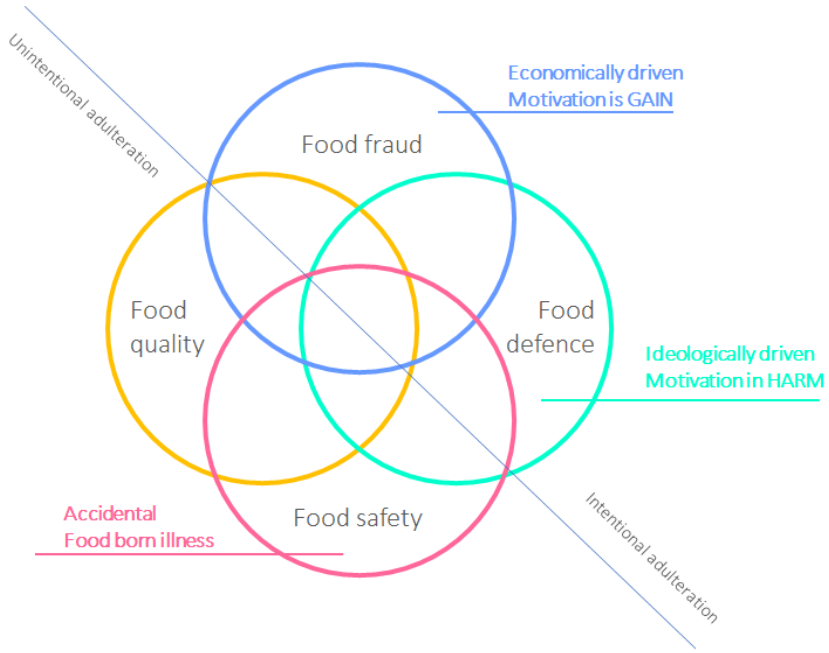


Figure 1. Graphical representation of the difference of food fraud, food quality, food defence and food safety. When the adulteration is intentional and with an economical gain there is a case of food fraud, but when the aim is to harm people there is a case of food defence. In case of unintentional adulteration, there is a food safety issue in the case of a security problem of the consumer, otherwise there is only a food quality issue.

If a fraud is unintentional, it is a "*food safety*" issue, because consumer health can be harmed, otherwise is only a "*food quality*" issue. When the action is intentional, we are facing a crime. If the crime aims is to harm people we are in front of "*food defence*", if the intention is an economical gain the action is evaluated as a "*food fraud*" (Morin and Lees, 2018). Several typologies of fraud can be found, as it can be seen in the figure below. The most common typology is Adulteration, which is referred to as Economically Motivated Adulteration (EMA). In May 2009 FDA

defined EMA as: "*the fraudulent, intentional substitution or addition of a substance in a product for the purpose of increasing the apparent value of the product or reducing the cost of its production*". Different types of adulteration can occur in food products, such as dilution, substitution, concealment and unapproved enhancement (Figure 2).

Dilution is "*the process of mixing an ingredient with high value with an ingredient with a lower value*"'. Examples include dilution of honey with sugar syrup and the addition of water to fruit juice and milk.

Substitution is the "*process of mixing a nutrient, an ingredient with high value with an ingredient with a lower value*". Examples of substitutions are substituting EVO oil with seed oil or substituting high-value fish species with lower value fish, especially in the form of fillets or processed products.

Concealment is the "*process of hiding the low quality of food ingredients or products*". An example is injecting poultry with hormones to conceal disease (Morin and Lees, 2018).

Unapproved enhancement is the "*process of adding unknown and undeclared compounds to food products in order to enhance their quality attributes*". Some examples are the addition of Sudan dyes in species such as chili powder and the addition of melamine in milk (Morin and Lees, 2018).

The other types of food frauds are:

Counterfeit: it is an "*Infringement of intellectual property rights*". It can be related to the packaging of a product, an illegal reproduction of a patent but also the copying of the brand name. A common counterfeit is label "Made in Italy" when the product is made abroad.

Grey market: this term includes "*production, theft, and diversion*" and involves unauthorised sales channels for products.

Mislabelling is defined as *"False claims or distortion of the information provided on the label/packaging".* In the field of seafood this is a common typology of fraud, where fillets of lower value are sold instead of the species of a higher value. Even if mislabelling may apply to all forms of food fraud, the expression is mainly used to indicate false information provided on the label.

Figure 2. Different typologies of food fraud (https://ec.europa.eu/food/safety/food-fraud/what-does-it-mean_en).

2.1.2. Food fraud in modern society

Food fraud has been increasing in the last few years and it is a global business worth over $50 billion annually (Smith, 2020). The most common fraud in the food field is the substitution of an animal or plant of high commercial value with a cheaper one. This phenomenon occurs in all fields, from meat and seafood to dairy products, oils and juice.

The scientific literature had raised interest in the food fraud phenomenon during the years. In the figure below (Figure 3) is shown the absolute number of publications containing the keyword "food fraud" over the years from the database "Pubmed". It seems that the 2013 horsemeat scandal has given a boost to scientific publications in this area.

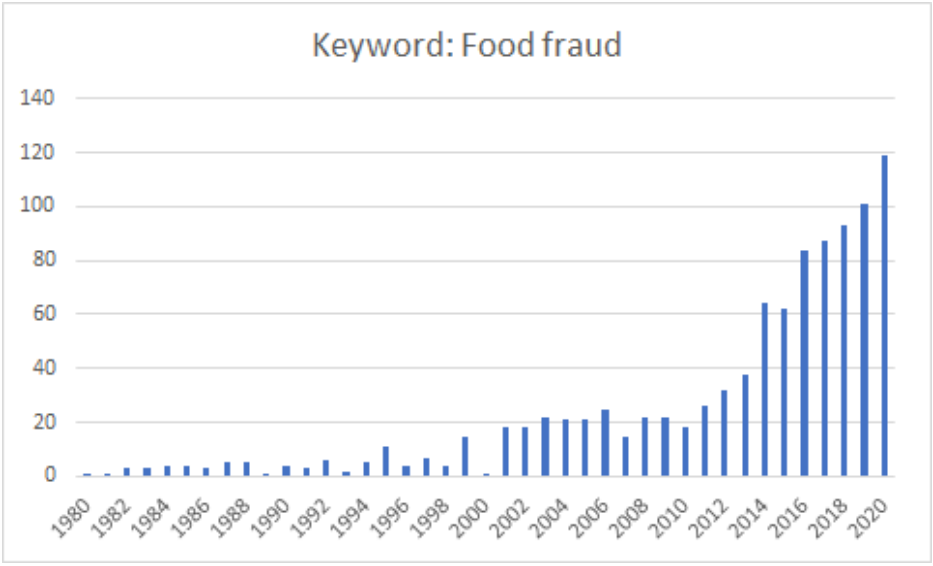


Figure 3. The absolute number of publications containing the keyword "food fraud" over the years from the database "Pubmed".

More specifically, from Scopus database, using terms such as “fish”, “honey”, “meat” etc in combination with “fraud” and “adulteration” it is possible to obtain the figure below (Ulberth, 2020). Also in this case, it is clear the increase of publications from 2013 (Figure 4).

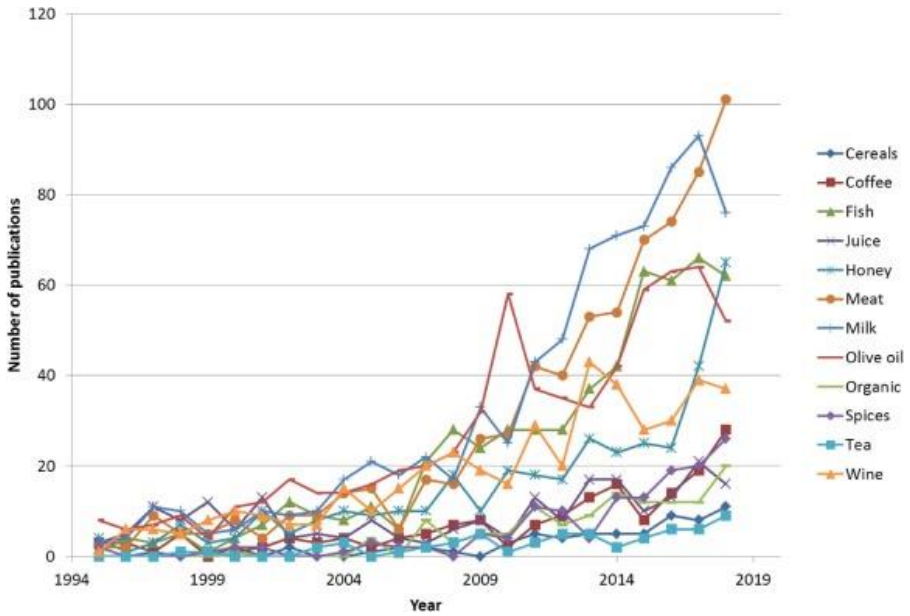


Figure 4. Number of publications related to food fraud retrieved from the Scopus database (Ulberth, 2020).

The most common fraud includes oil, fish, honey, milk and dairy products, meat products, grain-based foods, fruit juices, wine and alcoholic beverages, organic foods, spices, coffee, tea, and some highly processed foods (Hong et al., 2017).
Mislabelling have been frequently reported in the research publications: 57% in processed meat products (Di Pinto et al., 2015) and 80% in fish fillets sold in Italy

(Barbuto et al., 2010), up to 35% in meat products (Kane and Hellberg, 2016) and 39% in fish fillets (Liou et al., 2020) sold in the United States and up to 80% in dairy products (Di Pinto et al., 2017).

Concerning the herbal supplements field, a global survey shows that 27% of the herbal products commercialized in the global marketplace is adulterated when their content was tested against their labelled, claimed ingredient species (Figure 5). The most defrauded regions are South America (67% mislabelled products) and Australia (79% mislabelled products) (Ichim, 2019).

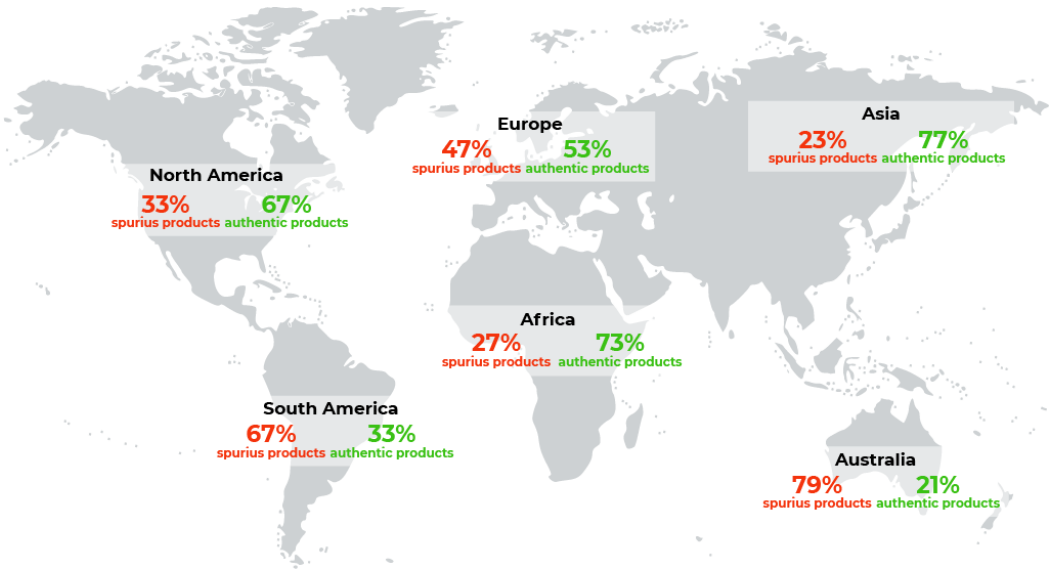


Figure 5. Percentage of mislabelling of herbal products in the North and South America, Europe, Africa, Asia and Australia.

Dietary supplements and spices are commonly exposed to fraud through the substitution of the corresponding plant species (De Mattia et al., 2011; Derz et al., 2020). The form of powder in which they are sold make authentication challenging.

Undeclared species substitution in food products might also represent an important health threat to allergic consumers because of the introduction of food allergens, such as different kinds of nuts and molluscs (Böhme et al., 2019) or poisonous plants (Mezzasalma et al., 2017).

Example of a common adulteration is the case of saffron (*Crocus sativus* L.). Saffron is an expensive spice (from 15.000 to 30.000 euro for kilo). This high cost is due to the harvest methodology. This spice is derived from the dried stigmas of *Crocus sativus*, which is harvested by hand. It also has an attractive power to the botanicals industries, due to its medical proprieties. The most expensive saffron came from Greek and Italy, followed by Spain and finally Iran. The adulteration of this spice is very common, and usually it is mixed with lower value products. In 2018, for example, in the UK a total of 87kg of saffron, coming from Spain, were seized with an estimated market value of EUR 783,000 (The EU Food Fraud Network and the Administrative Assistance and Cooperation System, 2019 Annual Report).

The European Commission publishes a monthly report on worldwide fraud and seizures carried out. Thirty-one cases of fraud were reported only in the October 2020 report. For example, the Moroccan authorities seized 1.2 million bottles of adulterated wine. These bottles were also imported illegally, expired or unlabelled, and could have endangered consumer health. Moreover, the Khyber Pakhtunkhwa Food Authority seized 3 tonnes of spices adulterated with hazardous colours, hay, and other unhealthy materials. Finally, the Italian Coast Guard, with the operation "Cephalopoda", seized almost 1700 kg of seafood products for a total amount of around 34000 euros. These products were mislabelled or missed the required traceability documentation (Food Fraud Summary October, 2020).

2.1.3. *Food EU regulations and food fraud*

Since 1996, in Europe, there have been problems in the food sector (such as BSE (Bovine Spongiform Encephalopathy), dioxins in food, etc.) which have put the health of the consumer at risk. For this reason, the European Commission (EC) has set itself the goal of stipulating rules for food safety and the first step was the publication of 2 books:

The Green Book (Brussels, 1997) which defines the general principles of food law in the EU (http://www.salute.gov.it/imgs/C_17_pubblicazioni_2643_allegato.pdf);

The White Paper (Brussels, 2000) which presents the legislative developments mentioned in the story Green Paper for food safety (http://www.salute.gov.it/imgs/C_17_pubblicazioni_1553_allegato.pdf).

Subsequently, on January 28, 2002, (Reg. (EC) n. 178/2002) the “General Food Law” was enacted. Has the general principles and requirements of food law and establishes the European Food Safety authority and the fixed procedure in the field of food safety (https: //eur-lex.europa. eu /LexUriServ/LexUriServ.do?uri=CONSLEG:2002R0178:20060428:IT:PDF).

Finally, with the UNI EN ISO 22005: 2008 "Traceability in agri-food supply chains" (https://irp-cdn.multiscreensite.com/8524317a/docs/1.3907234..pdf) the concept of "traceability system" is definitively specified .

To complete the picture of food safety, on 29 April 2004, the so-called “Hygiene Package” was published, which came into force on 1 January 2006 and is part of the “from farm to table” strategy. This package consists of a set of 4 legislative texts that regulate the hygiene of food production and the controls to which they must be subjected. The "Hygiene Package" is composed as follows:

Reg. (CE) n. 852/2004, on food hygiene (https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2004:139:0001:0054:it:PDF);

Reg. (CE) n. 853/2004, on specific hygiene rules for food of animal origin (https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32004R0853&from=HR);

Reg. (CE) n. 854/2004, on specific rules for the organization of official controls on products of animal origin intended for human consumption (https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32004R0854);

Reg. (CE) n. 882/2004, on official controls aimed at verifying compliance with feed and food legislation and with animal health and welfare regulations (http://www.salute.gov.it/imgs/C_17_normativa_1793_allegato.pdf).

With these 4 legislative texts, the "Hygiene Package" manages to take into consideration all levels of the food chain: from primary production to breeding, from hunting and fishing to production, from processing to distribution, up to the sale of the product to consumer. This method of analysis is called the “supply chain approach”.

Finally, the "Hygiene Package" outlines two professionals with the task of supervising compliance with these 4 regulations:

The Food Sector Operator (FBO), solely responsible for Reg. (EC) no. 852/2004 and 853/2004;

The Competent Authority (CA), for the official controls provided for in Reg. (EC) no. 854/2004 and 882/2004.

In 2013, to complete the "farm to table" strategy, the EC decided to issue a further package of five measures on animal and plant health, entitled "Smarter rules for safer food" (https://eur-lex.europa.eu/legal-content/it/TXT/PDF/?uri=OJ:L:2014:189:FULL&from=EN). It is based on a more risk assessment approach thus focusing on animal diseases and plant pests to provide safer products for consumers. Food supplements - intended as a concentrated source of substances with a nutritional or physiological effect but never therapeutic or preventive of diseases - are subject to the control of the EFSA (European Food Safety Agency), but the legislation in this field is less developed and restrictive. Only guidelines are available to ensure the safety of botanicals and derived products in order to obtain food supplements and it is recommended to follow the directions of the pharmacopoeia to achieve identification. In addition, there is a continuously updated list of substances that are reported to influence health when used in the preparation of foods or supplements.

2.2. Food traceability: the importance of the supply chain control

2.2.1. Covid-19 outbreaks consequences in the agri-food field

Covid-19 outbreaks have been, and will continue to have, a generic impact on the likelihood of many food fraud risks (Points and Manning, 2020). As a society, we need the food and agriculture sector to continue to operate despite the Covid-19 outbreaks (Cauhan et al., 2020). Nevertheless, the companies responsible for activities such as checking, inspecting, auditing and testing have been minimized and, in some cases, totally collapsed. Moreover, food companies are much more concerned with trying to guarantee an adequate food supply, instead of controlling the suppliers and ensure a safe and authentic food. The supply chain control is a key aspect to consider as part of the food fraud prevention program.

The fraud risk in these months is increased due to the supply chain disruption, the surplus of specific food commodities regionally, the changes in commodity prices, the reduction of supervision of regular procedures, the regional labour shortages in farming and finally the changing consumer purchasing behaviour (e commerce, local food) (Laborde et al., 2020).

The Opson 2020 report, an investigation led by Interpol and Europol, which cover the period from December 2019 to June 2020 (the pandemic period), revealed that 19 organised crime groups were dismantled and over 400 arrests made. The defrauded products were cereals and grains, meat, coffee, spices and dairy products (Elliot, 2020). High value products can be the most endangered. Products such as botanical ingredients used in food supplements may be particularly at risk due to the supply chain disruptions. There are some reported cases of food fraud due to supply chain disruptions. There was a case of mislabelled horsemeat in Europe (such as the 2013 scandal) (Taylor, 2020) and a case of false endemic strawberry in

Wien (Fresh Plaza, 2020). In India thousands of kilograms of toxic fish and shrimp were seized due to a supply chain shortage resulting in the half in fishing caused by lockdown measures, which “oblige” companies to add non-compliant food.
The supply chain traceability and surveillance are fundamental, especially in this difficult period. The Acheson Group and The Food Chemicals Codex group recommends increasing supply chain surveillance and testing (Everstine, 2020).

2.2.2. The blockchain technology for supply chain traceability

Blockchain technology emerged in 2008 as a core component of the bitcoin cryptocurrency (Bhardwaj and Kaushik, 2017) and is a shared and immutable data structure. It is defined as a digital register whose entries are grouped into blocks, concatenated in chronological order, and whose integrity is guaranteed using cryptography. Although its size is destined to grow over time it is immutable since, as a rule, its content once written is no longer modifiable or eliminable unless the entire structure is invalidated. These technologies are included in the broader family of Distributed Ledgers, systems that are based on a distributed ledger, which can be read and modified by multiple nodes on a network. The nodes involved are not required to know each other's identity or trust each other. In fact, to ensure consistency between the various copies, the addition of a new block is globally governed by a shared protocol. Once the addition of the new block has been authorized, each node updates its own private copy: the very nature of the data structure guarantees the absence of its future manipulation. Thanks to these characteristics, the blockchain is therefore considered an alternative in terms of security, reliability, transparency and costs to databases and registers managed centrally by recognized and regulated authorities (public administrations, banks, insurance companies, payment intermediaries, etc.).

The agri-food field gained interest in this technology and applied blockchain to trace a product "from farm to fork". The aim is to trace every single product all along the supply chain. A customer can scan a QR Code using the smartphone and could obtain all the quality and safety information about a product, even all the analysis results.

The computerisation and molecularization of traceability tools open the opportunity to use DNA analysis are emerging technologies to address food traceability. A digital copy of the DNA can then be attached to every item or product a company creates, which brings traceability to the item level, rather than to an entire batch. Furthermore, digital DNA can be cross-checked with the blockchain record to ensure the product's authenticity. This process allows producers to create a chain of custody. With this tool it is possible to prevent fraud, offering total traceability and guarantee the authenticity of the food.

Despite the great potential of this technology, a complex and long supply chain can be expensive, difficult to implement and hard to handle a huge volume of data. Furthermore, is necessary the digitalization of all the documents, because manual written documents lead to human error and it is difficult to quickly record the data (Galvez at al., 2018).

Figure 6. The blockchain technology applied to the agri-food supply chain.

In 2017, Walmart implemented a pilot test of blockchain technology developed by IBM with the aim to track pork in China (Kamath, 2018). It happened that a customer became ill, and Walmart thanks to blockchain technology was able to trace all the supply chain of that specific product.

Carrefour, in collaboration with Crystalchain e Connecting Food, developed in 2018 the first food blockchain technology in Europe. They started with chicken and in a second time they extended this technology to other products: the Cauralina tomatoes, the Loué farm eggs and Rocamadour DOC. All these products have a simple QR Code placed on the package, thanks to it is possible to see on a smartphone all the information.

Figure 7. Blockchain's chicken for Carrefour.

Also Coop Italia and IBM, with the project "Coopchain", collaborate for the application of the blockchain to the food supply chain and specifically to the production of Coop brand eggs. The solution allows for full transparency and visibility of the various actors involved in the Coop-branded egg production chain, involving a total of 2 million hens for over 200 million eggs produced per year.

By scanning the QR Code printed on the new package, like for Carrefour, and typing the specific code of the batch, the consumer can thus discover the history of the product and trace back from the point of sale to the farm, to the territory from which the egg comes and to the incubator from which it was born. In this way, consumers can ascertain that the eggs were produced in full compliance with animal welfare requirements, never closed in cages, and without the use of antibiotics, thanks also to the certification of two independent third parties.

Figure 8. Blockchain's egg for CoopItalia.

To measure the satisfaction of this solution in various Coop stores in Italy, more than a thousand consumers were interviewed and it emerged that all those who used the QR Code expressed very high levels of satisfaction and interest in both the scan done in-store and at home.

This trace technology is the future of the food industry because it can store the history of all transactions ever made and allows to recreate the history and identify the origin of a product (Galvez et al., 2018).

2.3. Genomics and Metagenomics approach for Food Quality

2.3.1. DNA analyses for food quality, authentication and traceability

Food authentication is made by a wide typology of analysis. Chemical methods, such as HPLC, LC-MS, GC-MS, HPTLC, spectroscopic data from an MS, NMR, UV, or IR experiment are widely used (Pawar et al., 2017). However, in most cases there are not universal standards especially for processed materials and this is one of the major limiting factors that prevents the widespread adoption of such chemical methods for universal traceability (Khan et al., 2012). Moreover, the chemical profile is often subject to phenotypic variables. For example, an aromatic plant can have different quantitative of aromas and qualitative variations depending on the cultivation area, harvesting, processing and conservation methods (Abdelmajeed et al., 2013). There is therefore a risk of not correctly identifying a matrix. Finally, in many cases the chemical parameters chosen do not necessarily correlate with organoleptic properties or nutritional values and often different species with different commercial values have the same markers compound, so chemical analyses cannot distinguish different species in commercial products (Raclariu et al., 2017).

Nevertheless, DNA is a universal molecule, specific and above all not influenced by external variables. For this reason, due to the reliability and reproducibility of DNA testing, is getting one of the most important analytical tools for food authentication. In the figure below was analysed the keywords used in 4280 publications with a word cloud generator. The size of the words is correlated with the frequency of being used as a keyword (Ulberth, 2020). DNA is the bigger word, meaning that DNA testing is the most used analysis in scientific literature.

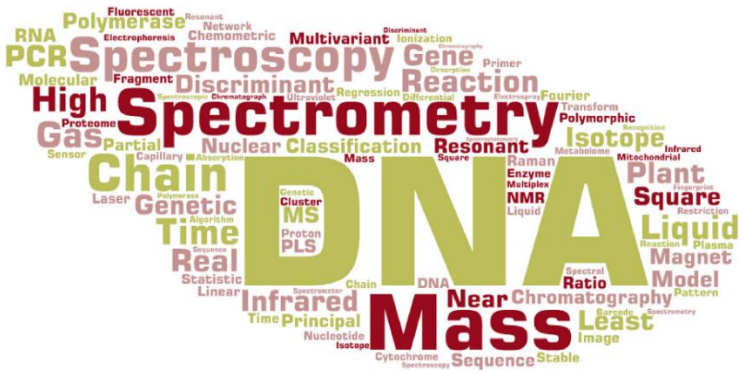


Figure 9. Analysis of analytical technology relevant keywords of 4280 publications with a word cloud generator (Ulberth, 2020).

Moreover, in the figure below is shown the absolute number of publications containing the keyword "DNA food" over the years from the database "Pubmed". There is a clear increase over the years, especially from 2003, when Paul Hebert presented for the first time the DNA barcoding analysis (Hebert et al., 2003).

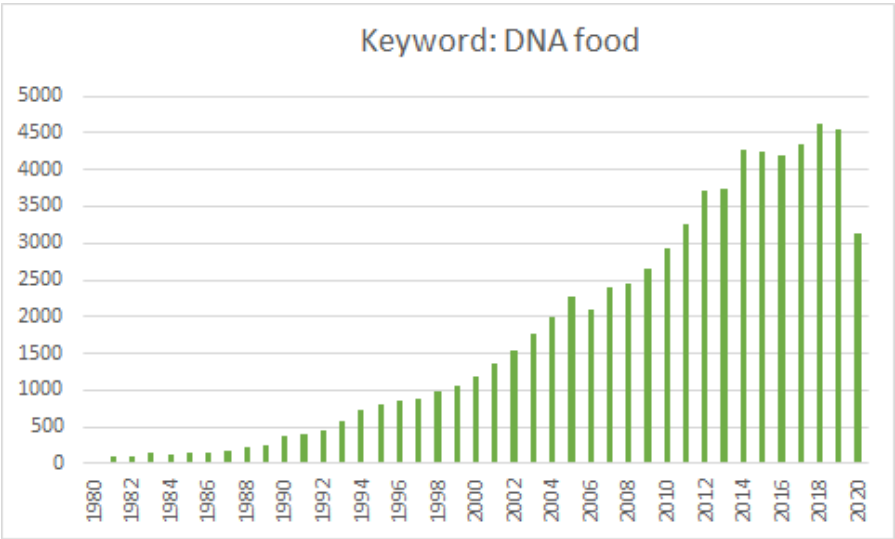


Figure 10. The absolute number of publications containing the keyword "DNA food" over the years from the database "Pubmed".

In the past were developed several DNA-based methods applied in the agri-food field, such as randomly amplified polymorphic DNA (RAPD) and microsatellites analysis (SSR) for variety and cultivar identification (Besnard at al., 2001; Caruso at al., 2007), quantitative PCR (qPCR), digital droplet PCR (ddPCR), isothermal amplification (LAMP, RPA, etc) and DNA barcoding for species identification and adulteration detection (Druml et al., 2015; De Castro et al., 2017), high resolution melting (HRM) for mutations and polymorphism detection (Druml and Markl, 2014) and finally High-Throughput Sequencing (HTS) for multispecies products analysis (Raclariu et al., 2018).

Figure 11. The most common DNA based analyses for food authentication.

2.3.2. DNA barcoding

DNA barcoding has been proposed as a standardized method by Paul Hebert in 2003 for species identification (Hebert et al., 2003). This method consists of sequencing and comparing orthologous DNA regions (Figure 12).

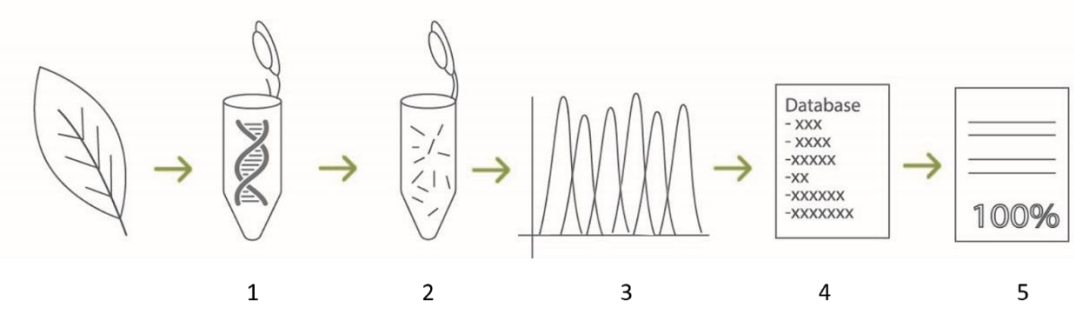


Figure 12. Methodological steps of DNA barcoding. The first step consists of DNA extraction (1). Subsequently, the region chosen is amplified by PCR (2). After visualization of the amplicon by using an agarose gel, the amplicon is sequenced by Sanger sequencing (3). Finally, the nucleotide sequence is compared to public databases (4) and the species is assigned to the most similar result (commonly In scientific literature the identity is ≥ 99%) (5).

DNA barcoding has the advantage of innovations: molecularization of identification processes, standardization of the procedure and computerization (Casiraghi et al., 2010). DNA barcoding is an important tool in food species identification and the supply chain traceability. One of the most challenging aspects of this analysis is searching for the most accurate gene that shows low variability within a certain taxon and a high level of interspecific variability. The most common regions are the gene for Cytochrome c oxidase I (COI) and Cytochrome b (cytb) for animal species (Hebert et al., 2003, Shen et al., 2013). Concerning the plants, are used the gene for maturase K (*matK*) and carboxylase (*rbcL*), the intergenic spacer of *psbA* and *trnH* genes (*psbA-trnH*) more recently the internal transcribed species (ITS) (Chen et al.,

2010; Newmaster and Subramanyam, 2017). The standard-length of amplicons are around 650 bp, but in processed products (such as canned fish) the amplification is challenging due to DNA degradation and fragmentation. In this context, a DNA mini-barcoding approach (100-200 bp) can successfully be applied (Shokralla et al., 2015).

The global project Barcode of Life System (BOLD) is a public database of barcodes for all species of life and has been given special attention to plants and fish. Due to a low intraspecies variability the discrimination between species belonging to the same genera is thus challenging (Böhme et al., 2019).

Despite this, DNA barcoding has been successfully applied in the agrifood sector for fraud and mislabelling identification. Seafood field is the most adulterated and DNA barcoding is a useful tool for fraud identification, in fresh, frozen and processed products (Delpiani et al., 2020; Xing et al., 2020; Barbosa et al., 2020; Calegari et al., 2020; Liou et al., 2020; Pardo et al., 2020; Deconinck et al., 2020). Several studies in the scientific literature show a high percentage of meat mislabelled such as sausage, ground meat and meat balls discovered by DNA barcoding technology (Kane et al., 2016; D'Amato et al., 2013; Hellberg et al., 2017).

The authenticity of plant-based food additives and supplements is another area of concern for consumers, highlighting the need for accurate methods to ensure quality. DNA barcoding has been applied to food authentication for herbal infusion (De Castro et al., 2017), spices (Parvathy et al., 2014), olive oil (Kumar et al., 2011), honey (Bruni et al., 2015), food supplements (Mosa et al., 2018) and poisonous plants (Mezzasalma et al., 2017; Cornara et al., 2018).

2.3.3. High-Throughput Sequencing (HTS)

High-Throughput Sequencing (HTS) or next generation sequencing (NGS) is a method that has revolutionized genomic research, because it can sequence millions of small fragments in parallel (Behjati and Tarpey, 2013). HTS technique can provide sequence data around a hundred times faster and cheaper than the conventional Sanger approach (Figure 13). Sequencers such as 454 Life Sciences/Roche (producing about a million sequences of length 800–1000 base pairs), Solexa/Illumina and Applied Biosystems SOLiD technology (producing over a billion sequences of length 50–500 base pairs) were produced as second-generation technologies and other competitive instruments appeared on the market such as the Ion Torrent and PacBio.

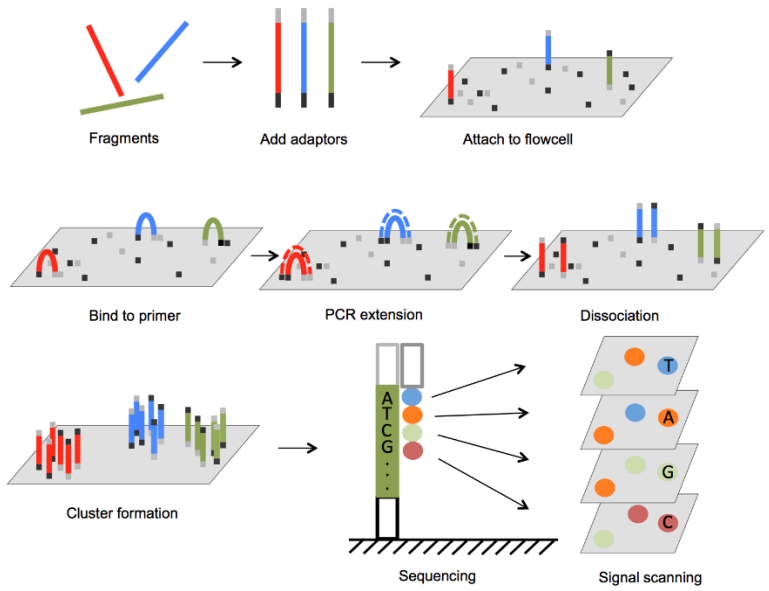


Figure 13. Methodological steps of Illumina sequencing process. By PCR, adaptors are added to the ends of sequence fragments. Single-stranded fragments are randomly bound to the inside surface of the flow cell channels and bridge PCR reactions amplify each bound fragment to produce clusters of fragments. During each sequencing cycle, one fluorophore

attached nucleotide is added to the growing strands. The laser excites the fluorophores in all the fragments that are being sequenced and an optic scanner collects the signals from each fragment cluster. Then the sequencing terminator is removed, and the next sequencing cycle starts. The data are aligned and compared to a reference, and sequencing differences are identified (Lu et al., 2015).

With the development of HTS, DNA metabarcoding has emerged. This technique combines DNA barcoding with HTS, a perfect approach for analysing multispecies products such as complex food matrices (Xing et al., 2019). In the last years, DNA metabarcoding have been used for food authentication and fraud detection in seafood products (De Battisti et al., 2014; Kappel at al., 2017), meat derived products (Bertolini et al., 2015; Xing et al., 2019), dairy products (Ribani et al., 2017;), candies (Muñoz-Colmenero et al., 2017), honey (Prosser and Hebert, 2017) and food supplements (Raclariu et al., 2018).

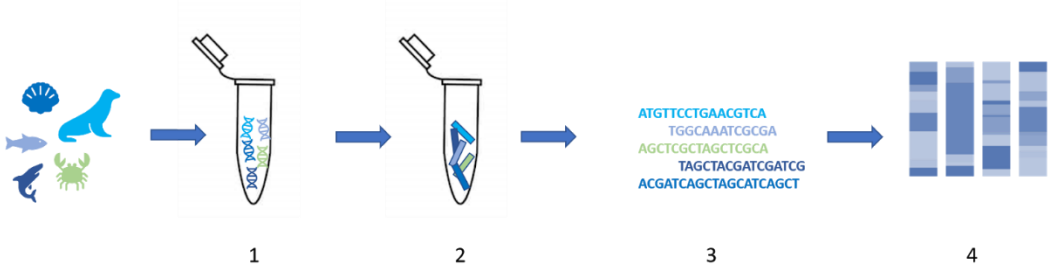


Figure 14. Methodological steps of metabarcoding. The first step consists of DNA extraction (1). Subsequently, the region chosen is amplified by PCR (2). After sequencing (3) all sample compositions are defined after bioinformatic analysis (4).

2.3.4. Isothermal technologies

DNA barcoding is a method used routinely in food authentication. Nevertheless, there are some critical aspects to consider for its application, including: long timescales, use of dedicated and sophisticated tools, since PCR requires sudden temperature changes (thermocyclers), visualization of the result requires DNA electrophoresis equipment and consequently scientific skills are needed to process the different phases.

Therefore, alternative methods have been developed that deal with these criticalities, specifically isothermal methods of DNA amplification such ah Loop-Mediated Isothermal Amplification (LAMP) and Recombinase Polymerase Amplification (RPA). These isothermal amplification techniques do not require a thermal cycler and can produce easily interpretable results, thus becoming accessible and applicable by unskilled personnel, even in the absence of equipped laboratories.

LAMP (loop-mediated isothermal amplification) is a technique that allows rapid amplification of DNA with high efficiency and specificity under isothermal conditions. It was first described in 2000 by Notomi and collaborators (Notomi et al., 2000) and the technique is based on the use of at least four primers (two pairs) which recognize six different regions of the target DNA. Thanks to the enzymatic activity of Bst DNA polymerase, the LAMP reactions can be performed in isothermal conditions and therefore it is enough to have a thermal block or water bath at a controlled temperature. Furthermore, the results are naked-eye visible, due to the change of colour of the reagents in case of amplification (see figure 15).

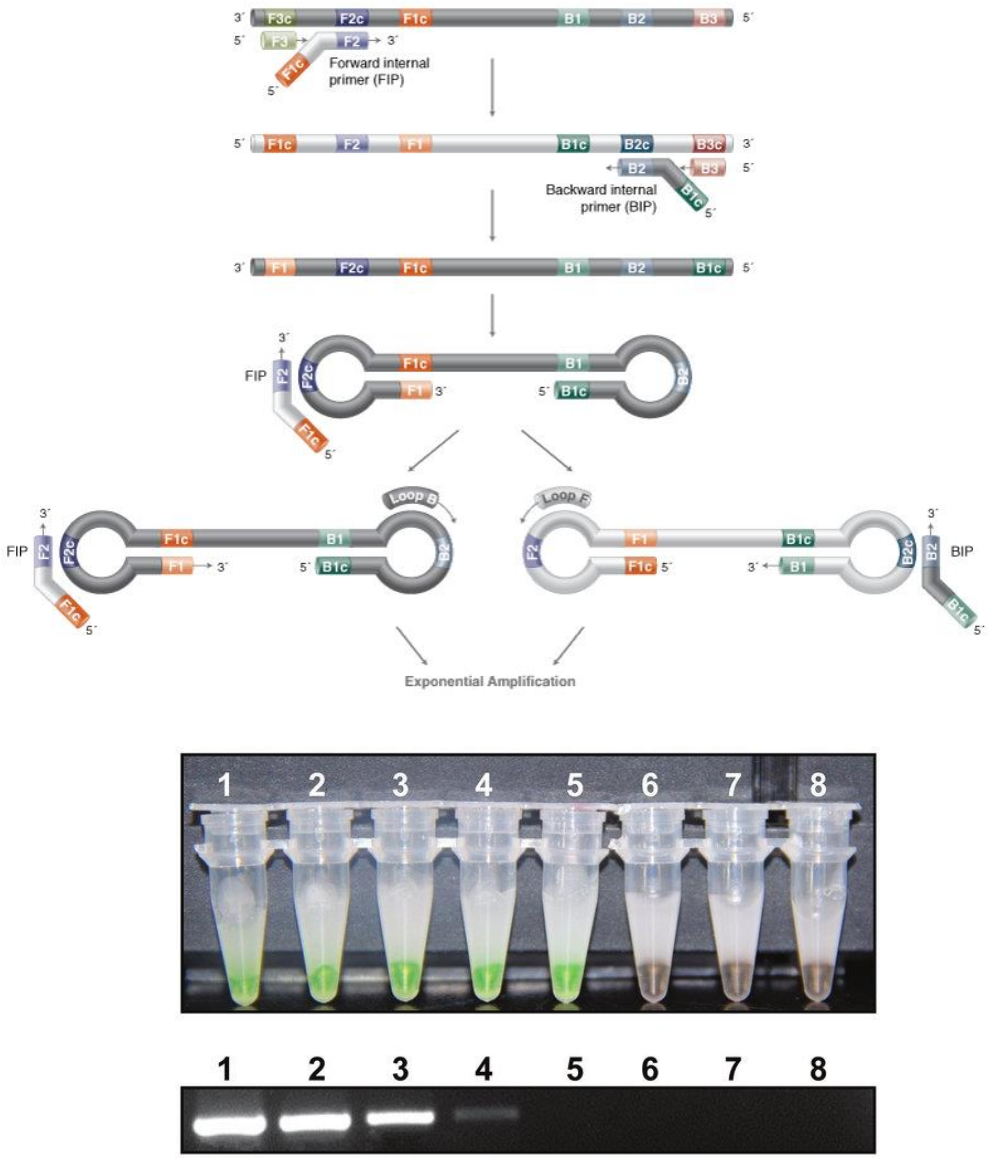


Figure 15. On the image above: Loop-mediated isothermal amplification (LAMP) uses 4-6 primers recognizing 6-8 distinct regions of target DNA. A strand-displacing DNA polymerase initiates synthesis and 2 of the primers form loop structures to facilitate subsequent rounds of amplification (Alhassan et al. 2015), on the image below: example of results visualisation (Li et al., 2015).

Recently, LAMP has been used for food authentication. It has been successful in spices and botanicals identification (Sasaki et al., 2007; Sasaki et al. 2008), and in allergen detection (Sheu et al., 2018).

The RPA (Recombinase Amplification Polymerase), developed by the biotech company TwistDx Ltd, uses a recombinase, a protein capable of catalyzing the hybridization process between small oligonucleotide primers and the homologous target region (Figure 16). In this way, the enzyme allows specific coupling between the primers and the double-stranded DNA target, thus directing the polymerase for the extension of the primers and the amplification of the DNA. During the RPA reaction, the primer-recombinase complex recognizes the specific site, the enzyme opens the double helix and allows the attachment of the primer to the target sequence. The primer is then extended, and the new filament produced undermines the original filament. The synthesized sequence can act as a template for the next cycle thus generating an exponential amplification. The reaction progresses rapidly and results in a detectable amplification of the target DNA in a few minutes (Piepenburg et al., 2006). The results are naked-eye visible thanks to the use of lateral flow (see figure 16).

RPA has been successfully applied for food authentication, such as herbal products (Liu et al., 2018) and spices (Mingming et al., 2019).

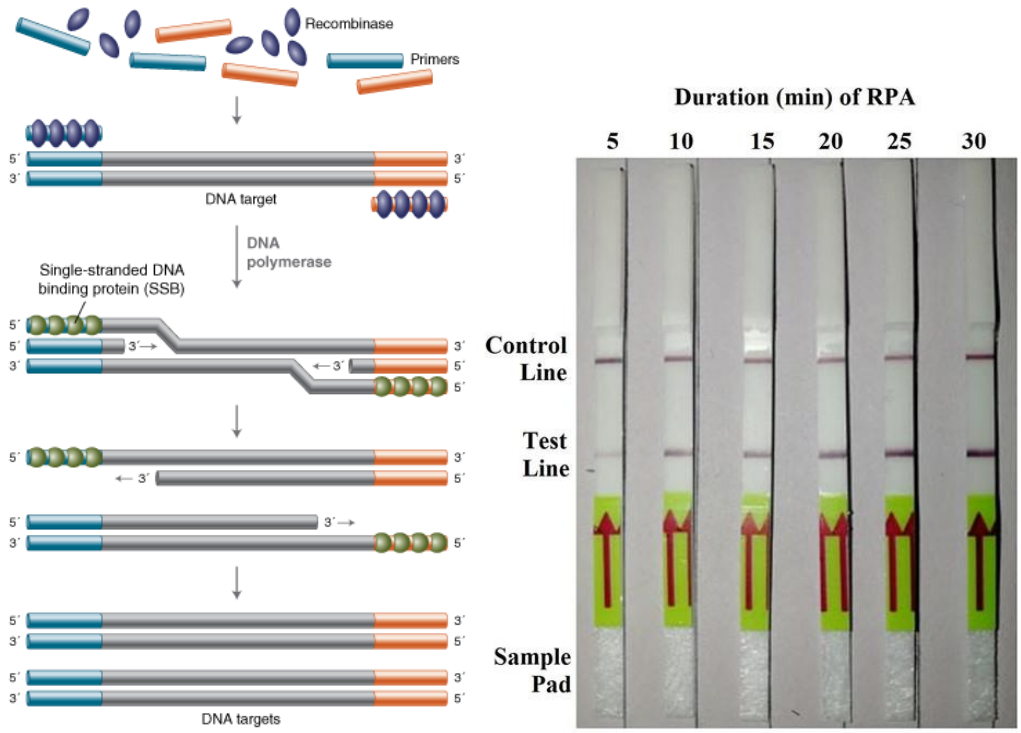


Figure 16. On the left: RPA utilizes a recombinase-primer complex, a strand-displacing polymerase, and single-stranded DNA binding proteins to amplify DNA (New England Biolabs); on the right: example of lateral flow visualisation (Dai et al. 2019).

Although isothermal techniques are fast, simple and cheap, the limitation is that they are not universal. If DNA barcoding allows identifying any unknown plant through an untargeted analysis, LAMP and RPA only allow a target analysis. Nevertheless, a ready-to-use kit could be a useful method for supply chain control.

2.4. Project aims and experimental design

Food fraud is an increasing problem and there is a demand from the consumer to better understand food products. Furthermore, due to globalization, the supply chain is long and wide. The product, from raw material to market, often undergoes several steps which need to be tracked. Often there is a correlation between the identity of a product with its quality and their organoleptic properties. Therefore, checking the identity of a raw material or a final product can ascertain its purity and quality. A suitable bio-molecular identification and traceability system could be used to assess the quality of raw materials up to the finished food products as well as to guarantee the consumer on the identity of the products purchased.

The main objective of this PhD project was to identify and test innovative biomolecular analyses for the identity assessment of raw materials and processed food products. These methods will be useful for assessing quality and safety in the agri-food industry from farm to fork. The stakeholders of this project are both companies, who want to guarantee high quality and safety products and final consumers, who have been increasingly careful about the food products. For this reason, the proposed methods must be simple, effective and possibly universal to be applied on a large scale.

FEM2-Ambiente, the company where the thesis project was carried out, currently offers a genetic identification service of species for food and herbal products: DNA barcoding. This analysis, although it is currently used as a routine analysis, has some defects.

Industrial treatments could alter DNA quality of raw material (Pecoraro et al., 2020; Raclariu et al., 2017; Parveen et al., 2016), therefore DNA barcoding could be challenging to apply. Examples of such products are canned fish (tuna, mackerel, sardines, etc.), highly processed meats (meat sauce, etc.) and food supplements. DNA

barcoding technique analyses large fragments of DNA, but in case of processed products DNA is fragmented and therefore small and DNA mini-barcoding could be more successful. For this reason, the first technical objective of this project was to develop a modification of the DNA barcoding approach to analyse highly processed products. Specifically, we evaluated whether these analyses can be applied to trace the plant component in food supplements from the starting raw material to the finished commercial products.

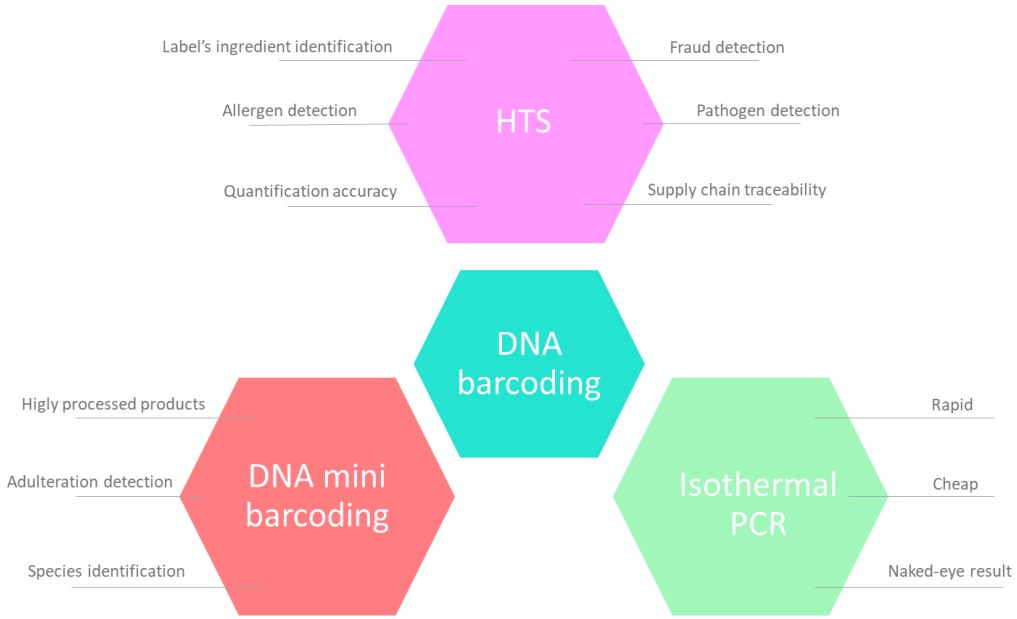


Figure 17. Experimental design.

A second limit that occurs in the food market is related to multispecies products. Many products contain a mix of different species, but DNA barcoding does not allow the analysis of multiple species at the same time. For this reason, the second objective of this thesis project was to test DNA metabarcoding (High-Throughput

Sequencing technique). This technique is capable of analysing several genomes at the same time and distinguishing them thus recognizing the species of the mixture of processed food products. Moreover, it was necessary to define bioinformatic tools to identify the best threshold for this analysis. We tried to answer these questions: Is it possible to identify all the species in a product? Can we quantify all the ingredients? Is it possible to identify allergens in low concentrations? Is it possible to identify pathogens in a product? Can we use the microbiome for supply chain traceability?

To achieve the first objective, we analysed innovative food complex matrices such as insect-based novel foods and herbal products like herbal teas. In order to test the ability of DNA metabarcoding as an early warning method of identifying allergens in processed products, samples with insect flour and different concentrations of allergens, more specifically gluten and soy, were created. With the aim to evaluate the accuracy of quantification, samples with different concentrations of five plant species were created. These samples were both created starting from raw plants (gr/gr) and genomic DNA (ng/ng). Finally, the microbiome of insect-based novel foods was analysed. The aim was to evaluate both the presence of pathogenic bacteria and to identify a bacterial pattern for the traceability of novel foods.

Further limitations of DNA barcoding are costs and time-consuming of analysis; it requires specialized personnel and very expensive tools.

To develop a rapid kit for carrying out on-site species identification analyses, a DNA amplification isothermal technique (LAMP) was tested on food products (truffle).

Overall, the obtained results allowed to develop innovative analysis to offer to the stakeholders, with the final goal to guarantee the traceability of the supply chain in the agri-food sector and assure high quality and safe products.

In the next chapter are presented scientific papers published in international journals with the results of the PhD project. Due to the desire to patent the ready-to-use kits, the results of the last part of the project have not been published.
I published five further works concerning food safety (Frigerio et al., 2019; Cornara et al., 2018) and food quality (Bruno et al., 2019; Cornara et al., 2020; Frigerio et al., 2020).

3. Publications

DNA barcoding to trace Medicinal and Aromatic Plants from the field to the food supplement

2019, Journal of Applied Botany and Food Quality 92:33-38

Jessica Frigerio a,b, Tommaso Gorini a, Andrea Galimberti a, Ilaria Bruni a, Nicola Tommasi a, Valerio Mezzasalm ab, Massimo Labra a*

a Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, 20126 Milano, Italy
b FEM2 Ambiente Srl, Piazza della Scienza 2, 20126 Milano, Italy

* Corresponding author: Massimo Labra, Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, 20126 Milan, Italy; Tel.: +39-02-64483472; E-mail address: massimo.labra@unimib.it

Key Words: Herbal products, ITS, Minibarcode, psbA-trnH, Phytoextracts, rbcL

Summary

The global market of food supplements is growing, along with consumers demand for high-quality herbal products. Nevertheless, substitution fraud and adulteration cases remain a common safety problem of global concern. In the last years, the DNA barcoding approach has been proposed as a valid identification method and it is now commonly used in the authentication of herbal and food products. The

objective of this study was to evaluate whether DNA barcoding can be applied to trace the plant species from the starting raw material to the finished commercial products. We selected a panel of 28 phytoextracts obtained through three different extraction methods (i.e., maceration, percolation and sonication) with different solvents (i.e., ethanol, deionized water and glycerol). Furthermore, we chose six plant species for which we collected and analysed all the intermediates of the industrial production. We sequenced and analyzed the sequence variability at DNA barcoding (*psbA-trnH, ITS*) and minibarcoding (*rbcL* 1-B) marker regions. Phytoextracts obtained through hydroalcoholic treatment, with the lower percentage of ethanol (<40%), and aqueous processing, at the lowest temperature, had major rate of sequencing and identification success. This study proves that DNA barcoding is a useful tool for Medicinal and Aromatic Plants (MAPs) traceability, that would provide consumers with safe and high-quality herbal products.

Introduction

Medicinal and Aromatic Plants (MAPs) and their preparations are products used in medicine, cosmetics and food industry, belonging to plants, fungi, algae or lichens (EFSA, 2009). Such products are prepared using plants or their parts to exploit their therapeutic and healthy properties (e.g., antioxidant, anti-inflammatory), as well as their flavor or scent (WHO 1999). According to a report published by the Persistence Market Research, the global market of herbal supplements had a value of USD 40 billion in the 2017 and is expected to reach a market valuation in excess of USD 65 billion by 2025 (PERSISTENCE MARKET RESEARCH, 2017). In the last years, the increasing consumption of natural food supplements and the growing awareness of consumers concerning the healthy benefits of these products have been progressively enhancing the market of MAPs (EFSA, 2009). Although most of the herbal products

used as food ingredients have been available to consumers since decades, the regulation of these products differs greatly among jurisdictions. While some countries consider MAPs as reliable ingredients for food production, others regulate them as healthy products or medicines. For example, in the European Union (EU), most products containing Medicinal and Aromatic Plants are sold as food supplements and regulated under the food law (SILANO et al., 2011); in Australia dietary supplements are considered medicinal products and in Canada they are subjected to the complex regulation of the Natural Health Products Directorate (NNHPD) of Health Canada (HEALTH CAN., 2015) as medical products (LOW et al., 2017). The lack of a clear and shared global regulation and the large market demand of high-quality plant-based items led to safety problems with the increase of substitution, fraud and adulteration cases. Anyway, more attention is required to guarantee an high quality level of MAPs which necessitates a stable raw material and its assurance. As a matter of fact, frequently, valuable plants are substituted with cheaper raw materials, such as the case of saffron substituted with safflower (BOSMALI et al., 2017). However, adulteration is not necessarily intentional, and herbal products may be altered due to inadvertent substitution, misidentification or confusion resulting from the use of different vernacular names in the countries of production.

According to the World Health Organization (WHO), the adulteration of herbal products is a potential threat to the consumers' safety. This condition opens two key issues which refer to the definition of suitable toxicological evaluations to estimate the risks for human health and the setup of an efficient identification system to trace the herbal products from the field to the traded MAPs items. Usually, macroscopic and microscopic examinations are the classic strategies adopted to verify the identity of fresh plants or the origin of plant portions. These tools can be used to trace the herbal products when the plants are processed immediately after being

harvested. However, when the herbs undergo drying, fragmentation and pulverization processes the morphological traits cannot be longer used to reliably assess the botanic source. Moreover, many herbal ingredients are obtained by infusion, maceration, distillation or pressing. In these cases, only dedicated chemical analyses of the complex mixtures could permit to achieve a reliable plant identification.

In the last years the DNA barcoding approach was proposed as a valid molecular identification method to provide species-level resolution and it is now more and more used in the authentication of taxonomic provenance of herbal and food products (NEWSMASTER et al., 2013; GALIMBERTI et al 2013; MOHAMMED et al., 2017). However, the most important limit of this molecular tool is that it can preferentially be adopted on unprocessed material (e.g., dry, fragmented and shredded plant portions) and several difficulties are encountered when dealing with extracts or with any other process that results in the degradation of the DNA. Recently, some manuscripts described the efficacy of minibarcode regions (i.e., the analysis of smaller genome portions - 100-150 bp - usually associated to the largest DNA barcodes) for the identification of processed plant extracts (RACLARIU et al., 2017, LITTLE, 2014). To date, any study addressed the efficacy of a DNA barcoding-based approach to trace herbal products along the entire production chain. In this work, we have selected Medicinal and Aromatic Plants in the form of phytoextracts obtained by several industrial companies and subjected to different kind of industrial processes and phytoextraction strategies. The objective of this survey was to evaluate whether or not the DNA barcoding approach (using standard barcodes or a minibarcode regions) could be applied to trace the plant species from the field to the finished commercial product in the case of food supplements. Therefore, we evaluated which are the industrial processes mostly affecting the efficacy of DNA analysis such as sample

pre-treatment methods, solvents used for extraction and which are the most suitable DNA markers to achieve a reliable MAPs traceability.

Material and methods

Study design

To test the efficacy of DNA traceability at the different steps of the industrial production chain of MAPs, we selected a panel of 28 plant commercial phytoextracts (Table 1) sold by three main European companies.

SAMPLES	**Industrial processing**		**DNA Extraction yield**		**DNA BARCODING MARKERS**		
	Phytoextraction Process	**Solvent**	**Value ng/µl**	**Standard Deviation**	***psbA - trnH***	***ITS***	***rbcL 1 - B***
Achillea millefolium L.	Sonication	EtOH < 40%	12.26	1.55			x
Echinacea pallida (Nutt.) Nutt.	Sonication		16.29	0.97			x

Harpagophytum procumbens (Burch.) DC. ex Meisn.	Sonication		12.94	1.58	x	x	x
Melissa officinalis L.	Percolation		44.63	1.32			x
Mentha x piperita L.	Sonication		12.3	0.83	x	x	x
Tilia platyphyllos Scop.	Sonication		9.78	1.28	x	x	x
Zingiber officinale Roscoe	Sonication		13.16	2.13	x	x	x
Arctium lappa L.	Maceration	EtOH > 40 %	1.23	0.2			
Echinacea angustifolia DC.	Maceration		2.83	0.5			
Melissa officinalis L.	Percolation		1.4	0.44			

Passiflora incarnata L.	Maceration		1.74	0.17	
Taraxacum officinale Weber ex F.H. Wigg.	Maceration		2.47	0.54	
Thymus vulgaris L.	Maceration		1.78	0.38	
Arctostaphylos uva-ursi (L.) Spreng.	Percolation	WATER	3.1	0.33	
Cetraria islandica (L.) Ach.	Percolation		2.27	0.94	
Echinacea purpurea (L.) Moench	Percolation		3.47	0.71	x
Epilobium angustifolium L.	Percolation		1.88	0.63	
Malva sylvestris L.	Percolation		2.34	0.69	

Arctostaphylos uva-ursi (L.) Spreng.	Sonication		13.36	1.05			x
Echinacea purpurea (L.) Moench	Sonication		46.41	0.81	x	x	x
Epilobium angustifolium L.	Sonication		14.78	0.97			x
Melissa officinalis L.	Sonication		12,73	0,76	x	x	x
Arctium lappa L.	Maceration	GLYCEROL	2.69	0.64			
Echinacea angustifolia DC.	Maceration		4.73	0.85			x
Melissa officinalis L.	Maceration		2.55	0.76			
Passiflora incarnata L.	Maceration		2.09	0.6			

Taraxacum officinale Weber ex F.H. Wigg.	Macera-tion	3.12	0.3
Thymus vulgaris L.	Macera-tion	2.71	0.8

Table 1: List of the analysed MAPs samples with details concerning their industrial processing to obtain the final phytoextracts. Average yield of DNA extraction (with stand-ard deviation) and assessment of positive sequencing of DNA barcoding markers (x) are also reported.

The selected items were obtained starting from 17 plant species (in the initial form of dried raw material) and were processed by the same companies adopting three main extraction procedures, namely maceration, sonication and percolation. Maceration consists in the solubilization of the plant material in different solvents like water and alcohol (e.g., ethanol), while percolation involves the slow descent of a solvent through the plant raw material until it absorbs the molecules of interest. Both methods rely on liquid filtration and concentration. Differently, the sonication provokes cellular cavitation and the release of the phytocomplexes in the used solvent. Three different extraction solvents were considered in this study and specifically, ethanol (i.e., alcoholic), deionized water (i.e., aqueous) and glycerol. During maceration the temperature was maintained under the threshold of 55 °C while in percolation higher temperatures (i.e., > 80 °C) were maintained, and sonication was mainly performed at 30-40 °C. After the percolation process, some phytoextracts are dried at very high temperatures (about 200 °C). Processing details for each tested phytoextract are shown in Table 1.

To evaluate the efficacy of DNA barcoding to trace the intermediates of industrial production after different steps of phytoextraction, we selected a panel of six commercial products obtained only by alcoholic and aqueous extraction procedures (Table 2).

Plant species		**Steps of the industrial production process**					
	Solvent	1	2	3	4	5	6
Achillea millefolium L.	20% Ethanol	Y	Y	N	Y	Y	Y
Zingiber officinale Roscoe	30% Ethanol	Y	Y	N	Y	Y	Y
Thymus vulgaris L.	60% Ethanol	Y	N	Y	N	N	-
Melissa officinalis L.	70% Ethanol	Y	N	Y	N	N	-
Echinacea purpurea (L.) Moench	Water	Y	Y	N	Y	Y	-

Melissa officinalis L.	Water	Y	Y	N	Y	Y	Y

Table 2: List of six commercial MAPs (phytoextracts) traced along their entire production chain. Each sample (intermediates of industrial production) was treated for DNA extraction and DNA barcoding analysis using the minibarcode region rbcL 1-B. Numbers indicate the industrial processing step as described in Figure 1. Y= correct plant identification by DNA barcoding at rbcL 1-B locus, N= DNA extraction or amplification failure, - = Sample was not collected and analysed.

For these samples, we collected and molecularly analysed through DNA barcoding, any intermediate of production (Figure 1).

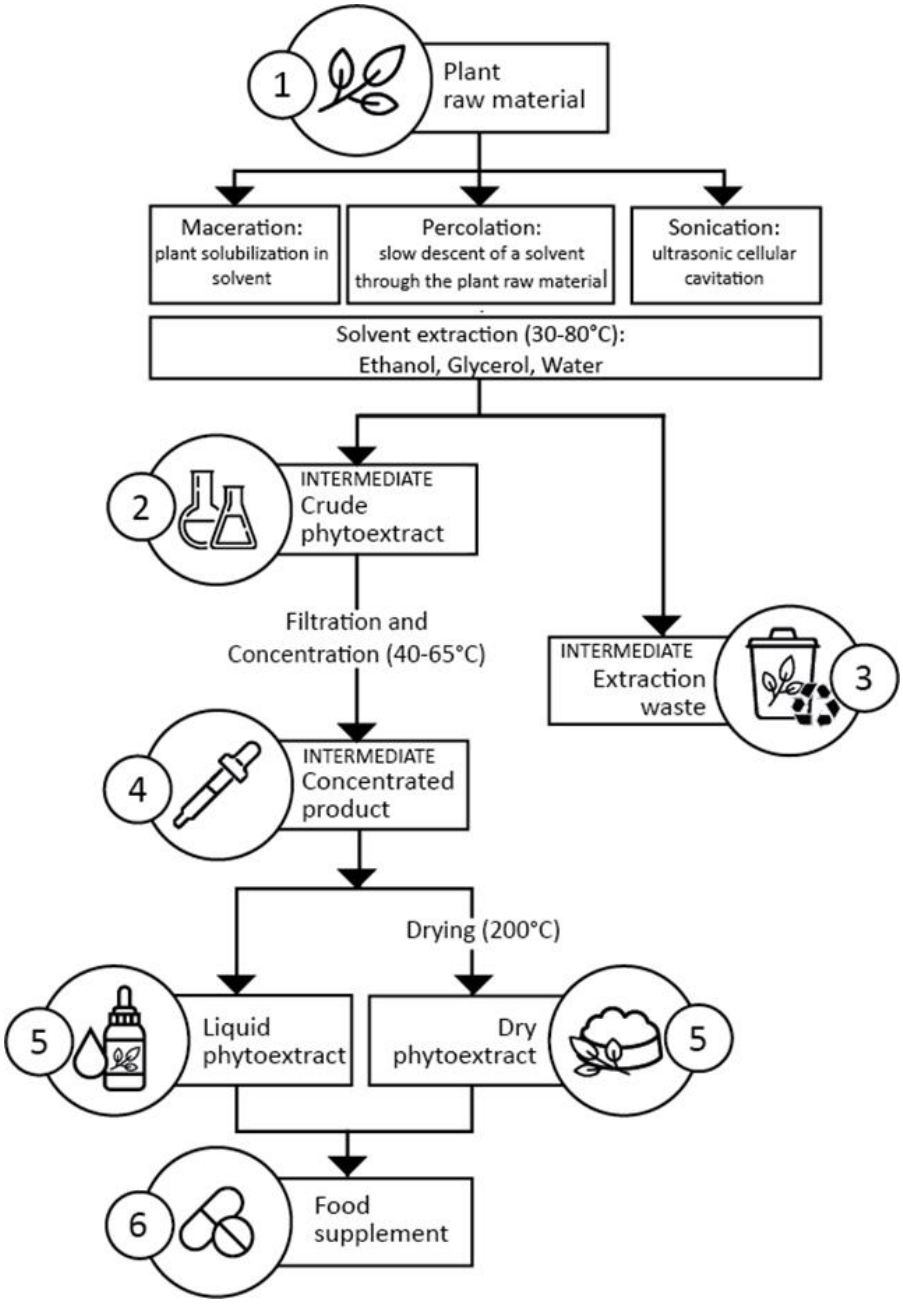


Figure 1: The industrial flowchart of MAPs production. The numbers indicate the intermediate steps of the industrial process for which the DNA barcoding efficacy has been verified (Table 2).

DNA Extraction and DNA barcoding analysis

All commercial products of Table 1 were tested for authenticity by sequencing three candidate markers, namely the standard DNA barcoding plastidial intergenic spacer psbA–trnH (STEVEN & SUBRAMANYAM, 2009), the nuclear ITS region (primers ITS p5–u4, CHENG et al., 2016) and the minibarcode region *rbcL* 1-B (LITTLE, 2014). Primer details and size of amplified fragments are provided in Table A.1.

A total of 50 mg of dried plant raw material, 150 µL of phytoextract (and intermediate products of phytoextraction, see Table 2) were treated for DNA extraction by using the EuroGOLD Plant DNA Mini Kit (Euroclone, Pero, Italy). Each commercial phytoextract product was subjected to DNA extraction in three replicates. Purified DNA concentration of each sample was estimated fluorometrically by using NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer (Thermo Scientific™).

A PCR amplification for each candidate marker was performed using puReTaq Ready-To-Go PCR beads (GE Healthcare Life Sciences, Italy) in a 25 µL reaction according to the manufacturer's instructions containing 1 µL 10mM of each primer and up to 3 µL of DNA template. PCR cycles consisted of an initial denaturation step for 7 min at 94 °C, followed by 35 cycles of denaturation (45 s at 94 °C), annealing (30 s at different temperatures; see Table A.1) and extension (1 min at 72 °C), and, hence, a final extension at 72 °C for 7 min.

In the case of the intermediates of production listed in Table 2, we amplified and sequenced only the minibarcode locus *rbc*L 1B.

Amplicons occurrence was assessed by electrophoresis on agarose gel using 1.5% agarose TAE gel stained with ethidium bromide and amplicon length was measured by comparison against 100 bp ladder. When a sample did not produce any band or showed multiple or non-specific amplicons, the reaction was repeated increasing

the amount of template DNA up to 10 µL. Purified amplicons were bidirectionally sequenced using an ABI 3730XL automated sequencing machine at Eurofins Genomics (Ebersberg, Germany). The 3' and 5' terminal portions of each sequence were clipped to generate consensus sequences for each sample. After manual editing, primer removal and pairwise alignment, the obtained sequences for dried raw material were submitted to the international GenBank through the EMBL platform (see Table A.2 for accession numbers).

For all the tested samples (Table1 and Table 2), the reliability of DNA barcoding identification was assessed by adopting a standard comparison approach against a GenBank database with BLASTn. Each barcode sequence was taxonomically assigned to the plant species with the nearest matches (maximum identity >99% and query coverage of 100%) according to BRUNI et al. (2015). We performed the identification separately for the three markers.

Results and discussion

Good DNA quality (i.e., A260/A230 and A260/A280 absorbance ratios within the range 1.8 - 2.2) and extraction yield (20-40 ng/µl) were obtained from all the 17 raw material samples. The three candidate genetic markers exhibited high PCR success and the obtained PCR products were successfully sequenced with high-quality bidirectional sequences. The BLASTn analysis suggested that all the obtained sequences corresponded with 100% maximum identity to the species declared by each company.

We are aware that multiple cases of 100% maximum identity within the same plant genus could occur, especially concerning the DNA minibarcoding *rbcL* 1B region. For example, this plastid region failed in discriminating *Echinacea purpurea (L.) Moench* from congenerics like *Echinacea angustifolia DC.* or *Echinacea pallida (Nutt.) Nutt.*

Such events suggest that in some conditions, the main limit of DNA minibarcoding relies on the reduced discrimination power among congenerics but it allows to detect plant contaminations when the adulterant/s belong to genera different from the target one. Nevertheless, it should be considered that when DNA content is expected to be low (of or low quality), the use of shorter DNA barcoding regions offers the best compromise between amplification universality, sequence quality and taxonomic discrimination (LITTLE, 2014).

Concerning the 17 dry raw material samples, our results agree with the assumptions of NEWMASTER and co-workers (2013) who suggested that a DNA barcoding approach could be successfully applied to verify the identity of commercial herbal products and to reveal cases of contamination or substitution. Therefore, when herbal products are directly used as ingredients of complex food, medicine or cosmetics items, they are subjected to "soft" processing actions such as cleaning, drying and cutting and the DNA barcoding (achieved using long barcode fragments > 300 bp) represents a useful tool to trace plant species during the processing (DE MATTIA et al., 2011).

As expected, the efficacy of DNA extraction and amplification decreased when we analyzed the 28 commercial phytoextracts and their intermediates of industrial processing. Overall, the DNA amount obtained after extraction processes ranged from 1.5 to more than 40 ng/µL (Table 1). The extracts obtained through hydroalcoholic treatment, with the lower percentage of ethanol (<40%), and aqueous processing, at the lowest temperature, contained more DNA than the other samples (Table 1 and Table 2).

In the samples where the DNA barcoding analysis worked well, no contamination or adulteration (i.e., the occurrence of DNA barcodes of other plant species) were observed. Unfortunately, in some groups of extracts, the molecular analysis did not

provide reliable DNA extraction or high-quality sequences. At the technical level, we hypothesize that in general, the high concentration of ethanol used in the industrial processing steps lead to DNA precipitation. This was confirmed by the data reported in Table 2, where samples of *Thymus vulgaris* L. and *Melissa officinalis* L. processed with ethanol at high concentration, showed residual DNA in the extraction waste rather than in the phytoextract. For this reason, both the DNA extraction and DNA barcoding authentication failed when applied to the successive intermediate products of industrial processing and DNA is no longer available in the final herbal supplements. In this case, we conclude that for this kind of industrial production, a DNA-based approach is not suitable to achieve a reliable traceability of the initial plant raw material. Similarly, high temperatures of water during aqueous extraction, followed by a drying step (about 200 °C) probably lead to DNA fragmentation and degradation (KARNI et al., 2013) as observed in five of the samples processed with a percolation procedure (Table 1). Conversely, the use of more lukewarm water (i.e., < 55° C) allows to achieve a successful DNA extraction, amplification and sequencing of DNA barcoding markers (Table 1). Moreover, such conditions also allow the traceability of the intermediate products of industrial processing as observed for *Echinacea purpurea* (L.) Moench and *Melissa officinalis* L. extracted using water as solvent.

Concerning glycerol extracts, although this solvent does not act directly on DNA molecules, it usually contains ethylhexylglycerin and phenoxyethanol, which are typically used as additives. According to Langsrud and co-workers (2016) these antibacterial agents could be responsible for DNA leakage. For this reason, also the analysis of the DNA minibarcode region did not produce amplicons in glycerine extracts (Table 1).

Concerning the industrial treatments, the sonication seems to keep the DNA of raw materials more intact than the other processes (i.e., maceration and percolation). Our results also show that sonicated samples contained higher amounts of DNA (i.e. from 9.73 to 44 ng/µl, Table 1) compared to the other categories, thus allowing a successful amplification and sequencing of the DNA minibarcode marker.

Concerning the quality of extracted genetic material, the purity of DNA is more important than the extraction yield to achieve a good amplification and then a reliable identification (SONG et al., 2017). It should also be considered that secondary metabolites, like polyphenols and polysaccharides, which are normally extracted along with DNA, may interfere with PCR amplification (SAHU, THANGARAJ, & KATHIRESAN, 2012). These molecules could bind DNA covalently and make the extraction products impure, with several problems for the successive molecular analysis. For example, tannic acids could bind and inactivate Taq polymerase (OPEL, CHUNG, & MCCORD, 2010). However, in our analysis we hypothesize that the main amplification problem for the phytoextracts is the fragmentation of DNA. In all the tested cases, the DNA minibarcode locus *rbcL*1-B was most easily amplified and sequenced (Table 1) than the other two DNA barcoding markers. This suggests that the DNA obtained from phytoextracts are richer in small DNA fragments (80-200 bp). Such condition is in line with the data reported in recent review articles (MOHAMMED et al., 2017) suggesting that DNA barcoding is a reliable and suitable technique only for the herbal product that preserve a good quality DNA and with poor fragmentation. In the other cases DNA minibarcoding is the most efficient and reliable tool for traceability purposes (SONG et al., 2017).

Nowadays, analytical chemistry methods (TLC, HPLC) represent the most used tools to verify the quality of MAPs, however, these approaches are usually directed to define the concentration of specific bioactive molecules or to estimate chemical

contaminants (e.g., heavy metals) rather than to identify the occurrence of plant contaminants (SGAMMA et al., 2017). Conversely, the DNA barcoding approach is globally recognized as one of the most reliable DNA-based approaches to identify species if a well populated reference dataset of DNA barcode sequences for the target taxa is available (GALIMBERTI et al., 2013). Moreover, in the case of contamination (or substitution), DNA analyses also tallow to simultaneously identify any species (i.e., DNA metabarcoding) using High Throughput Sequencing (HTS) sequencing systems (GALIMBERTI et al., 2015; MEZZASALMA et al., 2017). For these reasons, the Pharmacopoeia guidelines of some countries such as that of UK (BRITISH PHARMACOPOEIA COMMISSION, 2017) indicate the DNA barcoding as one of the official traceability systems in the sector of herbal products. Our data support this proposal and the ability of DNA minibarcode makers to provide a reliable tracing of the intermediate products of industrial production. However, it is important to underline that some industrial processes demanding high temperatures and the use of solvents, such as a high concentration of ethanol, can induce DNA degradation and make this molecular tool less effective.

In conclusion, this study leads to two main considerations about the future application of DNA barcoding as a quality control tool in the sectors where the Medicinal and Aromatic Plants constitute relevant ingredients (e.g., food, cosmetics and pharmacology). First of all, the current industrial trends promote the adoption of extraction processes from plant raw material, which rely on the reduction of energy consumption (i.e., low temperatures), and on the use of more 'green' solvents (e.g., water) to obtain exhausted waste products that can be used in other supply chains (e.g., fertilizers). The adoption and spread of this trend should lead to an increased integrity and quality of DNA in MAPs (and related intermediate products) and

therefore to enhance the success of DNA barcoding as a universal traceability system.

Secondly, the continuous advances in High Throughput Sequencing and the resulting possibility of exploring multiple short genetic regions simultaneously (i.e. 150-200 bp), could increase the sensitivity of a DNA-based identification.An HTS-DNA metabarcoding approach would allow to check the presence of several plant contaminants in the same sample, even if occurring at low concentrations (NEWMASTER et al., 2013). SGAMMA and co-workers (2017) proposed the introduction of DNA metabarcoding to evaluate the quality and authenticate herbal drug material in the industrial context. The authors proposed a dedicated DNA barcoding flowchart for industrial traceability purposes. Our results could be taken into account to improve this flowchart and to also adapt it to the traceability of intermediates of industrial production. Interestingly, VALENTINI and co-workers (2017) recently proposed an innovative nanoparticle-DNA barcoding hybrid system called NanoTracer that could potentially revolutionize the world of traceability as it allows for rapid and naked-eye molecular traceability of any food and requires limited instrumentation and cost-effective reagents.

This and other similar applications (AARTSE et al., 2017) open the opportunity to really boost the issue of herbal supplements traceability, not only with the industrial actors as the main stakeholders, but also involving the a wider cicle of specialists.

Acknowledgments

This work was supported by Regione Lombardia in the framework of the Program 'Accordi per la ricerca e l'innovazione', project name: Food Social Sensor Network Food NET, grant number: E47F17000020009. The funder had no role in conducting the research and/or during the preparation of the article. FEM2-Ambiente s.r.l. provided support in the form of a salary for authors Jessica Frigerio and Valerio Mezzasalma but did not play a role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript and only provided financial support in the form of research materials.

References

AARTSE, A., SCHOLTENS, I. M. J., VAN DER A, H. J. G., BOERSMA-GREVE, M. M., PRINS, T. W., VAN GINKEL, L. A., … BOVEE, T. F. H., 2017: Evaluation of a loop-mediated isothermal amplification (LAMP) method for rapid on-site detection of horse meat. Food Control, 81, 9-15. https://doi.org/10.1016/j.foodcont.2017.05.025

BOSMALI, I., ORDOUDI, S. A., TSIMIDOU, M. Z., & MADESIS, P., 2017: Greek PDO saffron authentication studies using species specific molecular markers. Food Res. Int. 100(Pt 1), 899–907. https://doi.org/10.1016/j.foodres.2017.08

BRUNI, I., GALIMBERTI, A., CARIDI, L., SCACCABAROZZI, D., DE MATTIA, F., CASIRAGHI, M., & LABRA, M., 2015: A DNA barcoding approach to identify plant species in multiflower honey. Food Chem. 170, 308–315. https://doi.org/10.1016/j.foodchem.2014.08.060

CHENG, T., XU, C., LEI, L., LI, C., ZHANG, Y., & ZHOU, S., 2016: Barcoding the kingdom Plantae: new PCR primers for ITS regions of plants with improved universality and

specificity. Mol. Ecol. Resour. 16(1), 138–149. https://doi.org/10.1111/1755-0998.12438

DE MATTIA, F., BRUNI, I., GALIMBERTI, A., CATTANEO, F., CASIRAGHI, M., & LABRA, M., 2011: A comparative study of different DNA barcoding markers for the identification of some members of Lamiacaea. Food Res. Int. 44(3), 693-702. https://doi.org/10.1016/j.foodres.2010.12.032

GALIMBERTI, A., BRUNO, A., MEZZASALMA, V., DE MATTIA, F., BRUNI, I., & LABRA, M., 2015: Emerging DNA-based technologies to characterize food ecosystems. Food Res. Int. 69, 424–433. https://doi.org/10.1016/J.FOODRES.2015.01.017

GALIMBERTI, A., DE MATTIA, F., LOSA, A., BRUNI, I., FEDERICI, S., CASIRAGHI, M., … LABRA, M., 2013: DNA barcoding as a new tool for food traceability. Food Res. Int. 50(1), 55–63. https://doi.org/10.1016/J.FOODRES.2012.09.036

EFSA, 2009: Guidance on Safety assessment of botanicals and botanical preparations intended for use as ingredients in food supplements. EFSA Journal, 7(9). https://doi.org/10.2903/j.efsa.2009.1249

HEALTH CAN., 2015: Natural and Non-prescription Health Products Directorate (NNHPD). Updated on May 1, Health Can., Ottawa. http://www.hc-sc.gc.ca/dhp-mps/prodnatur/index-eng.php

KARNI, M., ZIDON, D., POLAK, P., ZALEVSKY, Z., & SHEFI, O., 2013: Thermal Degradation of DNA. DNA Cell Biol. 32(6). https://doi.org/10.1089/dna.2013.2056

LANGSRUD, S., STEINHAUER, K., LÜTHJE, S., WEBER, K., GORONCY-BERMES, P., & HOLCK, A. L., 2016: Ethylhexylglycerin Impairs Membrane Integrity and Enhances the Lethal

Effect of Phenoxyethanol. PLOS ONE, 11(10), e0165228. https://doi.org/10.1371/journal.pone.0165228

LITTLE, D. P., 2014: A DNA mini-barcode for land plants. Mol. Ecol. Resour. 14(3), 437–446. https://doi.org/10.1111/1755-0998.12194

LOW, T. Y., WONG, K. O., YAP, A. L. L., DE HAAN, L. H. J., & RIETJENS, I. M. C. M., 2017: The Regulatory Framework Across International Jurisdictions for Risks Associated with Consumption of Botanical Food Supplements. Compr. Rev. Food Sci. F. 16(5), 821–834. https://doi.org/10.1111/1541-4337.12289

MEZZASALMA, V., GANOPOULOS, I., GALIMBERTI, A., CORNARA, L., FERRI, E., & LABRA, M., 2017: Poisonous or non-poisonous plants? DNA-based tools and applications for accurate identification. Int. J. Legal Med. 131(1), 1–19. https://doi.org/10.1007/s00414-016-1460-y

MOHAMMED ABUBAKAR, B., MOHD SALLEH, F., SHAMSIR OMAR, M. S., & WAGIRAN, A., 2017: Review: DNA Barcoding and Chromatography Fingerprints for the Authentication of Botanicals in Herbal Medicinal Products. Evid-Based Compl. Alt. 2017, 1–28. https://doi.org/10.1155/2017/1352948

NEWMASTER, S. G., GRGURIC, M., SHANMUGHANANDHAN, D., RAMALINGAM, S., & RAGUPATHY, S., 2013: DNA barcoding detects contamination and substitution in North American herbal products. BMC Med. 11(1), 222. https://doi.org/10.1186/1741-7015-11-222

OPEL, K. L., CHUNG, D., & MCCORD, B. R., 2010: A Study of PCR Inhibition Mechanisms Using Real Time PCR. J. Forensic Sci. 55(1), 25–33. https://doi.org/10.1111/j.1556-4029.2009.01245.x

PERSISTENCE MARKET RESEARCH, 2017. Global market study on botanical supplements: Drugs application segment to hold maximum value share during 2017–2025. https://www.prnewswire.com/news-releases/global-market-study-on-botanical-supplements-drugs-application-segment-to-hold-maximum-value-share-during-2017---2025-300484161.html

RACLARIU, A. C., PALTINEAN, R., VLASE, L., LABARRE, A., MANZANILLA, V., ICHIM, M. C., … DE BOER, H., 2017: Comparative authentication of Hypericum perforatum herbal products using DNA metabarcoding, TLC and HPLC-MS. Sci. Rep. 7(1), 1291. https://doi.org/10.1038/s41598-017-01389-w

SAHU, S. K., THANGARAJ, M., & KATHIRESAN, K., 2012: DNA Extraction Protocol for Plants with High Levels of Secondary Metabolites and Polysaccharides without Using Liquid Nitrogen and Phenol. ISRN Mol. Biol. 2012, 1–6. https://doi.org/10.5402/2012/205049

SGAMMA, T., LOCKIE-WILLIAMS, C., KREUZER, M., WILLIAMS, S., SCHEYHING, U., KOCH, E., … HOWARD, C., 2017: DNA Barcoding for Industrial Quality Assurance. Planta Med. 83(14/15), 1117–1129. https://doi.org/10.1055/s-0043-113448

SILANO, V., COPPENS, P., LARRAÑAGA-GUETARIA, A., MINGHETTI, P., & ROTH-EHRANG, R., 2011: Regulations applicable to plant food supplements and related products in the European Union. Food Funct. 2(12), 710. https://doi.org/10.1039/c1fo10105f

SONG, M., DONG, G.-Q., ZHANG, Y.-Q., LIU, X., & SUN, W., 2017: Identification of processed Chinese medicinal materials using DNA mini-barcoding. Chin. J. Nat. Med. 15(7), 481–486. https://doi.org/10.1016/S1875-5364(17)30073-0

STEVEN G., N., & SUBRAMANYAM, R., 2009: Testing plant barcoding in a sister species complex of pantropical Acacia (Mimosoideae, Fabaceae). Mol. Eco. Resour. 9, 172–180. https://doi.org/10.1111/j.1755-0998.2009.02642.x

VALENTINI, P., GALIMBERTI, A., MEZZASALMA, V., DE MATTIA, F., CASIRAGHI, M., LABRA, M., & POMPA, P. P., 2017: DNA Barcoding Meets Nanotechnology: Development of a Universal Colorimetric Test for Food Authentication. Angew. Chem. Int. Ed. 56(28), 8094–8098. https://doi.org/10.1002/anie.201702120

BRITISH PHARMACOPOEIA COMMISSION, 2015: Working Party (DNA): Identification Techniques SUMMARY MINUTES. https://www.pharmacopoeia.com/file/DNA---January-2016.pdf

WHO, 1999: Monographs on selected medicinal plants. http://apps.who.int/medicinedocs/pdf/s2200e/s2200e.pdf

The hidden 'plant side' of insect novel foods: a DNA-based assessment

2020. Food Research International, 128: 108751

Running Title: Molecular authentication of novel foods.

Author names and affiliations:

Jessica Frigerio[a]^, Giulia Agostinetto[b]^, Anna Sandionigi[b], Valerio Mezzasalma[a], Nadia Maria Berterame[b], Maurizio Casiraghi[b], Massimo Labra[b], Andrea Galimberti[b]*

[a] FEM2-Ambiente, Piazza della Scienza 2, I-20126 Milano, Italy.

[b] Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, I-20126 Milano, Italy.

^ Both authors contributed equally to this work

*** Corresponding author:**

Dr. Andrea Galimberti

Department of Biotechnology and Biosciences, University of Milano - Bicocca,

Piazza della Scienza 2, I-20126 Milano, Italy

E-mail: andrea.galimberti@unimib.it

Tel: +39 0264483412

Abstract

In the context of novel foods, a category for which the market demand is increasing worldwide, the consumption of edible insects and related insect-based products is expected to grow in the next years. Insects represent an important source of energy for the human diet but there is a lack of scientific knowledge about their processing to ensure safe food items to the consumer. In this study we adopted a combined DNA-based approach to verify the identity of the declared species in five categories of commercial insect-based products (mt *COI* DNA barcoding) and to characterize plant declared ingredients or contaminants (nu ITS2 DNA metabarcoding) with particular attention to putative elements of allergenic concern belonging, for example to the insect rearing substrate. Moreover, the same approach has been used to assess its sensitivity to cases of contamination and counterfeits to insect flour with low cost (and potentially allergenic) vegetable flours like wheat and soybean. Results show the success of insect DNA barcoding authentication even for highly processed products. Furthermore, the DNA metabarcoding analysis revealed a high efficacy as a screening method to identify both plant ingredients and vegetal traces belonging to insect farming or possible adulteration events, also acting as an early warning strategy for the occurrence of allergens of human concern. This approach could support the development of new risk assessment procedures for novel foods by regulatory authorities to ensure their quality, safety, and acceptance which will become more required in order to face the challenge of feeding the world population in the next decades.

Keywords

DNA barcoding; DNA metabarcoding; High-Throughput Sequencing; Insect flour; Plant allergens; Processed food

1. Introduction

Traditionally, edible insects are consumed in large parts of the world like Africa and Asia. In the last few years, they have increased in popularity as trendy foods in many Western countries (Sun-Waterhouse et al., 2016). Being rich in essential nutrients, they represent an important source of energy for human diets (Rumpold, & Schülter, 2013). Mean estimates show that the energy level of insects is around 400– 500 kcal per 100 g of dry matter, making it comparable with other protein sources (Payne, Scarborough, Rayner, & Nonaka, 2016). Protein is probably the most significant component of edible insects, with an average value ranging from 30% to 65% of the total dry matter. However, edible insects are also rich in micronutrients such as iron, zinc, and calcium (Dobermann, Swift, & Field, 2017), and preliminary studies have shown that insect farming has a lower environmental footprint compared to other livestock animals (Oonincx, & de Boer, 2012). Moreover, a recent study demonstrated that edible insects could represent a potential source of antioxidants with positive effects on human health (Di Mattia, Battista, Sacchetti & Serafini, 2019).

In very recent years, the use of insects in food fortification is emerging as a means of producing nutritious and acceptable food products for human consumption (Myers, & Pettigrew, 2018). In many countries a lot of insect-based products have recently become commercialized (e.g. pasta, biscuits, and energy bars based on insect flour in combination with fruits, nuts, and other ingredients), and their consumption trend is expected to grow steadily.

In Europe, edible insects are placed in the category of novel foods, and since the beginning of 2018, the Regulation (EU) 2015/2283 entered into force in attempt to regulate the production and safety of novel foods in Europe. This regulation establishes the requirements that enable Food Business Operators to bring new foods

into the EU market, while ensuring high levels of food safety to the consumers. However, concerning the possible risks for human health caused by insect-based products, the European Food Safety Authority has published an initial assessment (EFSA, 2015; SLU 2018), and it concluded that attention should be placed to the possible occurrence of biological and chemical hazards caused by novel foods.

Insects ingredients may cause allergic symptoms (Mazzucchelli et al., 2018; Pali-Schöll et al., 2019; Van Der Fels-Klerx, Adamse, Punt, & Van Asselt, 2018), and there is a cross-reactivity/cosensitisation between edible insects and crustaceans (Ribeiro, Cunha, Sousa-Pinto, & Fonseca, 2018). Moreover, at the microbiological level, the microbiota of insects is highly complex and, apart from the body surface and the mouthparts, the maximum microorganism diversity is in the mid-gut with poor or no data about its effect on consumer health (Schlüter et al., 2017; Walia, Kapoor, & Farber, 2018).

Another element that can negatively impact the safety of insect-based novel foods is the composition and quality of feed used to raise the insects and their rearing conditions. In many cases, these feeds are composed of vegetables and very few ingredients of animal origin, such as fishmeal and egg and milk based-products (EU No. 2017/ 893). However, to date no analytical systems are available to control the diet of marketed insects.

In general, according to a recent review (Schlüter et al., 2017), there is a lack of scientific knowledge about insect processing to ensure safe novel foods. Moreover, most insect-based products are consumed as flour or processed items (e.g. pasta and bars); therefore the insect morphological traits cannot be used to verify the product authenticity and consequently its safety. Considering the high price of insect flours, we cannot exclude the occurrence of deliberate or intentional

counterfeits (i.e. mix with low-cost flours, such as maize) as happens with other high-value food products, such as saffron (Petrakis, Cagliani, Polissiou, & Consonni, 2015).

Based on these assumptions, in this study we adopted, for the very first time, a DNA-based approach to analyse commercial novel food products to verify the identity of declared insect species and characterize trace amounts of plant material occurring in the same products in order to derive the source of the substrate, such as from insect diets and litter, with particular attention paid to putative elements of allergenic concern. Particularly, we used a region of the mtDNA *COI* marker to identify the insect species (DNA barcoding), and the nuclear ITS2 region to characterize insect diets and the vegetal composition of the tested products (High-Throughput Sequencing HTS DNA metabarcoding). Moreover, the same approach can be used to verify product contamination and counterfeiting insect flour with low cost vegetable flours. We prepared six mock mixtures composed of wheat and soybean flours at different concentrations to also verify the limit of detection of our approach in the context of possible contamination of insect-based novel foods.

2. Materials and methods

2.1. Insect commercial food products

A total of 13 commercial insect-based products, namely flour (n=3), pasta (n=3), crackers (n=2), protein bars (n=4), and pet food (n=1) were purchased via e-commerce from six different companies (Table 1). These categories offer an almost complete representation of the insect novel foods available in Europe. Based on the label information, these products contained only one insect species each, for a total of five species belonging to the orders Orthoptera (*Acheta domesticus* and *Gryllodes sigillatus*), Diptera (*Hermetia illucens*) and Coleoptera (*Alphitobius diaperinus*

and *Tenebrio molitor*). Reference insect samples (RI) for each species were also retrieved (Supplementary Table S1) from a certified pet shop (AGRIPETGARDEN S.r.l., Conselve, Italy).

Table 1. List of analysed insect-based products. For each sample, information found on the label about the category, the species of insect, and the plant ingredients are reported. F (flour); FP (Food Product).

Code	Category	Label Declared insect	Label Declared Ingredients
F_001	Flour	*Tenebrio molitor*	_
F_002	Flour	*Gryllodes sigillatus*	_
F_003	Flour	*Alphitobius diaperinus*	_
FP_004	Pasta	*Alphitobius diaperinus* (14%)	*Triticum durum, Ocimum basilicum* (1.5%); organic powdered egg whites.
FP_005	Pasta	*Alphitobius diaperinus* (14%)	*Triticum durum*; organic powdered egg whites.
FP_006	Pasta	*Tenebrio molitor* (10%)	*Oryza sativa* (43); *Cicer arietinum* (43%); organic powdered egg whites (4%).

FP_007	Cracker	*Acheta domesticus* (14%)	*Triticum aestivum*; *Sesamum indicum* (6%); *Olea europaea.*
FP_008	Cracker	*Tenebrio molitor* (10%)	*Triticum aestivum*; *Cocos nucifera; Avena sativa*; *Sesamum indicum* (12%); *Porphyra* sp. (1.2%).
FP_009	Protein bar	*Acheta domesticus* (5.2%)	*Phoenix dactylifera*; *Prunus dulcis*; *Musa* spp. (11%); *Theobroma cacao* (9%); *Vaccinium macrocarpon* (8%); *Anacardium occidentale*; *Cannabis sativa.*
FP_010	Protein bar	*Acheta domesticus*(5.5%)	*Phoenix dactylifera*; *Prunus dulcis*; *Prunus armeniaca* (22%); *Pisum sativum*; *Helianthus annuus*; *Lycium barbarum*(4.5%); *Salvia hispanica* (3.5%).
FP_011	Protein bar	*Acheta domesticus* (20%)	*Arachis hypogaea* (34%); *Cannabis sativa*; *Theobroma cacao*; *Agave* sp; *Beta vulgaris*; *Cinnamomum* sp. (1%).
FP_012	Protein bar	*Acheta domesticus* (10%)	*Ananas comosus* (30%); *Phoenix dactylifera*; *Anacardium occidentale*; *Cocos nucifera*; *Plantago* sp.; *Citrus limon.*
FP_013	Pet food	*Hermetia illucens* (25%)	*Ipomoea batatas*; *Saccharomyces cerevisiae*; *Lycium barbarum* (1%); *Rosmarinus officinalis*; plant base glycerin.

2.2. Mock mixtures

To test for the efficacy of DNA metabarcoding in characterizing the composition of the insect-based products at the qualitative and semi-quantitative levels, six mock mixtures were prepared (Supplementary Table S2). These were composed of insect flour (*T. molitor*) mixed with wheat (*Triticum aestivum*) flour (20 ppm, 200 ppm and 500 ppm of gluten) or soybean (*Glycine max*) flour (50 ppm, 200 ppm and 500 ppm of soybean proteins). Wheat and soybean flour concentrations were defined based on the alert threshold for allergens according to the EU regulation (No. 828/2014) and to Ballmer-Weber et al. (2017), respectively.

2.3. DNA extraction

For insect-based products and mock mixtures (see Table 1 and Supplementary Table S2), purified gDNA was obtained starting from 250 mg of samples by using the DNeasy PowerSoil Kit (QIAGEN, Hilden, Germany) following the manufacturer's instructions. All samples were prepared in three replicates. For reference insects (see Supplementary Table S1), purified gDNA was obtained starting from 25 mg of samples by using the DNeasy Blood & Tissue Kit (QIAGEN, Hilden, Germany) following the manufacturer's instructions. Purified DNA was checked for concentration and purity by using a Qubit 2.0 Fluorometer and Qubit dsDNA HS Assay Kit (Invitrogen, Carlsbad, California, United States).

2.4. Insect identification by DNA barcoding

The 658 bp mtDNA *COI* region was used to authenticate the animal component in the purchased insect-based products. This region was amplified and sequenced for all 13 samples and for the reference insect by using primer pairs described by Folmer, Black, Hoeh, Lutz, & Vrijenhoek (1994) (Supplementary Table S3) and the

protocol described in Bellati et al. (2011). The obtained sequences were submitted to the international GenBank through the EMBL platform (see Supplementary Table S1 for accession numbers). Each sequence was taxonomically assigned to the reference species (or to the declared one in the case of food items) by looking at the nearest matches with the BLAST algorithm using the following cut-off values/maximum identity >99% and query coverage of 100%.

2.5. Libraries preparation and sequencing

To characterize the plant composition of the investigated insect-based products and mock mixtures, the obtained gDNA extracts were sequenced at the DNA barcode ITS2 region (Chen et al., 2010). Amplicons were obtained using the same approach described by Biella et al. (in press) with Illumina adapter (Supplementary Table S3) using puReTaq Ready-To-Go PCR beads (GE Healthcare Life Sciences, Italy) following the manufacturer's instructions in a 25 µL reaction containing 1 µL 10 µM of each primer and up to 50 ng of gDNA. PCR cycles consisted of an initial denaturation step for 5 min at 94 °C, followed by 40 cycles of denaturation (30 s at 94 °C), annealing (30 s at 56 °C), and elongation (1 min at 72 °C), and, hence, a final elongation at 72 °C for 10 min. Amplicon DNA was checked for concentration by using a Qubit 2.0 Fluorometer and Qubit dsDNA HS Assay Kit (Invitrogen, Carlsbad, California, United States) (Supplementary Table S4) and amplicon length was measured by comparison against QX DNA Size Marker using the Qiaxcel Automatic electrophoresis system (QIAGEN, Hilden, Germany). Samples were sequenced by the Center for Translational Genomics and BioInformatics (Milan, Italy). The sequencing was carried out on the MiSeq sequencing platform (Illumina, San Diego, CA, USA) with a paired-end approach (2 x 300 bp).

2.6. Bioinformatics analysis

Raw Illumina reads were paired and merged using the PEAR algorithm (Zhang, Kobert, Flouri, & Stamatakis, 2013). Pre-processing was performed using VSEARCH 2.0 algorithm (Rognes, Flouri, Nichols, Quince, & Mahé, 2016): reads were filtered out if ambiguous bases were detected and lengths were outside the bounds of 100 bp; moreover, an expected error=1 was used as an indicator of read accuracy. Sequences were then dereplicated using --derep_fulllength. In order to decrease the false positive rate in the sequence population, a chimera detection analysis was performed on the obtained reference sequences. Since there is no reference database for ITS2 region for chimera detection, we used --uchime_denovo algorithm that carries out a de novo analysis without a reference. Plant features were obtained using --cluster_fast algorithm with a 100% sequence identity with at least a depth of 500x for each feature. A random sequence was chosen as the representative sequence of the cluster. Subsequently, DNA metabarcoding analysis was performed using the plugins of the QIIME2 suite (https://docs.qiime2.org/). The taxonomic assignment of the representative sequences was carried out using the classify-consensus-blast plugin implemented in QIIME2 (Camacho et al., 2009) against the local database, built with downloaded ITS2 sequences available in NCBI at 29th of January, 2019, adopting a percent identity > 0.99 and a query coverage > 0.90. To evaluate the occurrence of contaminants (*T. aestivum* and *G. max*) in insect flour (F_001) and their relative abundance, we generated a heat map representation of the significant discriminatory features (plant species) obtained with the bioinformatic pipeline. Sample and feature axes were also organized using a clustering approach. The heat map was generated with the *feature-table* QIIME2 plugin (McDonald et al., 2012). To evaluate the sensitivity of the approach in detecting species based on feature depth, we performed a qualitative analysis considering the results

of the taxonomy assignment described previously, assuming a depth of 500x, 100x, and 25x for each OTU, respectively.

Python script (v.3; Pandas and NumPy libraries) was used to calculate sequence abundance weighted OTU and taxa overlap respectively (Wen et al., 2017) among the technical replicates. To evaluate significative differences among samples belonging to mock mixtures, a PERMANOVA test (permutation-based ANOVA, PerMANOVA) with 999 permutation-based Bray Curtis distance metrics (Faith, Minchin, & Belbin, 1987) was performed using the *diversity* QIIME2 plugin, considering both OTUs and taxa composition. PerMANOVA Pairwise contrast was performed through the beta-group-significance command of *diversity* plugin (Anderson, 2001).

2.7. ELISA assays

The three flour samples (F001, F002, and F003) and the mock samples prepared with *T. aestivum* (MT_20 - MT_500) were also analysed by RIDASCREEN® Gliadin kit (R-Biopharm AG, Darmstadt, Germany, prod. no. R 7001), a sandwich enzyme-linked immunosorbent assay (ELISA) kit for gluten detection. The assay is based on the monoclonal antibody R5 (Méndez, Vela, Immer, & Janssen, 2005; Valdés, García, Llorente, & Méndez, 2003), which is specific for gliadin-fractions from wheat. The detection limit for gluten is 3 ppm (mg/kg). The manufacturer's instructions were followed.

3. Results

3.1. DNA barcoding authentication of insect ingredients

Good DNA yield (20-40 ng/µl) was obtained from all the replicates of the 13 collected samples and from the reference insect samples as well (see Supplementary

Table S4). The mt *COI* DNA barcoding sequencing results indicate that all the tested insect-based products were correctly labelled concerning insect composition with the only exclusion of FP009 and FP010 (protein bars). In both cases, the DNA barcoding analysis did not find occurrences of *Acheta domestica* as expected, but the sequences matched with the *COI* of the food parasite species *Ectomyelois ceratoniae* (Insecta: Pyralidae).

3.2. DNA metabarcoding characterization of plant composition

The High-Throughput Sequencing step produced a total of 8,142,444 raw reads, with an average of 126,299 reads (SD = 108,789; range 334 - 482,500) per sample. After the merging, quality filtering and dereplication steps, we retrieved a total of 868,414 reads, with an average of 13,677 reads (SD = 11,453; range 5 - 45,066) per sample. Details on the average and standard deviation statistics about raw and filtered reads obtained for each sample, considering replicates, are provided in Supplementary Table S4. After chimera detection and 100% cluster identity, with a depth of at least 500x for each feature, we obtained 120 OTUs (Operational Taxonomic Units). Negative controls for library sequencing were not included in the analysis since the very low amount of DNA copies. OTUs and taxa diversity were analysed separately for technical replicates at each sample. Both OTUs and taxa overlap (calculated with the weight of reads per OTUs and taxa) maintained a mean of 90% for all insect-products (with a standard deviation of 0.26).

The obtained taxonomic assignment and the distribution of the assigned taxa among the sample data

are depicted in Fig. 1. Overall, 120 OTUs were assigned to at least 26 plant species. Most of the assigned OTUs reached the species taxonomic level, however, in some

genera, such as *Triticum* and *Brassica*, the low interspecific variability did not allow the species to be identified.

The three flour samples were largely different. Specifically, F_001 is rich in *Daucus carota* reads (93.10%) followed by *Brassica sp*. (1.80%) and *Glycine max* (0.99%), F_002 contains mainly *Cicer arietinum* (43%), *Triticum sp.* (35%), and *Brassica sp.* (4%) reads. Finally, the F_003 flour shows many reads of *Glycine max* (33%), *Triticum sp.* (30%), and *Cicer arietinum* (11%). Moreover, all the flour samples showed a variable relative read abundance of *Cannabis sativa* (range 2.18-11%) and *Linum* sp. (range 0.76%-6%).

The assigned plant taxa in the 13 samples were grouped in Expected species (E), that include taxa listed on the product label, Rearing Substrate (RS) which includes the putative plant used both as feed and litter for insect farming, and Not Expected (NE) encompassing all the remaining species. Table 2 indicates the distribution of the assigned species among the above-mentioned categories.

Fig. 1. Relative abundance of the plant taxa recovered in the 13 insect-based products through ITS2 metabarcoding sequencing.

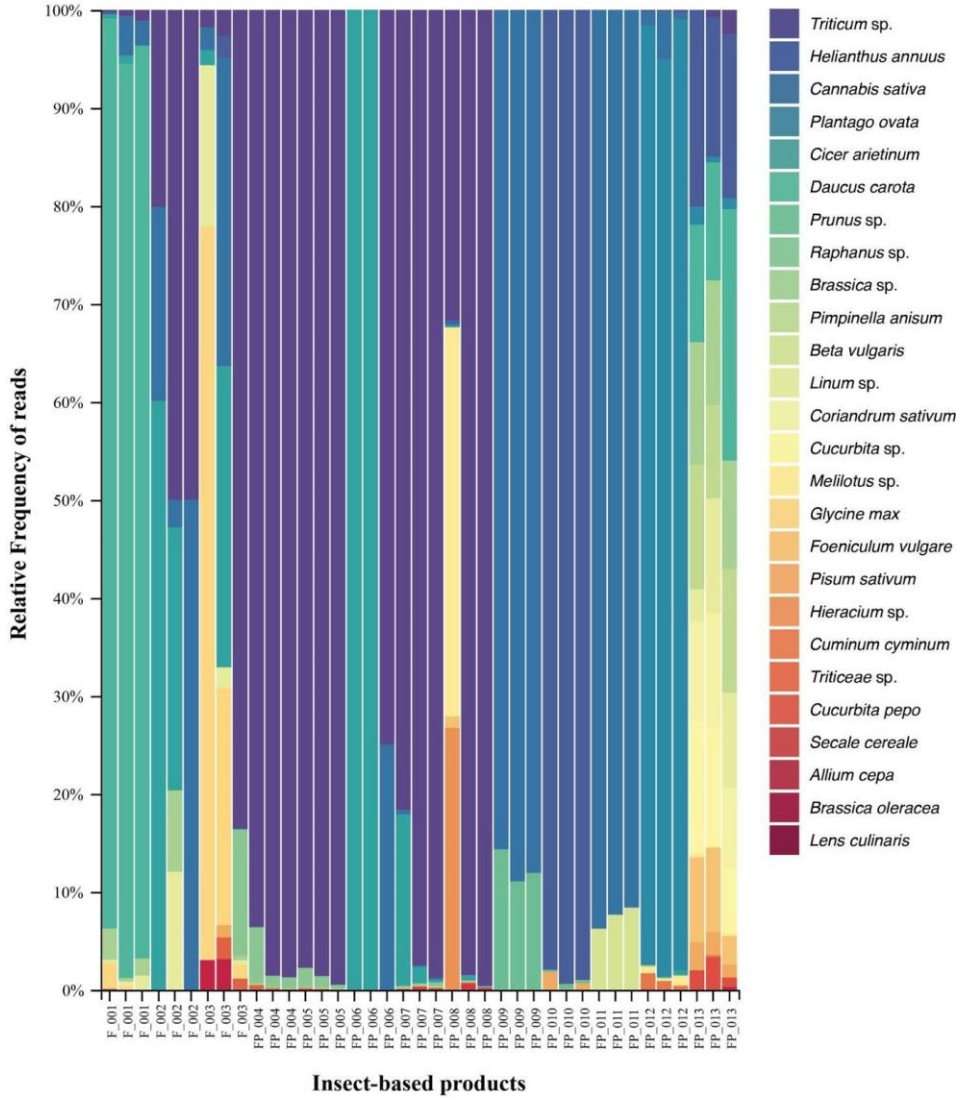


Table 2: List of the detected ingredients based on DNA metabarcoding assignment. For each insect-based product, the Expected species (E), the taxa belonging to the Rearing Substrate (RS), and the Not Expected (NE) species are indicated. The percentage values refer to the relative abundance of HTS ITS2 reads for each recognized ingredient.

Code	Detected ingredients		
	Expected (E)	**Rearing substrate (RS)**	**Not expected (NE)**
F_001	*T. molitor*	*Daucus carota* (93.1%); *Cannabis sativa* (2.2%); *Brassica* (1.8%); *Glycine max* (1*%); Linum* sp. (0.8%); *Cicer arietinum* (0.4%); *Triticum* sp. (0.6%); *Helianthus annuus* (0.1%).	*Melilotus* sp. (0.01%).
F_002	*G. sigillatus*	*Cicer arietinum* (43%); *Triticum* sp. (35%); *Cannabis sativa* (11%); *Linum* sp. (6%); *Brassica* (4%).	_
F_003	*A. diaperinus*	*Glycine max* (33%); *Triticum* sp. (30%); *Cannabis sativa* (11%); *Cicer arietinum* (11%); *Linum* sp. (6%); *Helianthus annuus* (1%); *Pisum sativum* (1%).	*Raphanus* sp. (4%); *Allium cepa* (1%).

FP_004	*A. diaperinus;Triticum* sp.(97%).	_	*Raphanus* sp. (3%).
FP_005	*A. diaperinus; Triticum* sp. (99%).	_	*Raphanus* sp. (1%).
FP_006	*T. molitor; Cicer arietinum* (67%).	*Triticum* sp. (25%); *Cannabis sativa* (8%).	_
FP_007	*A. domesticus; Triticum* sp. (92.7%).	*Cicer arietinum* (6.5%); *Cannabis sativa* (0.2%); *Brassica* sp. (0.2%); *Helianthus annuus* (0.06%).	*Beta vulgaris* (0.01%); *Raphanus* sp. (0.2%); *Secale cereale* (0.2%).
FP_008	*T. molitor; Triticum* sp. (74%).	*Cicer arietinum* (1%).	*Melilotus* sp. (13%); *Hieracium* sp. (9%); *Foeniculum vulgare* (1%); *Secale cereale* (1%).
FP_009	*Cannabis sativa* (87%); *Prunus* (12%)	_	*Ectomyelois ceratoniae*
FP_010	*Helianthus annuus* (97%);	_	*Ectomyelois ceratoniae*

	Pisum sativum (2%); *Prunus* sp. (1%).		
FP_011	*A. domesticus; Cannabis sativa* (93%); *Beta vulgaris* (7%).	_	_
FP_012	*A. domesticus; Plantago ovata* (96%).	*Cannabis sativa* (3%).	*Melilotus* sp. (1%); *Cuminum cyminum* (1%).
FP_013	*H. illucens*	*Helianthus annuus* (17%); *Daucus carota* (17%); *Pimpinella anisum* (12%); *Brassica* sp. (12%); *Linum* sp. (8%); *Pisum sativum* (2%); *Triticum* sp. (2%).	*Cucurbita* (11%); *Coriandrum sativum* (10%); *Foeniculum vulgare* (7%); *Cucurbita pepo* (2%); *Plantago ovata* (1%).

Concerning the processed insect-based products (from FP_004 - FP_013), the plant composition was clearly different among the tested categories. In the case of pasta, the first two samples (FP_004 and FP_005) mainly consisted of *Triticum* sp. (97% and 99% of reads respectively) according to the label information (Table 1). Interestingly, reads of the NE *Raphanus sp*. (3% and 1% respectively) were found.

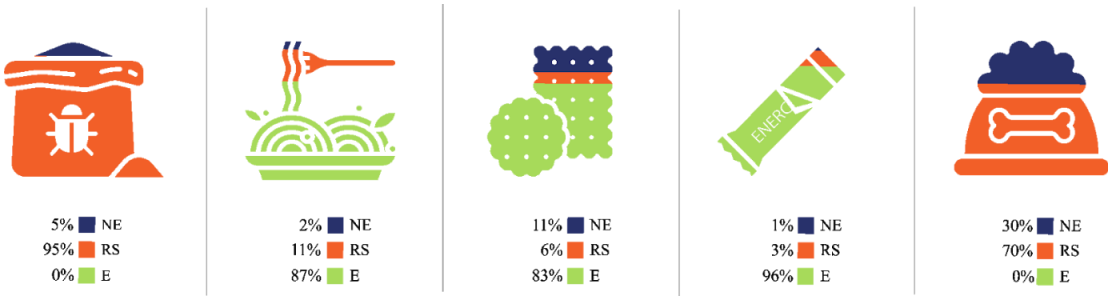
Regarding the F_005 pasta sample, two out of the three replicates were largely dominated by reads of the expected species *Cicer arietinum* (67%). Strangely, no OTUs belonging to *Oryza sativa* were found.

In the two cracker products, the reads of *Triticum* sp. were mainly detected (range 74 - 92.68%), followed by several NE species such as *Cicer arietinum* (range 1 - 6.5%), *Secale cereale* (range 0.19 - 1%), and other taxa occurring in traces (see Table 2).

In the four protein bars (belonging to two different companies), very few OTUs were obtained, and in most cases reads of E species were not found. Finally, the composition of pet food was very complex, and OTUs belonging to more than 12 species were detected (see Table 2).

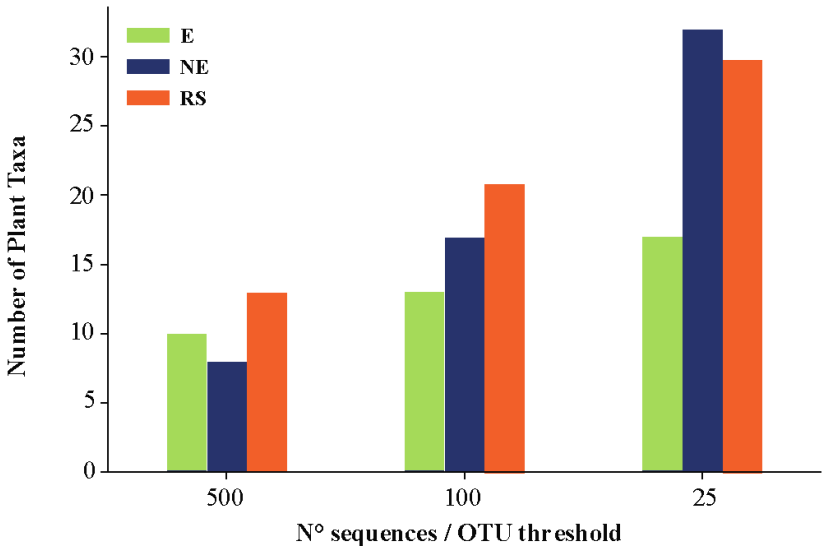
Overall, the plant taxa distribution among the five insect-based product categories is schematized in Fig. 2. The contribution of plant diet is appreciable in the insect flours which are the purest and least processed products. In Pasta, Crackers, and Protein bars the HTS analysis reveals the expected ingredient as the most abundant. Finally, in the pet food we detected the highest percentage of rearing substrate plant species.

Fig. 2. Comprehensive graphic summary of the detected plant taxa in the five insect-based product categories (from left: Flour, Pasta, Crackers, Protein Bars and Pet Food). E: expected species, RS: Rearing substrate species, NE: Not expected species.



To better characterize the composition of the processed products (from FP_004 to FP_013), we tested the identification performance reducing the reads filtering parameter from 500 to 100 and 25 reads per OTU. Fig. 3 shows the distribution of plant taxa among the three categories (i.e. E, RS, and NE) at different filtering thresholds. As expected, the whole number of assigned plant taxa increased with the decreasing threshold. Furthermore, the number of NE species increases dramatically compared to the increase of the Expected species.

Fig. 3. Numbers of plant taxa recovered in the 13 insect-based products and mocks (E, RS, and NE) through ITS2 metabarcoding sequencing using different thresholds of numbers of sequences per OTU (500, 100, and 25, respectively).

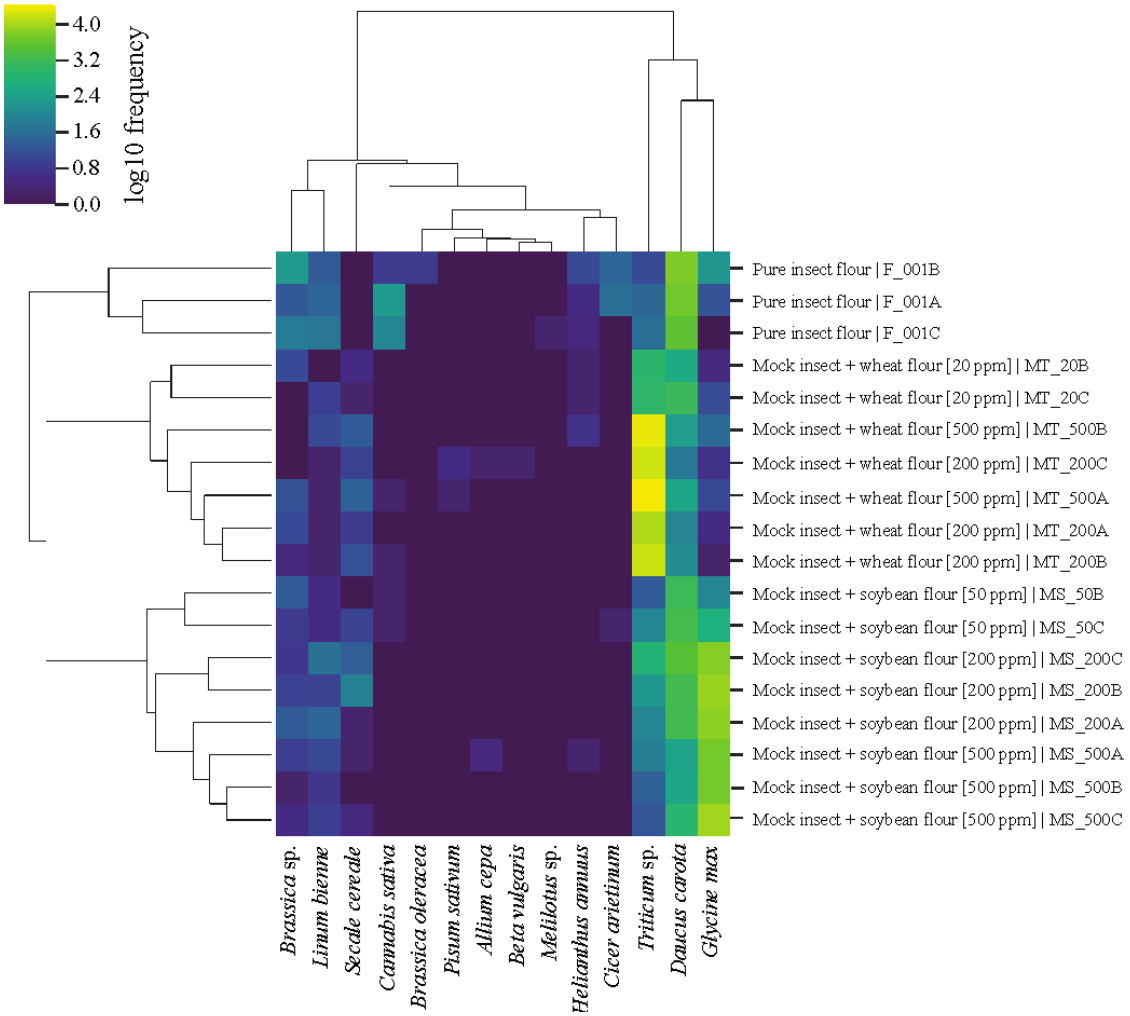


3.3. DNA metabarcoding characterization on flour mock mixtures

The results of the HTS DNA metabarcoding analysis performed on the six flour mock mixtures are shown in the heat map diagram of Fig. 4. Pure insect flours (F_001) cluster together and show the occurrence of several RS plants, especially *D. carota.* In these samples, the two contaminants *T. aestivum* and *G. max* are absent. Conversely, in both the types of flour mock mixture, the abundance of wheat and soybean reads increases along with the amount of the admixed contaminant flours. In the case of samples MT_200 and MT_500, the frequency of wheat reads reach the maximum level.

The PerMANOVA analysis shows that there is no statistical difference between samples belonging to the two categories of mock mixtures, both considering OTUs and taxa composition.

Fig. 4. Heat map diagram showing the abundance of plant taxa in the pure insect flours and mock mixtures. All the replicates, with the only exception of MT_20A and MS_50A, are shown. Color shading in the heat map indicates the abundance, expressed as log10 frequency, of each species in the sample.



3.4. ELISA assays

Flour samples (i.e., F_001, F_002 and F_003) and the mocks prepared with *T. aestivum* flour (i.e., MT_20, MT_200 and MT_500) were analysed for gluten detection. Sample F_002, the mock MT_200, and MT_500 were positive to gluten content. (Details are provided in Table 3).

Table 3. Results of ELISA analyses on insect flours and wheat mock mixtures.

Sample	ELISA results (mg/kg)	Gluten-free claim
F_001	<3	Yes
F_002	11	No
F_003	>20	Yes
MT_20	15	Yes
MT_200	>20	No
MT_500	>20	No

4. Discussion

The characterization of novel food composition is essential to ensure safety (Patel, Suleria, & Rauf, 2019) and our results suggest that the mtDNA *COI* region is useful for identifying the declared insect species in flours and in almost all the analysed insect-based products. However, this approach showed some limitations, since in two protein bars, the *COI* DNA barcoding analysis identified the parasite *Ectomyelois ceratoniae* instead of the declared *Acheta domestica.* The occurrence of *E.*

ceratoniae is not unexpected as its larvae typically parasitize raw food material, especially when almond occurs as an ingredient (Mortazavi, Samih, Ghajarieh, & Jafari, 2015). We cannot exclude that the industrial processing steps fragmented or degraded the *A. domestica* DNA in some way, making it not targetable by the used primer pair. The use of species-specific DNA probes (Tramuta et al., 2018) or another DNA metabarcoding approach targeting at the insect ingredients using short genetic regions (< 400 bp), would have detected the occurrence of both the moth and the insect species declared on product label (Frigerio et al., 2019).

Concerning the plant components, the ITS2 region was efficient in providing information on species composition including those taxa probably belonging to the insect rearing substrate. There are several open questions about the composition of the feeds used for insects farming and its influences on the quality and safety of the final products (Magara et al., 2019; Van Broekhoven, Oonincx, Van Huis, & Van Loon, 2015). The selection of suitable feeding substrates is very important to enhance the nutritional characteristics of the insects (Magara et al., 2019; Oonincx, Veenenbos, & van Loon, 2019; Oonincx, & de Boer, 2012; Van Broekhoven et al., 2015) and can also affect the total farming yield (Ganda, Zannou-Boukari, Kenis, Chrysostome, & Mensah, 2019). To date, the EU regulation (EC no 1069/2009) clarifies that the insect rearing substrate has to contain only products of non-animal origin. The circular economy strategies support the adoption of bio-waste and by-products of different agricultural and industrial origins claiming for potential benefits in terms of sustainability (EFSA, 2015). Therefore, the insect feeds are usually composed of vegetable ingredients derived from different agricultural supply chains which are likely impossible to be identified using morphological parameters. This was confirmed by our data which highlighted that the insect feed was almost completely characterized by horticultural plant sources, such as carrots, cabbages,

and chickpea (Magara et al., 2019). Therefore, the proposed molecular approach offers a universal diagnostic system to identify the composition of the rearing substrate and to verify compliance with the current and future regulations.

In agreement with Schlüter and co-workers (Schlüter et al., 2017), insects must be reared under a defined substrate to avoid contamination and possible food borne outbreak for the consumer (e.g., due to pathogenic microorganisms, toxins, and antinutrients). Our data suggest that the ITS2 barcode region was also able to identify the plant-based substrate used for insect rearing. For example, in the analysed flour samples, we detected DNA from hemp (*Cannabis sativa*) and linen *(Linum* sp.), which are commonly used as litter material (data confirmed by interviewed companies). We underline that *Cannabis sativa* is also used as an ingredient in some of the tested protein bars. Unfortunately, our method is not able to distinguish between the two sources.

In many processed products, we obtained most reads belonging to the expected highly abundant species, such as wheat and chickpeas, which likely hide the less represented OTUs. For example, the species that possibly constitute the rearing substrate are not very evident when compared to the analysed pure flours. Therefore, the main limitation of our analysis certainly resides in the sensitivity to detect the less abundant species due to primer bias. This issue has already been discussed by Bista and colleagues (Bista et al., 2018) and Krehenwinkel and colleagues (Krehenwinkel et al., 2019). Both studies agreed that a PCR-free whole genome sequencing could permit to avoid this effect. Furthermore, we demonstrated that different thresholds of OTUs size (in terms of number of reads) dramatically affect the final list of plant taxa recovered in the analysed samples. It is important to underline that in a context of food authentication and traceability, the adopted conservative criteria (i.e., n° sequences / OTU >500) is essential to preserve information on the

most representative species. Conversely, a deeper data exploration, using fewer conservative parameters (i.e., n° sequences / OTU < 500), retrieves information on trace species but increases the risk of including false positives. Probably, a multi-marker approach, coupled with a dedicated reference molecular database encompassing the expected plants, as well as the most common contaminants (or species prohibited by law) could improve the plant characterization of insect-based products and could be useful to exclude the false positives. Other authors (e.g., Zhang, Chain, Abbot, & Cristescu, 2018) demonstrated that the combined use of at least two barcode markers improves species detection. Another possible limitation of DNA metabarcoding resides in the completeness and reliability of the reference dataset that could lead to incorrect reads assignment (Murali, Bhargava & Wright, 2018). In our study, we chose the ITS2 barcode due to its higher capability of distinguishing congeneric plant species due to a higher mutation rate (Yu et al, 2017; Al-Juhani, 2019). Moreover, in recent years, the ITS2 database is growing exponentially and this improves the suitability of this locus to taxonomically assigning DNA metabarcoding data.

4.1. DNA metabarcoding to identify and quantify allergens

In our study, we tested the ability of DNA metabarcoding to find plant contaminants in edible insect flours. The obtained results suggested that we are able to identify DNA of *T. aestivum* or *G. max* in the tested mock mixtures, starting from 20 ppm and 50 ppm of allergenic proteins, respectively. According to European regulations (EU No. 828/2014), these concentrations are the maximum limits for commercial gluten and soybean-protein free products.

Therefore, the HTS DNA metabarcoding analysis detects low amounts of contaminant products and allergens, with a limit of detection even lower than the ELISA

analysis (i.e., 3 ppm). We underline that the three tested insect-based flours are declared as gluten-free products, but only F_001 and F_003 comply with the limit established by the European Commission.

Concerning the ability of DNA metabarcoding to quantify the 'putative plant contaminants', our study seems to indicate a weak relationship between the dry weight and the number of reads. However, there is a fervent debate about the effectiveness of providing quantitative inferences using HTS data. Some recent studies reported their findings in a quantitative manner where the relative read abundance is interpreted as the relative abundance of biomass (Lamb et al., 2019). Others use a frequency of occurrence approach, also referred to as weighted occurrence (Deagle et al., 2019), where the proportion of samples in which a given sequence was detected is used to infer a different sort of quantitative measure (De Barba et al., 2014).

As already noticed in our previous paper on processed food (Bruno et al., 2019), the amplicon DNA metabarcoding efficacy could be biased by the PCR amplification step using “universal” markers. The occurrence of bias during PCR amplification may cause the inaccurate estimation of quantities, and this was at least partially demonstrated for metazoans and plants (Balech et al., 2018; Thudi et al., 2012). This bias generates a variable number of template–primer mismatches across species, resulting in a final amplified DNA mixture that does not always reflect the original proportion of each species, limiting the quantitative potential of DNA metabarcoding (Bista et al., 2018; Piñol, Senar, & Symondson, 2019). Nevertheless, our analysis suggests that DNA metabarcoding has a relative quantitative ability, as already demonstrated by Lamb and colleagues (Lamb et al., 2019), and this methodology can be intended as an early warning method for allergen detection in food products.

4.2. DNA metabarcoding of insect-based novel food: an overview

DNA metabarcoding is currently used for food authentication (Prosser and Hebert, 2017; Utzeri et al, 2018; Galvin-King, Haughey, & Elliot, 2018). Moreover, Haynes and colleagues (Haynes, Jimenez, Pardo, & Helyar, 2019) recommend this approach to enhance the quality control along the food supply chain. In order to present an overview of DNA metabarcoding applied to insect food authentication and safety, we developed a SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis (Fig. 5). The strengths are related to the high efficiency of the technique. DNA metabarcoding can detect traces of ingredients due to its high sensitivity and allows to obtain simultaneously different information about food safety and quality. The main weaknesses are the higher cost compared to the current available analytical approaches like ELISA tests or target PCR assays. However, we cannot exclude that in the very next future, the panel of targets required to assess authenticity and safety of insect-based products will be so wide to make the HTS DNA metabarcoding approach much more convenient than the use of multiple single target tests. DNA tests indeed are related to DNA quality and quantity, so highly processed insect products like protein bars can be challenging to analyse. Finally, the results depend on the database which may be incomplete with a subsequent incorrect assignment. The opportunities are related to the novelty of insect food products. It is possible to create new economic opportunities in the analytical field with the goal to guarantee a safe product, stimulating the insect-based novel food market. HTS techniques, due to the completeness of the results, could also quickly support the compliance to forthcoming regulations. Finally, the first threat is information on insect feed composition is not reported on the product label and this impedes to compare the detected composition with the declared one. Furthermore, there is currently no scientific reference to DNA metabarcoding applied to the quality and

safety control of the new insect-based foods. In addition to that, DNA analyses are currently not mandatory, so this can interfere with the spread of DNA metabarcoding as a routine analysis.

Fig. 5. Overview of the main strengths, weaknesses, opportunities and threats (SWOT analysis) related to the use of DNA metabarcoding as tool for insect-based novel food products.

STRENGTHS	OPPORTUNITIES
• Higher sensitivity compared to EU-required analyses (e.g., ELISA) • More accurate food quality and traceability assessments than conventional methods • Increased food safety (multitarget allergen detection)	• Quickly support the compliance of future regulations • Create new economic opportunities in the analytical field • Stimulate the insect base novel food market
• Depends on DNA quality and quantity • Depends on completeness and accuracy of references databases • More expensive compared to EU-required analyses (e.g., ELISA)	• DNA authentication control of food products is not mandatory (e.g., in Europe) • Insect feed compositions and farming details are not required on the label • Lack of DNA metabarcoding references applied to insect-based novel food
WEAKNESSES	**THREATS**

5. Conclusions

Novel foods demand is increasing, and their consumption is expected to grow in the next years. This condition encourages an increase in the number of ingredient species and enhances the risk of misidentification, contamination, and counterfeiting. This is well documented in the case of fish, where more new species are available on the market, and in many countries the DNA barcoding approach is considered an essential tool to avoid fish frauds (Fox, Mitchell, Dean, Elliot, & Campbell, 2018).

Considering that many food products contain a mixture of species, we strongly encourage the adoption of DNA metabarcoding to better elucidate not only the food composition but also to assess trace elements belonging to different steps of the food supply chains. Unfortunately, this approach does not accurately estimate the biomass of the ingredient taxa, and although in our case the HTS DNA metabarcoding approach highlighted the occurrence of allergenic species (even at limit concentration values), we cannot use this method as an alternative to the standard ELISA test. This is also because the presence of DNA of a species is not necessarily correlated with the occurrence of allergens. However, we proposed the DNA metabarcoding analysis as a preliminary screening, especially for novel foods, because this method offers the ability to identify, at the semiquantitative level, several potential allergenic plants with a single analysis. Therefore, DNA-based analysis can be used to select which ELISA tests, and in general which of the more reliable toxicological assays for the detection of plant contaminants to use. In this sense, the DNA metabarcoding approach offers an opportunity to enhance the food safety of novel food products, such as those based on insect ingredients.

Data Statement

The dataset generated for this study will be submitted to the EBI metagenomics portal (https://www.ebi.ac.uk/metagenomics/) upon paper acceptance. In any case, we will provide the dataset if required.

Appendix A. Supplementary data

APPENDIX A: Supplementary Material

Table S.1. List of insect samples used as reference for DNA barcoding authentication.

Specimen	Species	Accession number
RI_001	*Acheta domesticus*	LR585071
RI_002	*Alphitobius diaperinus*	LR585072
RI_003	*Gryllodes sigillatus*	LR585073
RI_004	*Hermetia illucens*	LR585074
RI_005	*Tenebrio molitor*	LR585075

Table S.2. List of mock mixtures composed of insect flour (*T. molitor*) admixed with wheat and soybean contaminant flours.

Code	Contaminant species	Contaminant concentration (mg/kg)
MT_20	*Triticum aestivum*	20 ppm
MT_200	*Triticum aestivum*	200 ppm
MT_500	*Triticum aestivum*	500 ppm
MS_50	*Glycine max*	50 ppm
MS_200	*Glycine max*	200 ppm
MS_500	*Glycine max*	500 ppm

Table S.3. List of primer pairs used for DNA barcoding and metabarcoding analyses.

Primer name	5'-3'	Barcode locus	Reference
LCO1490	GGTCAACAAATCATAAAGATATTGG	COI	Folmer et al., 1994
HC02198	TAAACTTCAGGGTGACCAAAAAATCA		
S2F	ATGCGATACTTGGTGTGAAT	ITS	Chen et al., 2010
S3R	GACGCTTCTCCAGACTACAAT		

Table S.4. List of analysed specimens with DNA concentration (ng/µl) evaluated by using a Qubit 2.0 Fluorometer and Qubit dsDNA HS Assay Kit and details about the average number of reads for each specimen, considering raw and filtered reads (after merging, quality filtering and dereplication steps) and including replicates.

Specimen	DNA concentration (ng/µl)	Average number of raw reads	Average number of filtered reads
F__001A	53	38,186	5,252

F__001B	51		
F__001C	37.9		
F__002A	34.2	6,944	1,340
F__002B	43.4		
F__002C	32		
F__003A	31	83,306	14,056
F__003B	30		
F__003C	60		
FP_004A	100	97,395	9,979
FP_004B	130		
FP_004C	350		

FP_005A	112	304,191	25,231
FP_005B	40		
FP_005C	200		
FP_006A	200	103,443	8,055
FP_006B	300		
FP_006C	276		
FP_007A	42.8	157,591	18,127
FP_007B	38.3		
FP_007C	46.3		
FP_008A	43.4	203,698	21,307
FP_008B	36.6		

FP_008C	27.8		
FP_009A	87	158,204	22,390
FP_009B	76		
FP_009C	81		
FP_010A	95	312,664	26,646
FP_010B	155		
FP_010C	94		
FP_011A	160	108,473	9,270
FP_011B	125		
FP_011C	231		
FP_012A	39	162,537	14,163

FP_012B	34		
FP_012C	34		
FP_013A	30	144,571	25,150
FP_013B	33.4		
FP_013C	33.3		
MS_50A	47.6	40,883	5,940
MS_50B	38		
MS_50C	28.3		
MS_200A	42.5	91,712	11,939
MS_200B	26.2		
MS_200C	59		

MS_500A	36.8	54,081	5,753
MS_500B	37.8		
MS_500C	38		
MT_20A	40.8	108,620	13,827
MT_20B	36		
MT_20C	34.2		
MT_200A	50	89,601	7,888
MT_200B	60		
MT_200C	58		
MT_500A	60	133,585	13,554
MT_500B	43.7		

MT_500C	60

Acknowledgements

The authors are indebted with Paola Re for support in preparing the graphical part of the manuscript and with F. Braciola and M. Barnaby for advice and support during manuscript preparation.

Funding: This work was supported by Regione Lombardia in the framework of the Program 'Accordi per la ricerca e l'innovazione' under the project 'Food Social Sensor Network Food NET', grant number: E47F17000020009. The funder had no role in conducting the research and/or during the preparation of the article.

Conflicts of Interest: FEM2-Ambiente s.r.l., provided support in the form of a salary for authors J.F. and V.M. The company only provided financial support in the form of research materials

References

Al-Juhani, W. S. (2019). Evaluation of the capacity of the DNA barcode ITS2 for identifying and discriminating dryland plants. *Genetic Molecular Research,* 18(1): GMR18133. https://doi.org/10.4238/gmr18133

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26(1), 32-46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x

Balech, B., Sandionigi, A., Manzari, C., Trucchi, E., Tullo, A., Licciulli, F., Grillo, G., Sbisà, E., De Felici, S., Saccone, C., D'Erchia, A.M., Cesaroni, D., Casiraghi, M., Vicario, S. (2018). Tackling critical parameters in metazoan meta-barcoding

experiments: a preliminary study based on coxI DNA barcode. *PeerJ*, 6, e4845. https://doi.org/10.7717/peerj.4845

Ballmer-Weber, B. K., Holzhauser, T., Scibilia, J., Mittag, D., Zisa, G., Ortolani, C., Oesterballe, M., Poulsen, L., Vieths, S., Bindslev-Jensen, C. (2007). Clinical characteristics of soybean allergy in Europe: a double-blind, placebo-controlled food challenge study. *The Journal of Allergy and Clinical Immunology*, 119(6), 1489–1496. https://doi.org/10.1016/j.jaci.2007.01.049

Bellati, A., Tiberti, R., Cocca, W., Galimberti, A., Casiraghi, M., Bogliani, G., & Galeotti, P. (2014). A dark shell hiding great variability: a molecular insight into the evolution and conservation of melanic Daphnia populations in the Alps. *Zoological Journal of the Linnean Society*, 171(4), 697-715. https://doi.org/10.1111/zoj.12151

Biella, P., Tommasi, N., Akter, A., Guzzetti, L., Klecka, J., Sandionigi, A., Labra, M., Galimberti, A. (in press). Does a strong reduction of colony workforce affect the foraging strategy of a social pollinator? *Scientific Reports*. https://doi.org/10.1101/622910

Bista, I., Carvalho, G. R., Tang, M., Walsh, K., Zhou, X., Hajibabaei, M., Shokralla, S., Seymour, M., Bradley, D., Liu, S., Christmas, M., Creer, S. (2018). Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples. *Molecular Ecology Resources*, 18(5), 1020–1034. https://doi.org/10.1111/1755-0998.12888

Bruno, A., Sandionigi, A., Agostinetto, G., Bernabovi, L., Frigerio, J., Casiraghi, M., & Labra, M. (2019). Food Tracking Perspective: DNA Metabarcoding to Identify Plant Composition in Complex and Processed Food Products. *Genes*, 10(3), 248. https://doi.org/10.3390/genes10030248

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10, 421. https://doi.org/10.1186/1471-2105-10-421

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., Zhu, Y., Ma, X., Gao, T., Pang, X., Luo, K., Li, Y., Li, X., Jia, X., Lin, Y., & Luo, K. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PloS one*, *5*(1), e8613. https://doi.org/10.1371/journal.pone.0008613

De Barba, M., Miquel, C., Boyer, F., Mercier, C., Rioux, D., Coissac, E., & Taberlet, P. (2014). DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. *Molecular Ecology Resources*, 14(2), 306–323. https://doi.org/10.1111/1755-0998.12188

Deagle, B. E., Thomas, A. C., McInnes, J. C., Clarket, L. J., Vesterinen, E. J., Clare, E. L., Kartzinel, T.R., Eveson, J. P. (2018). Counting with DNA in metabarcoding studies: how should we convert sequence reads to dietary data? *Molecular Ecology,* 28(2), 391-406. https://doi.org/ 10.1111/mec.14734.

Di Mattia, C., Battista, N., Sacchetti, G., & Serafini, M. (2019). Antioxidant Activities in vitro of Water and Liposoluble Extracts Obtained by Different Species of Edible Insects and Invertebrates. *Frontiers in Nutrition*, 6, 106. https://doi.org/10.3389/fnut.2019.00106

Dobermann, D., Swift, J. A., & Field, L. M. (2017). Opportunities and hurdles of edible insects for food and feed. *Nutrition Bulletin*, *42*(4), 293-308. https://doi.org/10.1111/nbu.12291

EFSA (2015). Risk profile related to production and consumption of insects as food and feed. (2015). *EFSA Journal*, *13*(10), 4257. https://doi.org/10.2903/j.efsa.2015.4257

Faith, D. P., Minchin, P. R., & Belbin, L. (1987). Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio*, 69(1-3), 57-68. https://doi.org/10.1007/978-94-009-4061-1_6

Fernandez-Cassi, X., Supeanu, A., Jansson, A., Boqvist, S., & Vagsholm, I. (2018). Novel foods: a risk profile for the house cricket (Acheta domesticus). *EFSA Journal*, *16*, e16082. https://doi.org/10.2903/j.efsa.2018.e16082

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology,* 3(5), 294–299. https://doi.org/10.1371/journal.pone.0013102

Fox, M., Mitchell, M., Dean, M., Elliott, C., & Campbell, K. (2018). The seafood supply chain from a fraudulent perspective. *Food Security*, 10(4), 939-963. https://doi.org/10.1007/s12571-018-0826-z

Frigerio, J., Gorini, T., Galimberti, A., Bruni, I., Tommasi, N., Mezzasalma, V., & Labra, M. (2019). DNA barcoding to trace Medicinal and Aromatic Plants from the field to the food supplement. *Journal of Applied Botany and Food Quality*, 92, 33-38. https://doi.org/ 10.5073/JABFQ.2019.092.005

Galvin-King, P., Haughey, S. A., & Elliott, C. T. (2018). Herb and spice fraud; the drivers, challenges and detection. *Food Control*. 88, 85-97. https://doi.org/10.1016/j.foodcont.2017.12.031

Ganda, H., Zannou-Boukari, E. T., Kenis, M., Chrysostome, C. A. A. M., & Mensah, G. A. (2019). Potentials of animal, crop and agri-food wastes for the production of fly larvae. *Journal of Insects as Food and Feed*, 5(2), 59-67. https://doi.org/10.3920/JIFF2017.0064

Haynes, E., Jimenez, E., Pardo, M. A., & Helyar, S. J. (2019). The future of NGS (Next Generation Sequencing) analysis in testing food authenticity. *Food Control*, 101, 134-143. https://doi.org/10.1016/j.foodcont.2019.02.010

Krehenwinkel, H., Wolf, M., Lim, J. Y., Rominger, A. J., Simison, W. B., & Gillespie, R. G. (2017). Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. *Scientific Reports*, 7(1):17668. https://doi.org/10.1038/s41598-017-17333-x

Lamb, P. D., Hunter, E., Pinnegar, J. K., Creer, S., Davies, R. G., & Taylor, M. I. (2019). How quantitative is metabarcoding: A meta-analytical approach. *Molecular Ecology*, 28(2), 420–430. https://doi.org/10.1111/mec

Magara, H. J., Tanga, C. M., Ayieko, M. A., Hugel, S., Mohamed, S. A., Khamis, F. M., Salifu, D., Niassy, S., Sevgan, S., Fiaboe, K.K.M., & Roos, N. (2019). Performance of Newly Described Native Edible Cricket Scapsipedus icipe (Orthoptera: Gryllidae) on Various Diets of Relevance for Farming. *Journal of economic entomology*, 112(2), 653-664. https://doi.org/10.1093/jee/toy397

Mazzucchelli, G., Holzhauser, T., Cirkovic Velickovic, T., Diaz-Perales, A., Molina, E., Roncada, P., Rodrigues, P., Verhoeckx, K., & Hoffmann-Sommergruber, K. (2018). Current (Food) allergenic risk assessment: Is it fit for novel foods? Status quo and identification of gaps. *Molecular nutrition & food research,* 62(1), 1700278. https://doi.org/10.1002/mnfr.201700278

McDonald, D., Clemente, J. C., Kuczynski, J., Rideout, J. R., Stombaugh, J., Wendel, D., Wilke, A., Huse, S., Hufnagle, J., Meyer, F., Knight, R., & Caporaso, J. G. (2012). The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. *GigaScience* 1:7. https://doi.org/10.1186/2047-217X-1-7

Méndez, E., Vela, C., Immer, U., & Janssen, F. W. (2005). Report of a collaborative trial to investigate the performance of the R5 enzyme linked immunoassay to determine gliadin in gluten-free food. *European Journal of Gastroenterology & Hepatology*, 17(10), 1053–1063. https://doi.org/10.1097/00042737-200510000-00008

Mortazavi, S., Samih, M. A., Ghajarieh, H., & Jafari, A. (2015). Effect of some diets on demographic parameters of Ectomyelois ceratoniae (Zeller) (Lepidoptera: Pyralidae) in vitro. *Journal of Plant Protection Research*, 55(2), 212–219. https://doi.org/10.1515/jppr-2015-0028

Murali, A., Bhargava, A., & Wright, E. S. (2018). IDTAXA: A novel approach for accurate taxonomic classification of microbiome sequences. *Microbiome*, 6(1):140. https://doi.org/10.1186/s40168-018-0521-5

Myers, G., & Pettigrew, S. (2018). A qualitative exploration of the factors underlying seniors' receptiveness to entomophagy. *Food Research International*, 103, 163-169. https://doi.org/10.1016/j.foodres.2017.10.032

Oonincx, D. G. A. B., & Boer, I. J. M. de. (2012). Environmental Impact of the Production of Mealworms as a Protein Source for Humans – A Life Cycle Assessment. *PLOS ONE*, *7*(12), e51145. https://doi.org/10.1371/journal.pone.0051145

Oonincx, D. G. A. B., Laurent, S., Veenenbos, M. E., & Loon, J. J. A. van. (2019). Dietary enrichment of edible insects with omega 3 fatty acids. *Insect Science*, 0(0), 1-10. https://doi.org/10.1111/1744-7917.12669

Pali-Schöll, I., Verhoeckx, K., Mafra, I., Bavaro, S. L., Mills, E. C., & Monaci, L. (2019). Allergenic and novel food proteins: state of the art and challenges in the allergenicity assessment. *Trends in Food Science & Technology*, *84*, 45-48. https://doi.org/10.1016/j.tifs.2018.03.007

Patel, S., Suleria, H. A. R., & Rauf, A. (2019). Edible insects as innovative foods: Nutritional and functional assessments. *Trends in Food Science & Technology*, 86, 352–359. https://doi.org/10.1016/j.tifs.2019.02.033

Payne, C. L. R., Scarborough, P., Rayner, M., & Nonaka, K. (2016). Are edible insects more or less «healthy» than commonly consumed meats? A comparison using two nutrient profiling models developed to combat over- and undernutrition. *European Journal of Clinical Nutrition*, *70*(3), 285–291. https://doi.org/10.1038/ejcn.2015

Petrakis, E. A., Cagliani, L. R., Polissiou, M. G., & Consonni, R. (2015). Evaluation of saffron (Crocus sativus L.) adulteration with plant adulterants by 1H NMR metabolite fingerprinting. *Food chemistry*, 173, 890-896. https://doi.org/10.1016/j.foodchem.2014

Piñol, J., Senar, M. A., & Symondson, W. O. C. (2019). The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. *Molecular Ecology*, 28(2), 407–419. https://doi.org/10.1111/mec.14776

Prosser, S. W. J., & Hebert, P. D. N. (2017). Rapid identification of the botanical and entomological sources of honey using DNA metabarcoding. *Food Chemistry*, 1;214:183-191. https://doi.org/10.1016/j.foodchem.2016

Regulation (Ec) No 1069/2009 of the European Parliament and of the council of 11 December 2015. Available online https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2015:327:FULL&from=EN. (accessed on 19 April 2019).

Regulation (EC) No 764/2008 of the European Parliament and of the Council of 21 October 2009. Available onlinehttps://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:300:0001:0033:EN:PDF. (accessed on 19 April 2019).

Regulation (EU) 2017/893 of the European Parliament and of the council of 24 May 2017. Available online https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0893&from=en. (accessed on 19 April 2019).

Regulation (EU) No 142/2011 of 25 February 2011. Available online https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:054:0001:0254:EN:PDF. (accessed on 19 April 2019).

Regulation (EU) No 828/2014 of 30 July 2014. Available online https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014R0828&from=IT (accessed on 19 April 2019).

Ribeiro, J. C., Cunha, L. M., Sousa-Pinto, B., & Fonseca, J. (2018). Allergic risks of consuming edible insects: A systematic review. *Molecular nutrition & food research,* 62(1), 1700030. https://doi.org/10.1002/mnfr.201700030

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584. https://doi.org/10.7717/peerj.2584

Rumpold, B. A., & Schlüter, O. K. (2013). Nutritional composition and safety aspects of edible insects. *Molecular Nutrition & Food Research*, *57*(5), 802–823. https://doi.org/ 10.1002/mnfr.201200735

Schlüter, O., Rumpold, B., Holzhauser, T., Roth, A., Vogel, R. F., Quasigroch, W., Vogel, S., Heinz, V., Jäger, H., Bandick, N., Kulling, S., Knorr, D., Steinberg, P., & Kulling, S. (2017). Safety aspects of the production of foods and food ingredients from insects. *Molecular nutrition & food research*, *61*(6), 1600520. https://doi.org/10.1002/mnfr.201600520

Sun-Waterhouse, D., Waterhouse, G. I. N., You, L., Zhang, J., Liu, Y., Ma, L., Gao, J., & Dong, Y. (2016). Transforming insect biomass into consumer wellness foods: A review. *Food Research International*, *89*, 129–151. https://doi.org/1016/j.foodres.2016.10.001

Tramuta, C., Gallina, S., Bellio, A., Bianchi, D.M., Chiesa, F., Rubiola, S., Romano, A., & Decastelli, L (2018). A Set of Multiplex Polymerase Chain Reactions for Genomic Detection of Nine Edible Insect Species in Foods. *Journal of insect science*, Sep 1;18(5). https://doi: 10.1093/jisesa/iey087.

Thudi, M., Li, Y., Jackson, S. A., May, G. D., & Varshney, R. K. (2012). Current state-of-art of sequencing technologies for plant genomics research. *Briefings in Functional Genomics*, *11*(1), 3–11. https://doi.org/10.1093/bfgp/elr045

Utzeri, V. J., Ribani, A., Schiavo, G., Bertolini, F., Bovo, S., & Fontanesi, L. (2018). Application of next generation semiconductor based sequencing to detect the botanical composition of monofloral, polyfloral and honeydew honey. *Food Control*, 86, 342-349. https://doi.org/10.1016/j.foodcont.2017.11.033

Valdés, I., García, E., Llorente, M., & Méndez, E. (2003). Innovative approach to low-level gluten determination in foods using a novel sandwich enzyme-linked immunosorbent assay protocol. *European Journal of Gastroenterology & Hepatology*, 15(5), 465–474. https://doi.org/10.1097/01.meg.0000059119.41030.df

Van Broekhoven, S., Oonincx, D. G., Van Huis, A., & Van Loon, J. J. (2015). Growth performance and feed conversion efficiency of three edible mealworm species (Coleoptera: Tenebrionidae) on diets composed of organic by-products. *Journal of insect physiology*, *73*, 1-10. https://doi.org/10.1016/j.jinsphys.2014.12.005

Van der Fels-Klerx, H. J. (Ine), Adamse, P., Punt, A., & Van Asselt, E. D. (2018). Data Analyses and Modelling for Risk Based Monitoring of Mycotoxins in Animal Feed. *Toxins*, 10(2), 54. https://doi.org/10.3390/toxins10020054

Walia, K., Kapoor, A., & Farber, J. M. (2018). Qualitative risk assessment of cricket powder to be used to treat undernutrition in infants and children in Cambodia. *Food Control*, 92, 169–182. https://doi.org/10.1016/j.foodcont.2018.04.047

Wen, C., Wu, L., Qin, Y., Van Nostrand, J. D., Ning, D., Sun, B., Xue, K., Liu, F., Deng, Y., Liang, Y., & Zhou, J. (2017). Evaluation of the reproducibility of amplicon sequencing with Illumina MiSeq platform. *PloS one*, *12*(4), e0176716. https://doi.org/10.1371/journal.pone.0176716

Yu, N., Wei, Y. L., Zhang, X., Zhu, N., Wang, Y. L., Zhu, Y., Zhang H.P., Li F., Yang L., Sun J., Sun, A. D. (2017). Barcode ITS2: A useful tool for identifying Trachelospermum jasminoides and a good monitor for medicine market. *Scientific Reports*, 7(1):5037. 10.1038/s41598-017-04674-w

Zhang, G. K., Chain, F. J. J., Abbott, C. Lp., & Cristescu, M. E. (2018). Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities. *Evolutionary Applications*, 11(10), 1901–1914. https://doi.org/10.1111/eva.12694

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2013). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*(5), 614-620. https://doi.org/10.1093/bioinformatics/btt593

Tasting the differences: microbiota analysis of different insect-based novel food

2020. Food Research International, 137: 109426

Running title: Microbiota signature in insect-based food

Author names and affiliations:

Jessica Frigerio 1¶, Giulia Agostinetto 2¶, Andrea Galimberti 2, Fabrizio De Mattia 1, Massimo Labra 2, Antonia Bruno 2*

1FEM2-Ambiente, Piazza della Scienza 2, I-20126 Milano, Italy;
2Zooplantlab, Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, I-20126 Milano, Italy;

* Corresponding author
email: antonia.bruno@unimib.it

¶These authors contributed equally to this work

Highlight

DNA barcoding identifies insect species in raw and processed insect-based novel food products

DNA metabarcoding identifies the microbiota of raw and processed insect-based novel food products

A “core microbiota” characterizes food products depending on the insect
DNA metabarcoding can be a useful tool for insect-based novel food quality assessment

Abstract

Traceability, quality and safety of edible insects are important both for the producers and the consumers. Today, alongside the burst of edible insects in western countries, we are facing a gap of knowledge of insect microbiota associated with the microbial ecosystems of insect-based products. In this context, High-Throughput DNA Sequencing (HTS) techniques can give insight into the carryover of insect microbiota into final food products. In this study, we investigated the microbiota composition of insect-based commercial food products, applying HTS techniques coupled with bioinformatic analysis. The work aimed to analyse the microbiota variability of different categories of some insect-based commercial food products made of *A. domesticus* (house cricket), *T. molitor* (mealworm beetle), and *A. diaperinus* (lesser mealworm or litter beetle), including commercial raw materials and processed food items, purchased via e-commerce from different companies. Our data revealed that samples cluster per insect species based on microbiota profile and preliminary results suggested that a small number of prevalent bacteria formed a “core microbiota” characterizing the products depending on the insect. This microbial signature can be recognized despite the different food processing levels, rearing conditions and selling companies. Furthermore, differences between raw and processed food made of the same insect or similar product produced by different companies was found.

These results support the application of HTS analysis for studying the composition of insect-based commercial food products in a wider perspective, for food traceability and food quality control.

Keywords: DNA metabarcoding; Microbiota; High-Throughput Sequencing; Insect; Processed food; DNA barcoding

Introduction

Entomophagy is an emerging and fashionable diet issue in western countries. Insects are an important source of energy for human diets, because of their richness in essential nutrients (Rumpold & Schlüter, 2015). They have a protein content average value ranging from 30% to 65% of the total dry matter, and they are also rich in micronutrients such as iron, zinc and calcium (Dobermann, Swift & Field, 2017). Insects like *Alphitobius diaperinus* and *Tenebrio molitor* L. can be also used as a source for the production of fortified foods (Roncolini et al., 2019; Roncolini et al., 2020) facing the problem of the food demand of the growing world population (Baiano, 2020). Moreover, preliminary studies of Oonincx & de Boer (Oonincx & de Boer, 2012) stated that, compared to other livestock animals, insect farming has a lower environmental footprint.

Safety, traceability and quality of edible insects are of great interest both for the producers and the consumers, heavily affecting the acceptance of edible insects in the human diet (House, 2016). New tools for safety controls on these food items could also benefit institutions like food agencies, customs and health departments in the evaluation of new product development based on processed insects. In the European Union, the regulation (Regulation EU 2015/2283) has classified edible

insects as novel foods, which follow specific rules and require specific authorizations before allowing them to be distributed (Klunder, Wolkers-Rooijackers, Korpela, & Nout, 2016; Van Huis, 2012; Schlüter et al., 2017). Besides, food safety authorities and the scientific community are discussing whether edible insects can be a reliable solution or a problem to the food security (Belluco et al., 2015; Di Mattia, Battista, Sacchetti, & Serafini, 2019).

The potential safety risks of edible insects are chemical hazards including pesticides, heavy metals, allergens, toxins (mycotoxin and bacterial toxins) (Garofalo et al, 2019). There is a risk that harmful insect microbes are transmitted through the consumption of insect products (van der Spiegel, Noordam, & van der Fels-Klerx, 2013). Most of the insect microbiota are associated with gut (e.g., the intrinsic insect symbionts in the intestinal tract and the proximity of other anatomical compartments) or related to extrinsic sources, such as environment and rearing conditions (substrates and feed), handling, processing and preservation (ANSES, 2014). Especially, as stressed recently by the European Food Safety Authority (EFSA, 2015), spore-forming bacteria in processed edible insects (including freeze-dried, boiled and dried varieties) can be considered a dangerous source of biological contamination as well.

Garofalo and colleagues (Garofalo et al., 2017) explored the microbiota of marketed processed edible insects using culture-based methods and pyrosequencing. They described, among others, the microbiota of whole dried small crickets (*Acheta domesticus*) and whole dried mealworm larvae (*Tenebrio molitor*), revealing a great bacterial diversity and variability among individual insect species: some of the identified microbes may act as opportunistic pathogens in humans, such as *Listeria* spp., *Staphylococcus* spp., *Clostridium* spp. and *Bacillus* spp., while others represent food spoilage bacteria, as well as *Spiroplasma* spp. in mealworm larvae. The insect diet

and social behaviour have a great impact on the composition of the gut microbial community (Tinker & Ottesen, 2013), therefore different insect farm conditions result in different microbiological ecosystems. Although some authors such as Stoops and co-workers (Stoops et al., 2017) suggested that the microbial taxonomic composition varies mainly with insect species, the additional factors such as the growing substrates or contact with soil may play an important role in the composition of the insect gut microbiota (Klunder, Wolkers-Rooijackers, Korpela, & Nout, 2016; EFSA, 2015; Li et al., 2016). Considering the insect production system, industrial practices, such as post-harvest starvation and rinsing, can affect the microbial quality of the final insect products too (Wynants et al., 2018). Since all food products, including those insect-based, undergo processing, the risk for human safety should be measured throughout the various stages, from raw materials (i.e. insect flour) to final food products (Osimani et al., 2018). High-Throughput DNA Sequencing (HTS) offers a standardized and sensitive method to evaluate the microbial community changes by analysing a wide range of food products (De Filippis, Parente & Ercolini, 2019). The search for a microbial signature represents an opportunity to verify both food safety and food traceability strategy, indeed the microbial variation gives insight about rearing and processing products. The microbial variability allows obtaining more information besides the identification of the insect species, like the hygienic and sanitary conditions concerning the rearing systems. Moreover, the insect microbiota can be used to identify the geographical origin of a food product and used as a tracing signature, as previously demonstrated by recent studies (Bokulich et al., 2016; Mezzasalma et al., 2017). The microbial signature can then eventually be applied to management and control systems (Galimberti et al., 2019).

In this study, we evaluated the microbiota composition of insect-based commercial food products, applying HTS with complementary bioinformatics analysis. The aim

of this preliminary study was to analyse the microbiota variability of different categories of insect-based products made of *A. domesticus* (house cricket), *T. molitor* (mealworm beetle), and *A. diaperinus* (lesser mealworm or litter beetle) (including commercial raw materials like flours and processed food items), purchased via e-commerce from different companies. We sought to define if HTS can be a useful tool for insect-based novel food quality assessment.

Materials and methods

2.1 Insect food products

A total of 12 commercial insect-based products were purchased via e-commerce from five different companies. Referring to the label information, these products contained only one insect species each: *Acheta domesticus* (Order: Orthoptera), *Alphitobius diaperinus* (Order: Coleoptera), and *Tenebrio molitor (Order: Coleoptera)* (S1 Table).

Four out of 12 samples were pure insect flours, belonging to the species *A. diaperinus* (n=1) and *T. molitor* (n=3), and they have been categorized as insect raw material (dried insect product without other ingredients). In the case of *T. molitor*, flour samples derived from three different batches of the same product. Eight out of 12 samples represented processed food products: pasta (n=3), crackers (n=2) and protein bars (n=3). A detailed description of the samples can be found in Table 1.

Table 1. List of analysed insect-based products. For each sample, the information found on the label about the category, the species of insects, the percentage of insects present in the food product, the other ingredients declared on the label and the company origin are reported. R (Raw food products); FP (Processed Food product). *Different batches of the same product of T. molitor flour.

Sample type	Code	Label declared insect	Label declared ingredients	Company origin	Company name
Flour	R_001	*T. molitor* (100%)*	--	Netherlands	Company 1
	R_002	*T. molitor* (100%)*	--	Netherlands	Company 1
	R_003	*T. molitor* (100%)*	--	Netherlands	Company 1
	R_004	*A. diaperinus* (100%)	--	Netherlands	Company 1
Pasta	FP_005	*A. diaperinus* (14%)	*Triticum durum*, *Ocimum basilicum* (1.5%); organic powdered egg whites	France	Company 2
	FP_006	*A. diaperinus* (14%)	*Triticum durum*; organic powdered egg whites	France	Company 2

	FP_007	*T.o molitor* (10%)	*Oryza sativa* (43%); *Cicer arietinum* (43%); organic powdered egg whites (4%)	France	Company 3
Cracker	FP_008	*A. domesticus* (14%)	*Triticum aestivum*; *Sesamum indicum* (6%); *Olea europaea*	Great Britain	Company 4
	FP_009	*T. molitor* (10%)	*Triticum aestivum*; *Cocos nucifera*; *Avena sativa*; *Sesamum indicum* (12%); *Porphyra* sp. (1.2%	France	Company 3
Protein bar	FP_010	*A. domesticus* (5.2%)	*Phoenix dactylifera*; *Prunus dulcis*; *Musa* spp. (11%); *Theobroma cacao* (9%); *Vaccinium macrocarpon* (8%); *Anacardium occidentale*; *Cannabis sativa*	France	Company 2
	FP_011	*A. domesticus*(5.5%)	*Phoenix dactylifera*; *Prunus dulcis*; *Prunus armeniaca* (22%); *Pisum sativum*;	France	Company 2

			Helianthus annuus; *Lycium barbarum* (4.5%); *Salvia hispanica* (3.5%)		
	FP_012	*A. domesticus* (20%)	*Arachis hypogaea* (34%); *Cannabis sativa*; *Theobroma cacao*; *Agave* sp; *Beta vulgaris*; *Cinnamomum* sp. (1%)	Great Britain	Company 5

2.2 DNA extraction

High-quality genomic DNA was obtained starting from 250 mg of each sample of Table 1 using DNeasy PowerSoil Kit (QIAGEN, Hilden, Germany), according to manufacturer's instructions. Three replicates of DNA extraction were generated for each sample plus a negative control. Purified DNA was checked for concentration and purity by using a Qubit 2.0 Fluorometer and Qubit dsDNA HS Assay Kit (Invitrogen, Carlsbad, California, United States).

2.3 DNA barcoding characterization of insect samples

The 658 bp mtDNA COI region was used to validate the animal species declared on the label in the sampled insect-based products. This region was amplified and sequenced for all 12 samples according to the primer pairs presented by Folmer and colleagues (Folmer et al., 1994) and the protocol described in Bellati et al. (Bellati et al., 2014). Each sequence was defined as the nearest match with the BLAST

algorithm using the following cut-off values: maximum identity >99% and query coverage of 100%.

2.4 HTS library preparation and sequencing

To characterize the bacterial composition of the investigated insect-based products, 16S rRNA genes (V3 and V4 hypervariable regions) of the obtained gDNA extracts were sequenced using a High-Throughput DNA Sequencing approach. Amplicons were generated following the protocol described by Caporaso et al. (Caporaso et al., 2012) with Illumina adapters (S2 Table), with minor modifications as described in Frigerio et al. (Frigerio et al., 2020): we used PuReTaq Ready-To-Go PCR beads (GE Healthcare Life Sciences, Italy) according to manufacturer's instructions in a 25 µL reaction, containing 1 µL 10 mM of each primer and up to 50 ng of gDNA. The amplification profile consisted of an initial denaturation step for 5 min at 95 °C, followed by 25 cycles of denaturation (30 s at 95 °C), annealing (30 s at 55 °C), and elongation (30 s at 72 °C), and finally elongation at 72 °C for 5 min. Amplicon DNA was checked for concentration by using a Qubit 2.0 Fluorometer and Qubit dsDNA HS Assay Kit (Invitrogen, Carlsbad, California, United States) and amplicon length was measured by comparison against QX DNA Size Marker using the Qiaxcel Automatic electrophoresis system (QIAGEN, Hilden, Germany). Samples were sequenced by the Center for Translational Genomics and BioInformatics (Milan, Italy). The sequencing was performed on the MiSeq sequencing platform (Illumina, San Diego, CA, USA) with a paired-end approach (MiSeq Reagent Kit v3, 2 x 300 bp).

2.5 Bioinformatics analysis

Illumina reads were analysed with QIIME2, Quantitative Insights Into Microbial Ecology 2 program (ver. 2019.4; https://qiime2.org/) (Boylen et al., 2018).

Sequences were demultiplexed with native plugin and DADA2 (Divisive Amplicon Denoising Algorithm 2) (Callahan et al., 2016) was applied to obtain ASVs sequences (or features) (Callahan et al., 2017), trimming primers and performing a quality filter with an expected error of 2.0. Chimeric sequences were removed using the consensus method. Features with at least 10 representatives associated and detected in at least two samples were kept. The taxonomic assignment of representative sequences was carried out using the *feature-classifier* (https://github.com/qiime2/q2-feature-classifier) plugin implemented in QIIME2, using *classify-consensus-vsearch* method against the SILVA SSU non-redundant database (132 release), adopting a consensus confidence threshold of 0.8. Taxa bar plots were generated with the QIIME2 dedicated plugin *taxa* (https://github.com/qiime2/q2-taxa). As ASVs assigned to Cyanobacteria phylum (Class: Chloroplast) were considered potential plant contaminants, they were removed from the downstream analysis. Reads of mitochondrial or eukaryotic origin were also excluded. Overlap among technical replicates was calculated considering taxa at the family level weighted for abundances (Wen et al., 2017). Alpha diversity was carried out considering the presence/absence of ASVs and Shannon index. Statistical differences among samples belonging to the same insect species were calculated using alpha-group-significance plugin by QIIME2, performing also a pairwise contrast (Kruskal and Wallis, 1952). Beta diversity, instead, was carried out considering qualitative (Jaccard and unweighted UniFrac) and quantitative (Bray-Curtis and weighted UniFrac) distance metrics (Lozupone et al., 2011), using QIIME2 *core-metrics* plugin (https://github.com/qiime2/q2-diversity). Statistical differences were calculated by permutation based ANOVA (PerMANOVA) functions of *beta-group-significance* plugin (Anderson, 2001), with 999 permutations, considering insect species and sample type categories. A PerMANOVA Pairwise contrast was

performed with *beta-group-significance* plugin. Principal coordinates plots (PCoA) method was used to explore the structure of microbial communities. The phylogenetic tree necessary to calculate UniFrac distances was built on the alignment of ASVs representative sequences using align-to-tree-mafft-fasttree method by *phylogeny* plugin (https://github.com/qiime2/q2-phylogeny). Heatmap visualization was used to explore the abundance of bacteria families among samples and was generated by QIIME2. Core microbiota among insect samples was calculated considering the ceiling of the mean of species frequencies among samples and keeping a core threshold of 0.7 (minimum fraction of samples that a species must be observed in), performed with *core-features* plugin (https://github.com/qiime2/q2-feature-table). A Venn diagram was created starting from core microbiota results setting the threshold = 1, by calculating the number of shared and unique taxa per insect collapsed at the genus level. ANCOM analysis (Analysis of composition of microbiomes; Mandal et al., 2015) was performed to test differential abundances among genera distribution in the dataset, comparing samples with different insect composition. To avoid false discovery rates, only features shared in at least 25% of samples were considered.

Results

3.1 Sequencing output

All the replicates of the 12 collected samples showed good DNA quality (i.e., A260/A230 and A260/A280 absorbance ratios within the range 1.6 - 2.2) and good yield (20-40 ng/µl). The DNA barcoding (mt COI) sequencing results indicated that all the tested samples were composed of insects. Moreover, the BLAST analysis against reference insect DNA barcoding sequences confirmed that all samples

corresponded with the declared insect species (i.e., maximum identity > 99% with the declared species).

HTS analysis produced about 8,571,836 raw reads from the analysed samples, with an average of 119,053.,28 reads per sample (DS = 62,045.83). After quality filtering, merging reads, chimaera and contaminants removal, we obtained a total of 590 ASVs (Amplicon Sequence Variants). Negative controls (deriving from DNA extraction and amplification step) for library sequencing were not included in the analysis since they encompassed a very low number of DNA reads.

3.2 Microbial diversity analysis

From overlap calculations for technical replicates, family overlap resulted in a mean of 96%, with a standard deviation of 0.06.

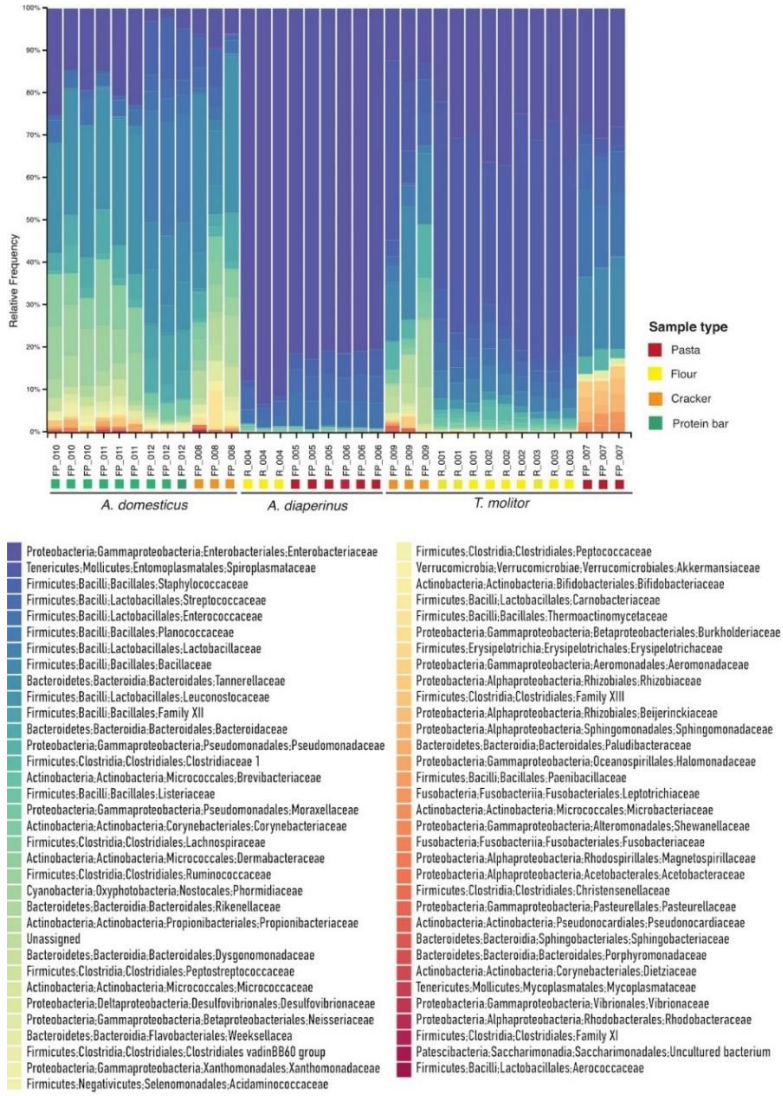
Both considering ASVs and Shannon metric, differences among samples derived from different insects were observed ($H=22.13$, p-value<0.01 and $H=29.93$, p-value<0.01, for ASVs and Shannon respectively; pairwise comparisons are visible in Table S3).

Samples belonging to raw material (flour) and food products (crackers, pasta and protein bars) showed a significant difference, considering both qualitative (Jaccard and Unweighted UniFrac) and quantitative metrics (Bray-Curtis and Weighted UniFrac) (p-value<0.01). Overall, we observed a significant difference among samples belonging to different insects (Jaccard metric: F-statistic=10.59, p-value=0.001; Unweighted UniFrac metric: F-statistic=10.57, p-value=0.001; Bray-Curtis metric: F-statistic=16.79, p=0.001; Weighted UniFrac metric: F-statistic=25.38; p-value=0.001). Results of pairwise comparisons are visible in Table S4.

3.3 Taxonomic composition analysis

A total of 9 bacterial phyla, 14 classes, 34 orders, and 66 families were identified (Fig 1, S5 Table).

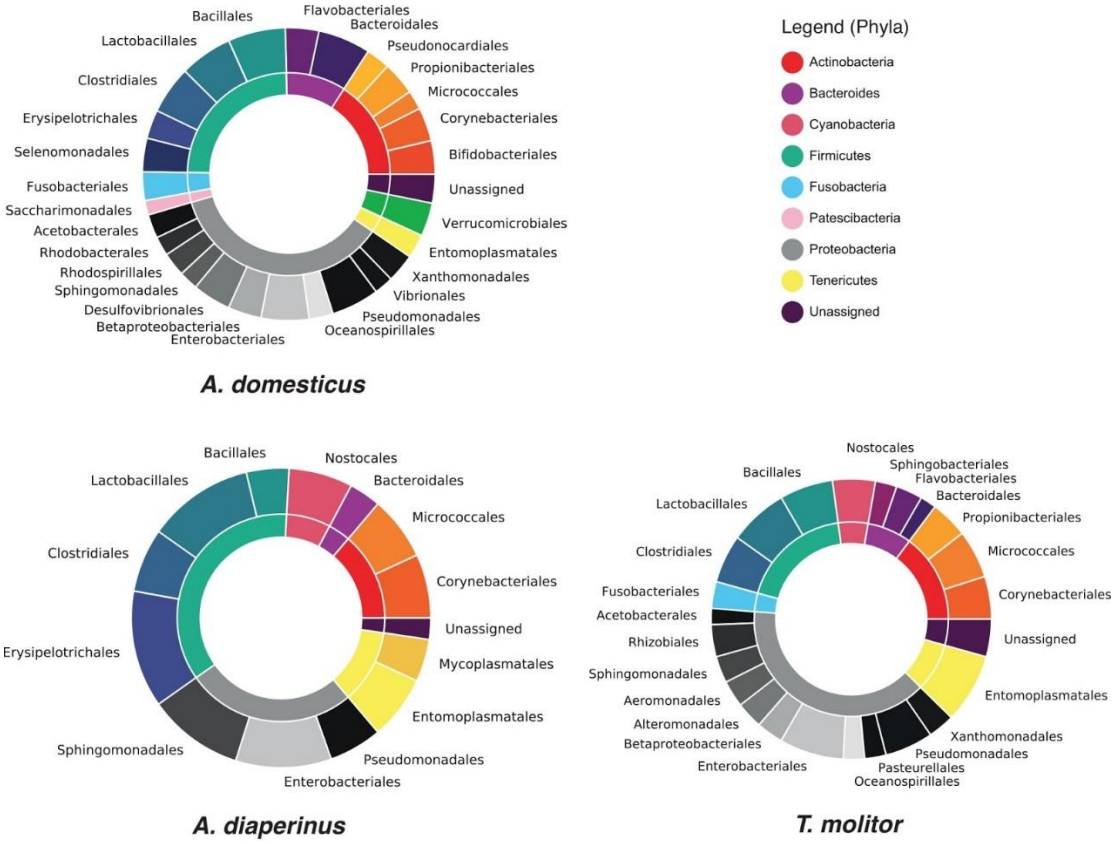
Fig 1. Relative abundance of bacteria families recovered in the insect-based products through 16S metabarcoding sequencing. Bacteria families are reported in gradient colors indicating relative abundances. For each sample, the sample type is reported (pasta: red square; flour: yellow square; cracker: orange square; protein bar: green square).



Taxonomic analysis revealed that most of the sequences in all the samples were associated with the phyla Proteobacteria (47%) and Tenericutes (26%), followed by Firmicutes (23%). 0.13% resulted in Unassigned taxa. Looking inside the taxonomic rank of class, the most abundant were Gammaproteobacteria, with 47% of sequences, followed by Mollicutes (26%) and Bacilli (22%). Enterobacteriales was the most abundant order, encompassing 45% of the sequences, distributed across all the samples, followed by Entomoplasmatales (26%), Lactobacillales (12%), Bacillales (10%), and Bacteroidales (2.6%). On the whole, the remaining 29 orders covered 4.4% of sequences. The Enterobacteriaceae family accounted for 45% of sequences, whereas Spiroplasmataceae represented 26% of sequences.

Considering taxa distribution per insect (Fig 2), we can notice differences in microbial composition, spanning from the phylum level to a deeper taxonomic resolution. Considering taxonomy per insect species, at the taxonomic level of order, we found that *A. domesticus*-based samples were dominated by Bacillales (54%), followed by Bacteroidales (21.2%), and Lactobacillales (8.9%), representing 84.1% of 28 orders. However, food products made with *A. diaperinus* had most of the sequences assigned to Enterobacteriales (89.6%), with the remaining 7% and 2.1% assigned to Lactobacillales and Bacillales, respectively, and 1.3% of sequences distributed in 11 orders. *T. molitor*-based food products showed 45.4% of sequences corresponding to Entomoplasmatales order, 29.5% to Enterobacteriales, 14.5% to Lactobacillales, and the remaining 10.6% to 21 different orders.

Fig 2. Donut charts of *A. domesticus*, *A. diaperinus*, and *T. molitor* microbial composition. Phyla in the inner circle and Orders in the outer circle are reported. Abundances are expressed as log frequency, in order to better show underrepresented taxa.



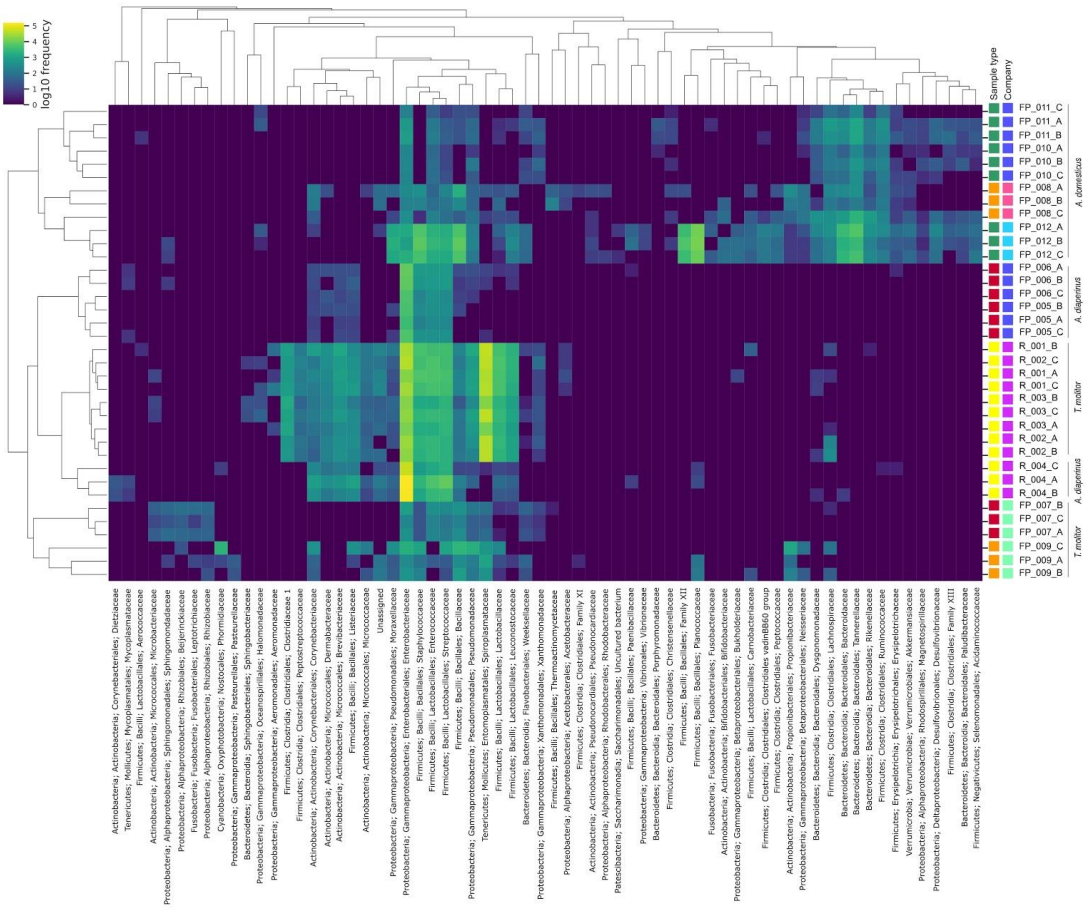
Focusing on specific features, we observed that the most abundant feature was assigned to an uncultured *Spiroplasma* (25%), reported exclusively in *T. molitor* samples. The sixth most abundant feature (3%), assigned to the genus *Kurthia* (Planococcaceae; Bacillales; Bacilli; Firmicutes) was detected only in *A. domesticus* protein bars produced by the British company 5, but not in samples belonging to the British company 4. Moreover, all and only the food products deriving from British company 5 are characterized by the presence of a specific feature assigned to

Exiguobacterium (Family XII; Bacillales; Bacilli; Firmicutes). Considering features shared between protein bars belonging to British company 5 and French company 2, some of the most abundants were assigned to Tannerellaceae (12,3%) (Bacteroidales; Bacteroidia; Bacteroidetes), followed by Bacteroidaceae (6%) (Enterobacteriales; Gammaproteobacteria; Proteobacteria), Enterobacteriaceae family (5%) (Enterobacteriales; Gammaproteobacteria; Proteobacteria) and Lachnospiraceae (2%) (Clostridiales; Clostridia; Firmicutes).

A feature assigned to an uncultured *Parabacteroides* (Tannerellaceae; Bacteroidales; Bacteroidia; Bacteroidetes) is unique for *A. domesticus* samples, whereas features assigned to *Enterobacter* (Enterobacteriaceae; Enterobacteriales; Gammaproteobacteria; Proteobacteria), a different microorganisms belonging to Enterobacteriaceae, and *Enterococcus* (Enterococcaceae; Lactobacillales; Bacilli; Firmicutes) were highly prevalent in *A. diaperinus* food products.

To better visualize the microbial variation among different food products, and which family mostly contribute distinguishing food products, a heatmap based on relative abundances was generated (Fig 3).

Fig 3. Heatmap diagram showing the relative abundance of families for each sample. Color shading in the heatmap indicates the abundance, expressed as log10 frequency, of each family in the sample. Sample type categories are flour (yellow), pasta (red), cracker (orange) and protein bar (green). Companies are represented in fuchsia (Company 1), blue (Company 2), aquamarine (Company 3), pink (Company 4) and light blue (Company 5).



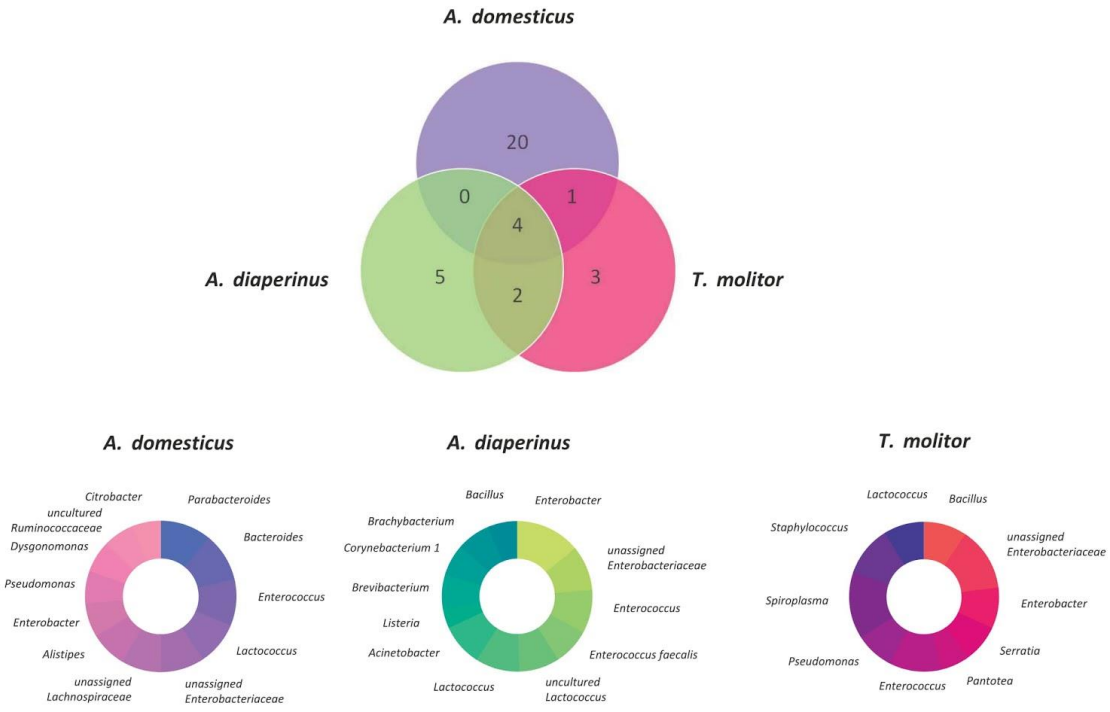
Analyzing the sample cluster dendrogram, two main clusters separate samples based on insect order, composed by *A. domesticus* (Orthoptera) food products and *T. molitor* plus *A. diaperinus* (both Coleoptera) food products. Subclusters

differentiated raw food products (flour) from processed food products (pasta, crackers and protein bars): flour made by the two insects of the Coleoptera order (i.e., *T. molitor* and *A. diaperinu*s) formed a distinct cluster that separated pasta and crackers samples based on the same insects. Moreover, the same food products constituted by different insects can be distinguished by family abundances in the heatmap: *A. diaperinus* pasta clustered separately from *T. molitor* pasta. Conversely, protein bars composed by the same insect (*A. domesticus*), but produced by different companies, are scattered in two different clusters, as also shown by microbial diversity analysis represented in the PCoA plot (Table S4).

3.4 Preliminary analysis on microbial signature

The preliminary analysis on *c*ore microbiota, defined as a group of shared microbial taxa or genes (Hamady & Knight, 2009; Turnbaugh et al., 2007), revealed the taxa shared by at least 70% and the 100% of samples of the category representing the insect used in the food products. Venn diagram, calculated from core microbiota results of the most conserved taxa (100% of samples per insect), highlighted the presence of unique and shared taxa considering insect species used in the food products analysed (Fig 4).

Fig 4. Venn diagram and donut charts of *A. domesticus*, *A. diaperinus*, and *T. molitor* core microbial composition. The Venn diagram in the upper part of the figure shows shared and unique taxa per insect. Taxa identified through core microbiota analysis are reported in the lower part of the figure. We considered the taxa found in 100% of the samples. In the case of *A. domesticus* and *A. diaperinus* the first twelve hits are reported, according to the frequency values listed in S4 Table.



In the case of *T. molitor*-based food products, we observed a core microbiota constituted by 21 taxa shared among > 70% of the samples and 10 taxa shared by all the samples. The 10 most conserved taxa (100% of samples) belonged to uncultured *Spiroplasma* sp., a taxon from Enterobacteriaceae family, *Enterococcus*,

Staphylococcus, *Enterobacter*, uncultured *Lactococcus*, *Pseudomonas*, *Bacillus*, *Serratia* and *Pantotea* (S6 Table).

On the other hand, *A. diaperinus*-based food products showed 14 shared taxa, both subsampling the 70% of samples or considering all the samples, indicating a highly conserved core microbiota. In contrast to *T. molitor*-based products, we reported the presence not only of *Enterococcus*, *Staphylococcus*, *Enterobacter*, *Lactococcus*, but also of *Enterococcus faecalis*, *Listeria*, *Brevibacterium*, *Corynebacterium*, *Brachybacterium*, *Acinetobacter*, and *Bacillus pumilus.* We reported as well the absence of *Spiroplasma*, *Pseudomonas*, *Serratia* and *Pantotea*.

Considering *A. domesticus*-based food products, all the samples shared 29 taxa, and 44 taxa are shared by 70% of samples. Among these, all the samples reported the presence of bacteria belonging to the family Lachnospiraceae and the genus *Parabateroides* (Family: Tannerellaceae).

Venn diagram analysis showed that, if four genera are shared among all the samples (a genus belonging to Enterobacteriaceae family, *Lactococcus*, *Enterobacter*, *Enterococcus*), 28 genera were unique considering the insect species. In particular, twenty genera were exclusively detected in all the samples of *A. domesticus*-based food products, and, among them, the three most abundant were *Parabacteroides*, *Bacteroides*, and a genus belonging to Lachnospiraceae family (see S6 Table for the complete list), thus confirming the explorative analyses described in the previous section. *Brevibacterium*, *Acinetobacter*, *Brachybacterium*, *Listeria*, and *Corynebacterium* were the genera unique for *A. diaperinus*-based food products, whereas *T. molitor*-based food products showed as unique genera *Spiroplasma*, *Pantoea*, and *Serratia*.

In order to explore variations in genera abundances among insect samples, ANCOM analysis was performed. Considering features shared in at least 25% of the dataset,

the analysis comprehended a total of 31 genus. ANCOM results showed 16 differential abundant genera among samples (S7 Table). In particular, ten were detected as insect-specific genera, according to core microbiome analysis. Further, for *T. molitor* samples, *Spiroplasma*, *Lactobacillus*, *Pediococcus* and a genus belonging to the *Clostridiaceae* family were identified, with a W-statistic of 30, 30, 24 and 23 respectively. Genera *Parabacteroides*, two uncultured bacteria belonging to the Ruminococcaceae family, *Bacteroides*, a genus belonging to Lachnospiraceae family and *Citrobacter* were peculiar of *A. domesticus* samples, with a W-statistic of 30, 30, 28, 29, 29, and 27, respectively. *Enterobacter* (W-statistic=30), *Corynebacterium* (W-statistic=25), and *Listeria* (W-statistic=25) were differentially distributed among the dataset, characterizing only *A. diaperinus* and *T. molitor* samples. Regarding genera that were shared among all insect species, *Lactococcus* (W-statistic=25), *Staphylococcus* (W-statistic=24), and a genus belonging to the Enterobacteriaceae family (W-statistic=23) were differentially distributed. Further, median abundances of Lactococcus were 35.5 in *A. domesticus* samples, 547.0 in *A. diaperinus* samples and 1,917.0 in *T. molitor* samples. *Staphylococcus*, instead, showed a median abundance of 501.0 in *A. diaperinus* samples and 1,475.0 in *T. molitor* samples, while Enterobacteriaceae medians were 144.0, 2,108 e 11,659, for *A. domesticus*, *A. diaperinus* and *T. molitor* samples, respectively (ANCOM results and distribution of genera among insect samples are visible in details in S7 Table).

Discussion

In this study, we characterized through the application of HTS techniques the microbial composition of insect-based food products made of *A. domesticus*, *T. molitor*, and *A. diaperinus*, purchased via e-commerce. We selected both raw and

processed food products, considering the availability on the market, from different selling companies.

Our preliminary data revealed that a small number of prevalent bacteria formed a "core microbiota" for each insect, which can potentially be used as biomarkers to identify insect ingredient origin in food products.

A recent study (Cambon et al., 2018) showed that a resident microbiota in *T. molitor* gut exists, thus supporting our hypothesis tested with core microbiota analysis. In particular, this study identified a resident *T. molitor* microbiota consisting of *Pseudomonas*, *Serratia* and genera belonging to the Enterobacteriaceae family. Noteworthy, this evidence is in accordance with the data we obtained in our study, as a further confirmation of our results.

If there was a significant insect component, the core microbiota would reflect the physiology of the organisms, the diet and rearing conditions. By contrast, if the level of food processing affected the microbiota, the organism could be difficult to identify searching for a microbial signature. Nevertheless, we identified shared features constituting the core microbiota of specific insects. In addition to that, despite the processing level, we found exclusive taxa in all the samples of specific insects. Noteworthy, our results showed that in *A. domesticus* processed food (i.e. protein bars and crackers) microbiota is composed by a robust core of microorganisms that is conserved and is similar in composition to what was reported in other studies on raw food (i.e. fresh crickets): Vandeweyer and colleagues (Vandeweyer, Crauwels, Lievens, & Van Campenhout, 2017) showed that *A. domesticus* is abundantly colonised by (Para)bacteroides species (Johnson, Moore, & Moore, 2009), confirming the first two hits we obtained through core microbiota analysis.

Interestingly, in this study *A. domesticus* core microbiota harbored bacteria belonging to the Lachnospiraceae family too. This evidence may prove beneficial when

edible insects will be introduced in the western diet and it is worth further studies: Lachnospiraceae are found, among others, in our digestive tract and are involved in fibre digestion. Menni and colleagues indeed discovered the association between Lachnospiraceae and lower long term weight (Menni et al., 2017). Furthermore, the exposure to antibiotics (such as β-lactam antibiotics and fluoroquinolones) eliminates Lachnospiraceae from gut microbiota. This lead to the gut becoming a prime target for opportunistic infections such as the one caused by *Clostridium difficile*, but restoring Lachnospiraceae into the intestines of infected patients has been shown to help cure *C. difficile* infections (Lagier et al., 2012; Segata et al., 2012; Song et al., 2013; Seekatz, 2018). It is conceivable that in the processed food we found only DNA and not viable cells and more investigations are needed, also focusing on prebiotic effects. In a recent study, the impact of an insect-based diet (cricket) on the human gut microbiota revealed increased levels of *Bifidobacterium animalis*. This could be due to cricket chitin which may function as a prebiotic (Stull et al., 2018). *T. molitor* flour in *in vitro* fecal models promoted the growth of Bacteroidaceae and Prevotellaceae, but not of *Clostridium histolyticum* group or Desulfovibrionales and Desulfuromonales (Carvallo et al., 2019).

On the other hand, exclusively all the samples based on *T. molitor* source are dominated by Spiroplasmataceae family (Phylum: Tenericutes; Class: Mollicutes), in particular bacteria belonging to *Spiroplasma* genus. *Spiroplasma* are found in the gut or hemolymph of insects where they can act as endosymbionts, impacting on host reproduction or host defence system. These findings are consistent with studies on fresh mealworm larvae (Vandeweyer, Crauwels, Lievens, & Van Campenhout, 2017) deriving from different companies.

A. diaperinus samples are dominated by *Enterobacter*, both flour and pasta, produced by different companies. These findings are in agreement with previous studies on fresh larvae (Wynants et al., 2017) and minced meat-like products (Stoops et al., 2016). *A. diaperinus*-based pasta clustered separately from flour samples made of the same insect, but in the same main cluster including food products belonging to Coleoptera. A similar behaviour can be seen in the case of *T. molitor* pasta and flour samples.

Concerning food safety, it is worth mentioning the presence, considering the 20 most abundant bacteria classified at the genus level, of sequences assigned to *Bacillus* in most of the samples (80%). The capacity to form endospore, resistant to heat and desiccation, deserve attention even if there is no confirmation of viability assay. There are currently no regulations for microbiological criteria of edible insects or their products in Europe, but a 5 Log10 (CFU/g) was defined as a safety threshold. Fasolato and colleagues found the presence of vital *Bacillus* in edible processed insects. Even if median values were lower than 4 Log10 CFU/g, some products showed higher lever (maximum 6.6 log10 CFU/g) (Fasolato et al., 2018).

Considering differential abundances of shared genera among samples, such as genera belonging to Enterobacteriaceae family or *Lactococcus*, differences may be caused by matrix peculiarities. In particular, food treatments as freezing or boiling processes can cause cell lysis and DNA degradation, thus affecting High-Throughput DNA Sequencing (HTS) output (De Filippis, Parente & Ercolini, 2019; Osimani et al, 2018). Further, high abundances of features were found in *T. molitor* and *A. diaperinus* samples, where we have a predominance of flour samples, a matrix obtained from only grinding treatment.

With the increasing availability of insect-based processed food products in the market, including a higher number of samples in the analyses will help in disentangling

the microbial dynamics behind food processing, and allowing the food products traceability at a finer scale.

Overall, our results showed that insect-based food products cluster based on their microbial signature. Even in the case of processed food in which there is more than one constituent (i.e., plant ingredients, see Table 1) that could interfere with its microbial contribution in the clustering process, we identified a shared pattern highlighted by core microbiota analysis and unique taxa that can be used as biomarkers. We also showed that differences exist in comparing raw vs processed food considering both qualitative and quantitative metrics. Recent studies (Bruno et al., 2019) reported the possibility to track the composition of plant processed food despite critical issues mostly deriving from the starting composition (i.e., variable complexity in taxa composition) of the sample itself and the different processing level (i.e., high or low DNA degradation). Other studies (Garofalo et al., 2017), investigating the microbial composition of commercial food products based on insects, never explored if any variability can be correlated with highly processed food such as pasta, crackers or protein bars. Our data clearly showed that processed food can be analysed searching for a microbial signature and that raw food products (i.e., flours) had a significant different microbiota compared to the processed ones (i.e., pasta, crackers and protein bars), even if maintaining unchanged a core of bacteria.

Highly processed food products represent one of the challenges of food traceability because of DNA degradation during food processing and, as a consequence, the limits in applying the common DNA barcoding techniques. Thus, DNA metabarcoding, based on HTS techniques combined with powerful tools for data analysis, can provide new perspectives for unveiling the composition of processed food, to

retrace food origin and food quality control (Bruno et al., 2019; Parente et al., 2019; De Filippis, Parente & Ercolini, 2019).

The identification of a microbial signature for traceability purposes was suggested also by forensic scientists as a natural consequence of the application of HTS technologies in a wider perspective (Bishop, 2019): with the globalisation of trade, food traceability is a hot topic and identifying a microbial signature in these products can provide a deeper insight into the "food ecosystem" (Galimberti et al., 2015; Bokulich, Lewis, Boundy-Mills, & Mills, 2016; Galimberti et al., 2019; Parente et al., 2019).

Conclusions and future perspectives

The application of high-throughput molecular techniques coupled with bioinformatic analyses allowed us to detect and identify the diversity of microbial communities in some raw and processed novel food products available on e-commerce. This study shows the value of the application of HTS analysis for unveiling the composition of microbiota in processed food containing insect ingredients. We were able to identify with our preliminary analysis a microbial signature, depending on the insect, suggesting that a resident microbiota is conserved despite the different food processing levels, rearing conditions, and selling companies. We are now facing a striking imbalance between available technologies and knowledge gaps on "food ecosystem": especially in the case of insect flour and insect-based products, as a future perspective we should consider the whole food production chain, taking into consideration that the microbial communities inhabiting surfaces, interacting with foods and being part of food themselves are influenced all along the supply chain, from rearing, in the case of insects, to the final processed product. HTS approach is a valuable tool to protect food safety as routine monitoring analysis, from

the identification of insect microbiota along the food production processing chain and characterization of the raw ingredients to the final processed food products. This tool can be applied to a wider range of food products to improve food source traceability too. Further studies are needed to improve our knowledge on the influence of rearing conditions and processing on the edible insect associated with the microbiota.

Acknowledgements

The authors are grateful to Valerio Mezzasalma for assistance during experimental procedures and manuscript preparation and to Filippo Bargero for English revision. Thanks also to Simone Bosaglia and FlatIcon community for graphic contribution.

Funding

This work was supported by Regione Lombardia in the framework of the Program 'Accordi per la ricerca e l'innovazione' under the project 'Food Social Sensor Network Food NET', grant number: E47F17000020009. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. FEM2-Ambiente s.r.l. provided support in the form of a salary for authors Fabrizio De Mattia and Jessica Frigerio, but did not play a role in study design, data collection and analysis, decision to publish, or preparation of the manuscript and only provided financial support in the form of research materials.

Competing interests

FEM2-Ambiente s.r.l., provided support in the form of a salary for authors J.F. and F.D.M. The company only provided financial support in the form of research materials.

Data availability statement

The dataset generated for this study was submitted to the EBI metagenomics portal (https://www.ebi.ac.uk/metagenomics/). BioSamples accessions are: SAMEA6262225-SAMEA626226 (Study ID: PRJEB35480).

Appendix A. Supplementary data

S1 Table. Insects used in the processed food analysed in this study.

Scientific name	*Acheta domesticus* (Linnaeus, 1758)	*Tenebrio molitor* (Linnaeus, 1758)	*Alphitobius diaperinus* (Panzer, 1797)
Taxonomy	Order: Orthoptera Family Gryllidae	Order: Coleoptera Family: Tenebrionidae.	Order: Coleoptera Family: Tenebrionidae
Description	House cricket, native to Southwest Asia, widespread in tropical and temperate zones. Species	Known as mealworm. It has a cosmopolitan distribution, being common	Known as lesser mealworm or litter beetle. It has a cosmopolitan distribution, being common in Europe, as a pest of grain storages and poultry farms.

	is native to most of the European countries.	in Europe, as a pest of grain storages.	
Growth	Adults grow up to 20-22 mm, both sexes are fully winged. Adult females are slightly bigger with prominent ovipositor protruding from the abdomen. Crickets are greyish yellow in color.	The adult beetles are up to 15-18 mm long. It is shiny black or brown with reddish brown elytra. The eggs are oval, whitish, about 1.5 mm long. The larvae resemble larvae of other mealworms, at the final stage measuring up to 25 mm in length.	The adult beetles are 6 mm long, oval. It is shiny black or brown with reddish brown elytra. Color is variable among individuals and subpopulations and changing with age. The antennae are paler at the tips and are covered in tiny, yellowish hairs. The elytra have shallow longitudinal grooves. The eggs are narrow, whitish, about 1.5 mm long. The larvae resemble larvae of other mealworms, at the final stage measuring up to 11 mm in length.
Incubation period (days from egg-laying to hatch)	11	10-12	10-12
Time to maturity (days from hatch	32-49	280-400	280-400

to max body weight)			
Resistance	Species is resistant to environmental conditions, and is very productive in mass culture, tolerating high population densities. The species is however very susceptible to the Cricket Paralysis Virus.	Species is resistant to environmental conditions, and is very productive in mass culture, tolerating high population densities.	Species is resistant to environmental conditions and is very productive in mass culture.
Protein and fat content	Protein content in larvae and imagines varies from 60 to 70% (d.m.), with a fat content of 20-25 % (d.m.)	Protein content in larvae varies from 50 to 65% (d.m.), with a fat content of 30-40 % (d.m.) highly depending on the feed and rearing conditions.	Protein content in larvae varies from 50 to 65% (d.m.), with fat content of 30-40 % (d.m.) highly depending on the feed and rearing conditions.

Adapted from http://ipiff.org/wp-content/uploads/2019/03/IPIFF_Guide_A4_2019-v5-separate.pdf

S2 Table. List of primer pairs used for DNA barcoding and metabarcoding analyses.

Primer name	5'-3'	Barcode locus	Reference
LCO1490	GGTCAACAAATCATAAAGATATTGG	COI	[17]
HC02198	TAAACTTCAGGGTGACCAAAAAATCA		
340F	CTACGGGNGGCWGCAG	16S	[19]
806R	GACTACHVGGGTATCTAATCC		

S3 Table. Results of alpha microbial diversity.

Pairwise comparison of ASVs counts between samples among to the same insect (*A. domesticus* n=12; *A. diaperinus* n=9; *T. molitor* n=15)

Group 1	Group 2	H	P-value	Q-value
A. domesticus	*A. diaperinus*	14.76	< 0.01	< 0.01
A. domesticus	*T. molitor*	5.60	0.018	0.018
A. diaperinus	*T. molitor*	14.83	< 0.01	< 0.01

Pairwise comparison of Shannon index between samples among to the same insect (*A. domesticus* n=12; *A. diaperinus* n=9; *T. molitor* n=15)

Group 1	Group 2	H	P-value	Q-value
A. domesticus	*A. diaperinus*	14.72	< 0.01	< 0.01
A. domesticus	*T. molitor*	18.02	< 0.01	< 0.01
A. diaperinus	*T. molitor*	16.20	< 0.01	< 0.01

S4 Table. Results of beta microbial diversity.

PCoA Emperor plots based on the Bray-Curtis metric. Food samples were compared based on insect order (red: Coleoptera; blue: Orthoptera) and insect species (sphere: *A. domesticus*; triangle: *T. molitor*; square: *A. diaperinus*).

Raw materials (flour, n=12) vs food products (crackers, pasta, protein bars; n=24)	**Pseudo-F**	**P-value**
Jaccard	5.84	0.001
Unweighted UniFrac	5.73	0.001

Bray-Curtis	6.31	0.001
Weighted UniFrac	8.00	0.002

Differences among samples belong to different insects (*A. domesticus* n=12; *A. diaperinus* n=9; *T. molitor* n=15)	**Pseudo-F**	**P-value**
Jaccard	10.39	0.001
Unweighted UniFrac	11.37	0.001
Bray-Curtis	16.87	0.001
Weighted UniFrac	25.63	0.001

Pairwise comparisons to test differences among samples belong to different insects (*A.domesticus* n=12; *A.diaperinus* n=9; *T.molitor* n=15).

Insects	**Group 1**	**Group 2**	**Pseudo-F**	**P-value**	**Q-value**

Jaccard	*A. domesti-cus*	*A. diaperi-nus*	11.99	0.001	0.001
	A. domesti-cus	*T. molitor*	8.47	0.001	0.001
	A. diaperinus	*T. molitor*	12.60	0.001	0.001
Unweighted UniFrac	*A. domesti-cus*	*A. diaperi-nus*	13.68	0.001	0.001
	A. domesti-cus	*T. molitor*	11.07	0.001	0.001
	A. diaperinus	*T. molitor*	7.31	0.001	0.001
Bray-Curtis	*A. domesti-cus*	*A. diaperi-nus*	16.36	0.001	0.001
	A. domesti-cus	*T. molitor*	12.36	0.001	0.001
	A. diaperinus	*T. molitor*	25.76	0.001	0.001

Weighted UniFrac	*A. domesti-cus*	*A. diaperi-nus*	29.01	0.001	0.001
	A. domesti-cus	*T. molitor*	24.83	0.001	0.001
	A. diaperinus	*T. molitor*	20.90	0.001	0.001

References

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. Austral Ecology, 26(1), 32–46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x

ANSES OPINION of the French Agency for Food, Environmental and Occupational Health & Safety on "the use of insects as food and feed and the review of scientific knowledge on the health risks related to the consumption of insects" | Anses - Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail. (n.d.). Retrieved April 22, 2020, from https://www.anses.fr/en/con-tent/opinion-french-agency-food-environmental-and-occu pational-health-safety-use-insects-food-and

Baiano, A. (2020). Edible insects: An overview on nutritional characteristics, safety, farming, production technologies, regulatory framework, and socio-economic and ethical implications. Trends in Food Science & Technology, 100, 35–50. https://doi.org/10.1016/j.tifs.2020.03.040

Bellati, A., Tiberti, R., Cocca, W., Galimberti, A., Casiraghi, M., Bogliani, G., & Galeotti, P. (2014). A dark shell hiding great variability: A molecular insight into the evolution and conservation of melanic Daphnia populations in the Alps. Zoological Journal of the Linnean Society, 171(4), 697–715. https://doi.org/10.1111/zoj.12151

Belluco, S., Losasso, C., Ricci, A., Maggioletti, M., Alonzi, C., & Paoletti, M. G. (2015). Edible insects: A food security solution or a food safety concern? Animal Frontiers, 5(2), 25–30. https://doi.org/10.2527/af.2015-0016

Bishop, A. H. (2019). The signatures of microorganisms and of human and environmental biomes can now be used to provide evidence in legal cases. FEMS Microbiology Letters, 366(3). https://doi.org/10.1093/femsle/fnz021

Bokulich, N. A., Lewis, Z. T., Boundy-Mills, K., & Mills, D. A. (2016). A new perspective on microbial landscapes within food production. Current Opinion in Biotechnology, Vol. 37, pp. 182–189. https://doi.org/10.1016/j.copbio.2015.12.008

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology, Vol. 37, pp. 852–857. https://doi.org/10.1038/s41587-019-0209-9

Bruno, A., Sandionigi, A., Agostinetto, G., Bernabovi, L., Frigerio, J., Casiraghi, M., & Labra, M. (2019). Food tracking perspective: Dna metabarcoding to identify plant composition in complex and processed food products. Genes, 10(3). https://doi.org/10.3390/genes10030248

Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME Journal, 11(12), 2639–2643. https://doi.org/10.1038/ismej.2017.119

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods, 13(7), 581–583. https://doi.org/10.1038/nmeth.3869

Cambon, M., Ogier, J.-C., Lanois, A., Ferdy, J.-B., & Gaudriault, S. (2018). Changes in rearing conditions rapidly modify gut microbiota structure in Tenebrio molitor larvae. BioRxiv, 423178. https://doi.org/10.1101/423178

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., … Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME Journal, 6(8), 1621–1624. https://doi.org/10.1038/ismej.2012.8

De Carvalho, N. M., Walton, G. E., Poveda, C. G., Silva, S. N., Amorim, M., Madureira, A. R., … Jauregi, P. (2019). Study of in vitro digestion of Tenebrio molitor flour for evaluation of its impact on the human gut microbiota. Journal of Functional Foods, 59, 101–109. https://doi.org/10.1016/j.jff.2019.05.024

De Filippis, F., Parente, E., & Ercolini, D. (2018). Recent Past, Present, and Future of the Food Microbiome. Annual Review of Food Science and Technology, 9(1), 589–608. https://doi.org/10.1146/annurev-food-030117-012312

Di Mattia, C., Battista, N., Sacchetti, G., & Serafini, M. (2019). Antioxidant Activities in vitro of Water and Liposoluble Extracts Obtained by Different Species of Edible Insects and Invertebrates. Frontiers in Nutrition, 6. https://doi.org/10.3389/fnut.2019.00106

Dobermann, D., Swift, J. A., & Field, L. M. (2017). Opportunities and hurdles of edible insects for food and feed. Nutrition Bulletin, 42(4), 293–308. https://doi.org/10.1111/nbu.12291

Fasolato, L., Cardazzo, B., Carraro, L., Fontana, F., Novelli, E., & Balzan, S. (2018). Edible processed insects from e-commerce: Food safety with a focus on the Bacillus cereus group. Food Microbiology, 76, 296-303. https://doi.org/10.1016/j.fm.2018.06.008

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology, 3(5), 294–299. https://doi.org/10.1071/ZO9660275

Frigerio, J., Agostinetto, G., Sandionigi, A., Mezzasalma, V., Berterame, N. M., Casiraghi, M., … Galimberti, A. (2020). The hidden 'plant side' of insect novel foods: A DNA-based assessment. Food Research International, 128. https://doi.org/10.1016/j.foodres.2019.108751

Galimberti, A., Bruno, A., Mezzasalma, V., De Mattia, F., Bruni, I., & Labra, M. (2015). Emerging DNA-based technologies to characterize food ecosystems. Food Research International, Vol. 69, pp. 424–433. https://doi.org/10.1016/j.foodres.2015.01.017

Galimberti, A., Casiraghi, M., Bruni, I., Guzzetti, L., Cortis, P., Berterame, N. M., & Labra, M. (2019). From DNA barcoding to personalized nutrition: the evolution of food traceability. Current Opinion in Food Science, Vol. 28, pp. 41–48. https://doi.org/10.1016/j.cofs.2019.07.008

Garofalo, C., Osimani, A., Milanović, V., Taccari, M., Cardinali, F., Aquilanti, L., … Clementi, F. (2017). The microbiota of marketed processed edible insects as revealed by high-throughput sequencing. Food Microbiology, 62, 15–22. https://doi.org/10.1016/j.fm.2016.09.012

Garofalo, C., Milanović, V., Cardinali, F., Aquilanti, L., Clementi, F., & Osimani, A. (2019). Current knowledge on the microbiota of edible insects intended for human consumption: A state-of-the-art review. Food Research International, 125, 108527. https://doi.org/10.1016/j.foodres.2019.108527.

Hamady, M., & Knight, R. (2009, July). Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. Genome Research, Vol. 19, pp. 1141–1152. https://doi.org/10.1101/gr.085464.108

House, J. (2016). Consumer acceptance of insect-based foods in the Netherlands: Academic and commercial implications. Appetite, 107, 47–58. https://doi.org/10.1016/j.appet.2016.07.023

Johnson, J. L., Moore, W. E. C., & Moore, L. V. H. (1986). Bacteroides caccae sp. nov., Bacteroides merdae sp. nov., and Bacteroides stercoris sp. nov. isolated from human feces. International Journal of Systematic Bacteriology, 36(4), 499–501. https://doi.org/10.1099/00207713-36-4-499

Klunder, H. C., Wolkers-Rooijackers, J., Korpela, J. M., & Nout, M. J. R. (2012). Microbiological aspects of processing and storage of edible insects. Food Control, 26(2), 628–631. https://doi.org/10.1016/j.foodcont.2012.02.013

Lagier, J. C., Million, M., Hugon, P., Armougom, F., & Raoult, D. (2012). Human gut microbiota: repertoire and variations. Frontiers in Cellular and Infection Microbiology, Vol. 2, p. 136. https://doi.org/10.3389/fcimb.2012.00136

Li, L., Xie, B., Dong, C., Wang, M., & Liu, H. (2016). Can closed artificial ecosystem have an impact on insect microbial community? A case study of yellow mealworm (Tenebrio molitor L.). Ecological Engineering, 86, 183–189. https://doi.org/10.1016/j.ecoleng.2015.09.015

Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J., & Knight, R. (2011, February). UniFrac: An effective distance metric for microbial community comparison. ISME Journal, Vol. 5, pp. 169–172. https://doi.org/10.1038/ismej.2010.133

Mandal, S., Van Treuren, W., White, R. A., Eggesbø, M., Knight, R., & Peddada, S. D. (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. Microbial Ecology in Health & Disease, 26(0). https://doi.org/10.3402/mehd.v26.27663

Menni, C., Jackson, M. A., Pallister, T., Steves, C. J., Spector, T. D., & Valdes, A. M. (2017). Gut microbiome diversity and high-fibre intake are related to lower long-term weight gain. International Journal of Obesity, 41(7), 1099-1105. https://doi.org/10.1016/j.fm.2018.06.008

Mezzasalma, V., Sandionigi, A., Bruni, I., Bruno, A., Lovicu, G., Casiraghi, M., & Labra, M. (2017). Grape microbiome as a reliable and persistent signature of field origin and environmental conditions in Cannonau wine production. PLoS ONE, 12(9). https://doi.org/10.1371/journal.pone.0184615

Oonincx, D. G. A. B., & de Boer, I. J. M. (2012). Environmental Impact of the Production of Mealworms as a Protein Source for Humans – A Life Cycle Assessment. PLoS ONE, 7(12), e51145. https://doi.org/10.1371/journal.pone.0051145

Osimani, A., et al (2018). Bread enriched with cricket powder (Acheta domesticus): A technological, microbiological and nutritional evaluation. Innovative Food Science and Emerging Technologies, 48, 150-163. https://doi.org/10.1016/j.ifset.2018.06.007.

Parente, E., De Filippis, F., Ercolini, D., Ricciardi, A., & Zotta, T. (2019). Advancing integration of data on food microbiome studies: FoodMicrobionet 3.1, a major

upgrade of the FoodMicrobionet database. International Journal of Food Microbiology, 305, 108249. https://doi.org/10.1016/j.ijfoodmicro.2019.108249

REGULATION (EU) 2015/ 2283 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL - of 25 November 2015 - on novel foods, amending Regulation (EU) No 1169/ 2011 of the European Parliament and of the Council and repealing Regulation (EC) No 258/ 97 of the European Parliament and of the Council and Commission Regulation (EC) No 1852/ 2001. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32015R2283&from=IT.

Risk profile related to production and consumption of insects as food and feed. (2015). EFSA Journal, 13(10), 4257. https://doi.org/10.2903/j.efsa.2015.4257

Roncolini, A., et al (2019). Protein fortification with mealworm (Tenebrio molitor L.) powder: Effect on textural, microbiological, nutritional and sensory features of bread. Plos One, 14(2), e0211747. https://doi.org/10.1371/journal.pone.0211747.

Roncolini, A, et al. (2020). Lesser mealworm (Alphitobius diaperinus) powder as a novel baking ingredient for manufacturing high-protein, mineral-dense snacks. Food Research International 131, 109031 https://doi.org/10.1016/j.foodres.2020.109031

Rumpold, B. A., & Schlüter, O. K. (2013). Nutritional composition and safety aspects of edible insects. Molecular Nutrition and Food Research, Vol. 57, pp. 802–823. https://doi.org/10.1002/mnfr.201200735

Schlüter, O., Rumpold, B., Holzhauser, T., Roth, A., Vogel, R. F., Quasigroch, W., … Engel, K. H. (2017). Safety aspects of the production of foods and food ingredients

from insects. Molecular Nutrition and Food Research, Vol. 61. https://doi.org/10.1002/mnfr.201600520

Seekatz, A. M., Theriot, C. M., Rao, K., Chang, Y. M., Freeman, A. E., Kao, J. Y., & Young, V. B. (2018). Restoration of short chain fatty acid and bile acid metabolism following fecal microbiota transplantation in patients with recurrent Clostridium difficile infection. Anaerobe, 53, 64–73. https://doi.org/10.1016/j.anaerobe.2018.04.001

Segata, N., Haake, S., Mannon, P., Lemon, K. P., Waldron, L., Gevers, D., … Izard, J. (2012). Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. Genome Biology, 13(6), R42. https://doi.org/10.1186/gb-2012-13-6-r42

Song, Y., Garg, S., Girotra, M., Maddox, C., Von Rosenvinge, E. C., Dutta, A., … Fricke, W. F. (2013). Microbiota dynamics in patients treated with fecal microbiota transplantation for recurrent Clostridium difficile infection. PLoS ONE, 8(11). https://doi.org/10.1371/journal.pone.0081330

Stoops, J., Crauwels, S., Waud, M., Claes, J., Lievens, B., & Van Campenhout, L. (2016). Microbial community assessment of mealworm larvae (Tenebrio molitor) and grasshoppers (Locusta migratoria migratorioides) sold for human consumption. Food Microbiology, 53, 122–127. https://doi.org/10.1016/j.fm.2015.09.010

Stoops, J., Vandeweyer, D., Crauwels, S., Verreth, C., Boeckx, H., Van Der Borght, M., … Van Campenhout, L. (2017). Minced meat-like products from mealworm larvae (Tenebrio molitor and Alphitobius diaperinus): microbial dynamics during production and storage. Innovative Food Science and Emerging Technologies, 41, 1–9. https://doi.org/10.1016/j.ifset.2017.02.001

Stull, V. J., Finer, E., Bergmans, R. S., Febvre, H. P., Longhurst, C., Manter, D. K., … Weir, T. L. (2018). Impact of Edible Cricket Consumption on Gut Microbiota in Healthy Adults, a Double-blind, Randomized Crossover Trial. Scientific Reports, 8(1). https://doi.org/10.1038/s41598-018-29032-2

Tinker, K. A., & Ottesen, E. A. (2018). The hindgut microbiota of praying mantids is highly variable and includes both prey-associated and host-specific microbes. PLoS ONE, 13(12). https://doi.org/10.1371/journal.pone.0208917

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-liggett, C., Knight, R., & Gordon, J. I. (2007). The human microbiome project: exploring the microbial part of ourselves in a changing world. Nature, 449(7164), 804–810. https://doi.org/10.1038/nature06244.The

Van der Spiegel, M., Noordam, M. Y., & van der Fels-Klerx, H. J. (2013). Safety of Novel Protein Sources (Insects, Microalgae, Seaweed, Duckweed, and Rapeseed) and Legislative Aspects for Their Application in Food and Feed Production. Comprehensive Reviews in Food Science and Food Safety, 12(6), 662–678. https://doi.org/10.1111/1541-4337.12032

Van Huis, A. (2013). Potential of Insects as Food and Feed in Assuring Food Security. Annual Review of Entomology, 58(1), 563–583. https://doi.org/10.1146/annurev-ento-120811-153704

Vandeweyer, D., Crauwels, S., Lievens, B., & Van Campenhout, L. (2017). Metagenetic analysis of the bacterial communities of edible insects from diverse production cycles at industrial rearing companies. International Journal of Food Microbiology, 261, 11–18. https://doi.org/10.1016/j.ijfoodmicro.2017.08.018

Wen, C., Wu, L., Qin, Y., Van Nostrand, J. D., Ning, D., Sun, B., … Zhou, J. (2017). Evaluation of the reproducibility of amplicon sequencing with Illumina MiSeq platform. PLOS ONE, 12(4), e0176716. https://doi.org/10.1371/journal.pone.0176716

Wynants, E., Crauwels, S., Lievens, B., Luca, S., Claes, J., Borremans, A., … Van Campenhout, L. (2017). Effect of post-harvest starvation and rinsing on the microbial numbers and the bacterial community composition of mealworm larvae (Tenebrio molitor). Innovative Food Science and Emerging Technologies, 42, 8–15. https://doi.org/10.1016/j.ifset.2017.06.004

Wynants, E., Crauwels, S., Verreth, C., Gianotten, N., Lievens, B., Claes, J., & Van Campenhout, L. (2018). Microbial dynamics during production of lesser mealworms (Alphitobius diaperinus) for human consumption at industrial scale. Food Microbiology, 70, 181–191. https://doi.org/10.1016/j.fm.2017.09.012

Paper to be submitted to Plants (MDPI)

DNA-based herbal teas authentication: a ITS2 and *psbA-trnH* multi-marker DNA metabarcoding approach

Author names and affiliations:

Jessica Frigerio 1, Valerio Mezzasalma 1, Giulia Agostinetto 2, Antonia Bruno 2, Lorenzo Guzzetti 2, Fabrizio De Mattia 1, Massimo Labra 2*

1 FEM2-Ambiente, Piazza della Scienza 2, I-20126 Milano, Italy;
2 Zooplantlab, Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, I-20126 Milano, Italy;

* Corresponding author
email: massimo.labra@unimib.it

Abstract

Medicinal plants have been widely used in traditional medicine due to their therapeutic properties. Although they are mostly used as herbal infusion and tincture, the employment as ingredients of herbal and food supplements is increasing. Despite this, fraud and adulteration are a widespread problem in the herbal and food industry. In our study, we aim at evaluating DNA metabarcoding as a extensive and sensitive tool to identify product composition, regardless of the number of taxa and their quantity. In order to accomplish this, we analysed fifteen commercial products with DNA metabarcoding, using two barcode regions: *psbA-trnH* and ITS2. Our

results show that on average, 70% of the declared ingredients have been identified. Concerning the use of two barcode markers, the ITS2 marker appears to identify more species (n=60) than *psbA-trnH* (n=35), with an ingredients identification rate of 49% for ITS2 versus 42% for *psbA-trnH*. However, the use of both markers has made it possible to almost double the capacity for identifying and assigning species, reaching for some samples to identify 100% of the declared species on the label. Some species (n=12) are identified only by one marker rather than the other. Additionally, in order to evaluate the quantitative ability of HTS to compare the plant component to the corresponding assigned reads, we created in laboratory six mock mixtures of plants both in biomass and gDNA. Our analysis suggests that DNA metabarcoding has a relative quantitative ability. These results support the application of HTS analysis for studying the composition of herbal teas from a wider perspective, for botanicals traceability and quality control.

Introduction

Medicinal plants have been used in traditional medicine for centuries due to their therapeutic properties. Although they are mostly consumed as herbal infusions and tinctures, employment as ingredients of herbal and food supplements is increasing worldwide (Lu et al., 2018). Consumers' awareness about healthy diets and their benefits is expanding the botanical and herbal supplements market (Lu et al., 2018). In the United States, the market of herbal supplements worth over US$ 7.4 billion per year (Smith et al., 2018) and the EU market accounts for a value of € 1.8 billion (Marieschi et al., 2009). Despite this, fraud and adulteration are a widespread issue in the herbal and food industry (Lupien, 2005). Medicinal plants are usually sold as herbal tea or as an ingredient, and fraud and adulteration are difficult to identify (Anthoons et al., 2021). The substitution of high-value plants with cheaper ones has

been widely reported in the literature (Nithaniyal et al., 2016; Mosa et al., 2018; Grazina et al., 2020). Moreover, adulteration with toxic plants has also been recorded and may lead to severe health risks (Cornara et al., 2017; Garzo et al., 2002) (eg. pyrrolizidine alkaloids (PA) are toxic for humans). These compounds are in high concentration in plants that can be accidentally in herbal teas (Wiedenfeld, 2010). Detecting adulteration and identifying the botanicals species present in herbal mixtures is fundamental to guarantee the consumers safety. Multiple methods, mainly based on morphological and chemical characterization, have been proposed in plant pharmacopoeia. However, these methods fail when the morphological features are lost or when the chemical profiles are shared among congenerics (Liu et al., 2018, Raclariu et al., 2017).

In the last decade, the use of molecular tools for the authentication of food products has drastically increased (Shaw et al., 2018). In this context, the biomolecular analysis of DNA barcoding has become more and more important over time (Galimberti et al., 2019). Although food authentication using DNA barcoding is well supported and validated when used to identify single species (De Mattia et al. 2011; Hellberg et al., 2017; Galimberti et al., 2014), the characterization of plant complex mixtures and/or processed products is still a challenge. DNA could undergo degradation processes due to industrial treatments. As a consequence, in several processed foods DNA could be highly degraded and fragmented (Bruno et al., 2019). To analyse these complex matrices, the DNA barcoding approach was combined with High-Throughput DNA Sequencing Technologies (HTS) which offers the opportunity to simultaneously sequencing multiple DNA amplicons (Haynes et al., 2019). Nevertheless, DNA metabarcoding still has some limitations such as amplification biases, accidental laboratory contamination when DNA is in low concentration and the difficulty of ingredients quantification (Shokralla et al., 2012; Elbrecht and

Leese, 2015; Bruno et al., 2018). In our study, we wanted to (i) evaluate if DNA metabarcoding can be a universal and sensitive tool to identify all the species in a product. In order to accomplish this, we analysed fifteen commercial products with DNA metabarcoding analysing two barcode regions, the nuclear ITS2 and for the very first time (in a DNA metabarcoding context), the plastidial intergenic spacer *psbA-trnH*. Additionally, in order to (ii) evaluate the quantitative ability of HTS, we compared the declared occurrence and abundance of plant components in a commercial herbal product to the corresponding assigned reads, creating in laboratory six mock mixtures of plants both starting from raw plants (biomass) and genomic DNA (gDNA).

Material and Methods

Sampling of herbal teas and assembling of mock mixtures

Fifteen samples of commercial herbal teas (Table 1) from five different companies were collected from supermarkets.

ID LAB	Company	Sample typology
HT_001	Company 1	Purifying Herbal Tea
HT_002	Company 1	Digestive Herbal Tea
HT_003	Company 1	Slimming Herbal Tea
HT_004	Company 2	Laxative Herbal Tea
HT_005	Company 2	Aromatic Herbal Tea

HT_006	Company 2	Purifying Herbal Tea
HT_007	Company 3	Aromatic Herbal Tea
HT_008	Company 3	Aromatic Herbal Tea
HT_009	Company 3	Depurative Herbal Tea
HT_010	Company 4	Relaxing Herbal Tea
HT_011	Company 4	Digestion Herbal Tea
HT_012	Company 4	Flat Stomach Herbal Tea
HT_013	Company 5	Laxative Herbal Tea
HT_014	Company 5	Sleep Herbal Tea
HT_015	Company 5	Draining Herbal Tea

Table 1. In the table are shown the ID specimen for all samples, the company of production and the sample typology.

In order to evaluate the ability of DNA metabarcoding to quantify the ingredients in herbal products, we set up an assay composed both by biomasses and genomic DNA of five plants in different proportions (Table 2). For these mock mixtures, common medicinal plant species were chosen: *Althaea officinalis* (roots), *Arnica montana* (flowers), *Ilex paraguariensis* (leaves), *Paullinia Capuana* (seeds) and *Solidago virgaurea* (aerial parts). As reported in several studies (Paranaiba et al., 2020; Lu et

al., 2020), the isolation of DNA can be challenging from some tissue such as wood and roots. For this reason, different parts of the plant were collected: roots, flowers, leaves, seeds and aerial parts. To test for the difference of quantification of plants in correlation to the features obtained before or after the extraction and to identify any bias, these mock mixtures were created with both plant sample quantities expressed as weight of dry material and with different concentrations of genomic DNA. In detail, for mock mixtures created by raw samples (QR_016, QR_017 and QR_018), plants were weighed and mixed in order to obtain the percentage expressed in Table 1. The weights and proportions chosen are consistent with the normal formulation of herbal teas, also in correlation with possible contaminants.
Concerning mocks mixture created by gDNA (QG_019, QG_020 and QG_021), DNA was individually extracted from each dry plant as indicated in the next paragraph and individually quantified using Qubit Fluorometer 4.0 (Thermofisher). Each plant species was identified by DNA barcoding analysis. Finally, we prepared different dilutions according to the percentage described in Table 2 of each DNA extract and we composed the artificial mixtures starting from the DNA extracts.

Species	Plant section	QR_016	QR_017	QR_018	QG_019	QG_020	QG_021
Althaea officinalis	Roots	20%	6%	35%	20%	6%	35%
Arnica montana	Flowers	20%	6%	6%	20%	6%	6%
Ilex paraguariensis	Leaves	20%	75%	20%	20%	75%	20%
Paullinia Capuana	Seeds	20%	6%	30%	20%	6%	30%
Solidago virgaurea	Aerial parts	20%	6%	9%	20%	6%	9%

Table 2. In the table are shown the percentage of the plants used for the mock mixture creation. Samples QR_016, QR_017 and QR_018 were created starting from the raw plants, samples QG_019, QG_020 and QG_021 were created starting from the genomic DNA.

DNA extraction and quantification

DNA extractions for herbal teas, positive control and mocks mixture were carried out using the commercial kit DNeasy PowerPlant (QIAGEN), following the manufacturer's instructions. We started from 50 mg of dry sample material that was homogenized via a mortar and liquid nitrogen; after lysis and wash steps, DNA was eluted in 50 uL of elution buffer and samples were stored at -20 °C. Three technical replicates of DNA extraction were created for each sample and negative controls of extraction were created. Genomic DNA concentration was evaluated by Qubit 4.0 Fluorometer (Thermofisher). DNA concentration (ng/µl) for all samples are indicated in Supplementary Table 1.

Libraries preparation and sequencing

In order to improve the ability to identify all the ingredients at the species level, two universal markers of DNA barcoding were analysed in this study: *psbA-trnH* and ITS2. ITS2 was selected as it was used in many DNA metabarcoding studies on plants and herbal teas (Biella et al., 2019; Richardson et al., 2015, Raclariu et al., 2017) while *psbA-trnH* was selected because it has a high intraspecific variability (Bolson et al., 2015). Amplicons were obtained using the same approach described by Bruno et al. (Bruno et al., 2019) with Illumina adapter (Supplementary Table S2) using pu-ReTaq Ready-To-Go PCR beads (GE Healthcare Life Sciences, Italy), following the manufacturer's instructions in a 25 µL reaction containing 1 µL 10 µM of each primer and up to 30 ng of gDNA. PCR cycles consisted of an initial denaturation step for 5 min at 94 °C, followed by 40 cycles of denaturation (30 s at 94 °C), annealing (30 s at 56 °C), and elongation (1 min at 72 °C), and, hence, a final elongation at 72 °C for 10 min. Amplicon DNA was checked for concentration by using a Qubit dsDNA HS Assay Kit (Invitrogen, Carlsbad, California, United States) (Supplementary Table S1) and amplicon length was measured by comparison against QX DNA Size Marker using the Qiaxcel Automatic electrophoresis system (QIAGEN, Hilden, Germany). Samples were sequenced by IGA Technology Services (Udine, Italy). The sequencing was carried out on the MiSeq sequencing platform (Illumina, San Diego, CA, USA) with a paired-end approach (2×300 bp).

Bioinformatic analysis

Illumina reads were analysed with QIIME2 (ver. 2020.8; https://qiime2.org/) (Boylen et al., 2018). After demultiplexing, primers were trimmed; ITS2 sequences were filtered with a minimum length of 100. Sequences were merged, dereplicated and chimaeras were removed via de-novo method (Rognes et al., 2016). A

clustering with a 1.0 of similarity threshold was applied. Sequences shared in at least two samples and with at least 250 representatives were kept. The taxonomic assignment of representative sequences was carried out via BLAST (Altschul et al., 1990), considering only reference sequences belonging to *Viridiplantae* rank (TaxID: 33090), adopting an identity of 97% and a coverage of 90%. Taxa bar plots were generated with the QIIME2 dedicated plugin *taxa* (https://github.com/qiime2/q2-taxa).

Statistical analysis

To identify the similarity existing between gDNA and plant mock samples, a PCoA analysis through the vegan package on R (version 3.3.3) was performed. A Bray-Curtis distance was applied to define the similarity matrix. Furthermore, a Linear Mixed Effect Model (LME) was performed on Bray-Curtis distances to compare the level of similarity to real samples posed by the different sample origins (gDNA or raw plant extracts). A random effect was posed on the replicates since each plant composition was tested three times and replicates were found to significantly affect the distance occurring with real samples. Packages exploited were nlme and lme4. Graphs were produced with ggplot2 package.

Results

DNA metabarcoding characterization of commercial herbal teas

HTS analysis produced about 12,358,533 raw pair reads from the analysed samples, with an average of 111,338.13 reads per sample (min=14,165; max=502,912). After quality filtering, merging reads, chimaera and clustering, we obtained a total of 508 ITS2 and 235 *psbA-trnH* features. Negative controls for library sequencing were not

included in the analysis since the very low amount of DNA reads. Overall, a total of 83 taxa were identified, of which 35 were found only by the *psbA-trnH* marker and 60 by the ITS2 marker. Considering the species declared on the label, only 12 species were identified by both markers (*Arctium lappa, Arnica montana, Betula sp., Camellia sinensis, Glycyrrhiza sp., Ilex paraguariensis, Matricaria chamomilla, Melissa officinalis, Passiflora incarnata, Paullinia cupana, Raphanus sativus* and *Senna alexandrina*) (Figure 1). Most of the features were assigned to the species taxonomic level, however, in some genera, such as *Glycyrrhiza* and *Mentha*, the low interspecific variability did not allow the species to be identified. On average, 70% of the declared ingredients have been identified in the analysed products (see Table 3). Despite this, some products belonging to the same companies have a higher percentage of ingredients identification (e.g. company 5, n=100%) than others (e.g. company 3, n=44%) as can be seen in Figure 2. Overall, our data reported a mixed composition reflecting, at least in part, the complexity of the herbal teas. DNA metabarcoding results suggested that in most cases the declared species were detected. In all samples, except for HT_013, we also detected undeclared elements. For example, in the HT_001 sample, we found a high percentage (n=17%) of *Pimpinella anisum*, a plant typically used in herbal teas but not declared on the label. It is interesting to note that in the sample HT_002, coming from the same company (Company 1), the presence of *Pimpinella anisum* is declared on the label and we found the presence of this ingredient. Also in the HT_012 sample a high percentage (n=21%) of a species commonly used for botanicals, *Melilotus albus*, was found.

ID LAB	**Declared species**	**Assigned species (*psbA-trnH*)**	**Assigned species (ITS2)**
HT_001	*Agropyron Repens Beauv. 20%, Taraxacum officinale Weber 20%, Arctium Iappa L. 15%, Cichorium intybus L. 15%, Melissa officinalis L. 15%, Cynara scolymus L. 15%*	*Melissa officinalis 52%, Arctium lappa 46%, Reichardia ligulata 2%*	*Melissa officinalis 51%, Pimpinella anisum 17%, Althaea officinalis 10%, Arctium lappa 8%, Helminthotheca echioides 5%, Arctium tomentosum 2%, Cynodon dactylon 3%, Taraxacum officinale 3%*
HT_002	*Foeniculum vulgare Mill. 20%, Glycyrrhiza glabra L. 20%, Pimpinella anisum L. 20%, Mentha piperita L. 20%, Citrus sinensis L. var. dulcis 15%, Matricaria chamomilla L. 5%*	*Glycyrrhiza sp. 69%, Mentha sp. 26%, Matricaria chamomilla 4%, Aloysia citrodora 1%.*	*Foeniculum vulgare 54%, Pimpinella anisum 37%, Glycyrrhiza glabra 7%, Matricaria chamomilla 2%*

HT_003	*Camelia sinensis (L.) Kuntze 20%, Prunus Cerasus L. 20%, Citrus limon Burm. F. 20%, Betula pendula Roth 15%, Agropyron Repens Beauv. 15%, Vitis vinifera L. 10%*	*Betula sp. 49%, Camellia sinensis 30%, Vitis vinifera 21%*	*Betula sp. 90%, Camellia sinensis 2%, Chenopodium album 1%, Cynodon dactylon 2%, Filipendula ulmaria 3%, Polyspora axillaris 1%, Tilia platyphyllos 1%.*
HT_004	*Senna alexandrina 40%, Rhamnus frangula 20%, Matricaria camomilla, Foeniculum vulgare Mill.*	*Senna alexandrina 97%, Rhamnus frangula 1%, Matricaria chamomilla 2%*	*Foeniculum vulgare 52%, Matricaria chamomilla 21%, Senna alexandrina 23%, Capsella bursa-pastoris 2%, Chrozophora tinctoria 2%,*
HT_005	*Echinacea angustifolia 30%, Citrus x limon, Althaea officinalis, Rosa canina, Hibiscus sabdariffa, Sambucus nigra 10%*	*Echinacea angustifolia 94%, Monstera deliciosa 4%, Portulaca oleracea 1%, Rumex obtusifolius 1%*	*Althaea officinalis 74%, Echinacea angustifolia 26%.*

HT_006	*Urtica dioica 30%, Arctium lappa L. 20%, Taraxacum officinalis Weber, Citrus x limon, Malva officinalis*	*Arctium lappa 62%, Senna alexandrina 24%, Galium sp. 4%, Lathyrus pratensis 2%, Mentha sp. 4%, Rumex obtusifolius 4%.*	*Urtica dioica 43%, Malva sp. 35%, Taraxacum officinale 19%, Foeniculum vulgare 2%, Matricaria chamomilla 1%*
HT_007	*Camellia sinensis (L.) 51%, Mentha 29%, Glycyrrhiza glabra L. 8,25%, Mentha piperita 3,9%, Aloe vera*	*Mentha sp. 74%, Camellia sinensis 17%, Glycyrrhiza sp. 9%*	*Glycyrrhiza glabra 84%, Amaranthus viridis 2%, Camellia sinensis 5%, Convolvulus arvensis 5%,Ipomoea sp. 1%, Morus alba 2%, Polyspora axillaris 1%*
HT_008	*Camellia sinensis (L.) 62,9%, Zingiber officinalis 22%, Peach 1%, Ginseng 1%, Aloe vera.*	*Camelia sinensis 100%*	*Camellia sinensis 30%, Zingiber officinale 23%, Eleutherococcus senticosus 1%, Ocimum sp. 25%, Polyspora sp. 10%, Ipomoea sp. 7%,*

			Achyranthes aspera 1%, Erechtites hieraciifolius 1%, Setaria palmifolia 1%
HT_009	*Zingiber officinale Roscoe, Citrus limon L., Malva sylvestris L., Cymbopogon citratus Stapf, Glycyrrhiza glabra L.*	*Glycyrrhiza glabra L. 100%*	*Malva sylvestris 94%, Glycyrrhiza glabra 3%, Ocimum sp. 2%, Zingiber officinale 1%.*
HT_010	*Matricaria chamomilla L. 44,4%, Melissa officinalis L. 22,2%, Betula pendula Roth. /pubescens Ehrh, Passiflora incarnata L., Lavandula officinalis chaix var 5,6%*	*Matricaria chamomilla 51%, Lavandula sp. 14%, Melilotus albus 11%, Melissa officinalis 6%, Passiflora sp. 16%, Raphanus sativus 2%.*	*Matricaria chamomilla 93%, Melilotus officinalis 1%, Melissa officinalis 2%, Passiflora incarnata 1%, Raphanus sativus 3%.*
HT_011	*Illicium verum - frutto 27%, Mentha piperita 25%, Melissa officinalis, Glycyrrhiza glabra,*	*Glycyrrhiza glabra 60%, Lavandula sp. 25%, Mentha sp. 7%, Illicium verum*	*Glycyrrhiza glabra 90%, Melissa officinalis 8%, Lavandula sp. 2%*

	Lavandula officinalis, Cinchona officinalis, Gentiana lutea 2%.	*7%, Melissa officinalis 1%*	
HT_012	*Foeniculum vulgare 40%, Illicium verum 40%, Carum carvi, Mentha piperita 9%.*	*Mentha sp. 41%, Eschscholzia californica 20%, Melilotus albus 21%, Melissa officinalis 18%,*	*Foeniculum vulgare 73%, Carum carvi 27%*
HT_013	*Senna alexandrina 40%; Rhamnus frangula 15%; Matricaria chamomilla L. 15%; Foeniculum vulgare 15%; Malva officinalis 15%*	*Senna alexandrina 98%, Rhamnus frangula 1%, Matricaria chamomilla 1%.*	*Foeniculum vulgare 38%, Malva sp. 34%, Matricaria chamomilla 16%, Senna alexandrina 12%*
HT_014	*Passiflora incarnata L., Escholtzia californica Cham., Matricaria chamomilla L., Tilia platyphyllos Scop., Ocimum basilicum L.*	*Eschscholzia californica 53%, Passiflora incarnata 30%, Matricaria chamomilla 4%,*	*Matricaria chamomilla 83%, Passiflora incarnata 7%, Panicum miliaceum 2%, Papaver rhoeas 1%, Tilia sp. 4%, Vicia villosa 1%,*

		Ocimum basilicum 13%	*Capsella bursa-pastoris 2%*
HT_015	*Camellia sinensis K., Filipendula ulmaria Max, Foeniculum vulgare Mill, Mentha spicata L.*	*Mentha sp. 69%, Portulaca oleracea 17%, Camellia sinensis 12%, Eschscholzia californica 2%*	*Filipendula ulmaria 50%, Foeniculum vulgare 47%, Digitaria ciliaris 2%*

Table 3. List of the detected ingredients in market samples, based on DNA metabarcoding assignment. For each sample, the declared species, assigned species for the DNA barcode marker *psbA-trnH* and ITS2 and the company are indicated. The percentage values refer to the relative abundance of HTS reads for each recognized ingredient.

Common infesting herbaceous species such as *Cynodon dactylon, Helminthotheca echioides, Chenopodium album, Digitaria ciliaris, Panicum miliaceum,* and *Lathyrus pratensis* were found in several samples in a range of percentage of 2-5%. The presence of such plants could be harmful, as some species could be poisonous (Cornara et al., 2018) or cause allergies (Frigerio et al., 2019).

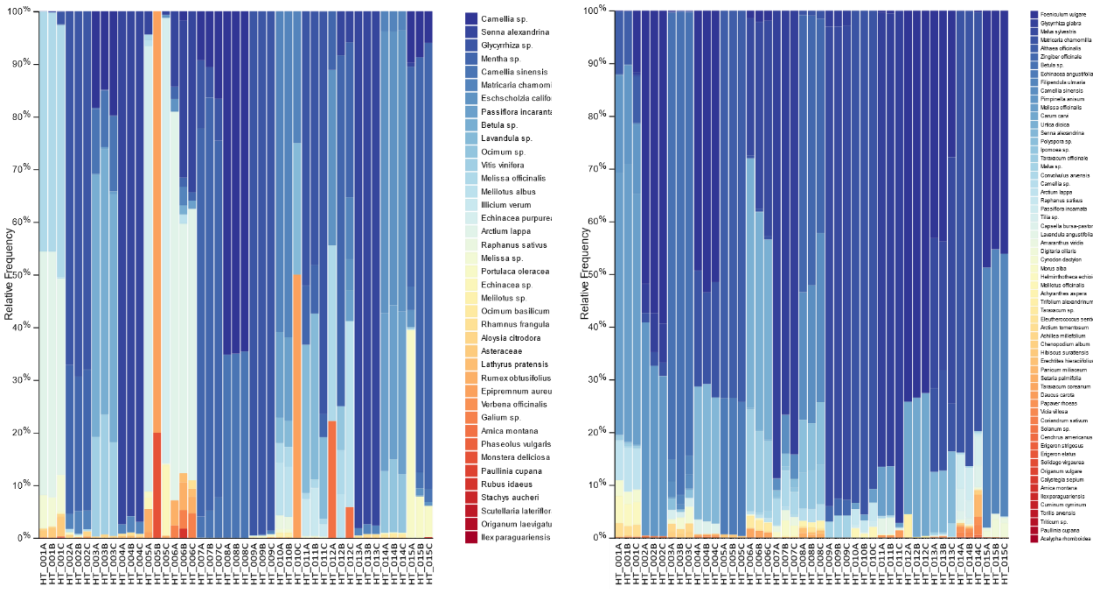


Figure 1. Relative abundance of the plant taxa recovered in the 15 herbal tea products through *psbA-trnH* (left) and ITS2 (right) metabarcoding sequencing.

The ITS2 marker appears to identify more species (n=60) than psbA-trnH (n=35), with an ingredient identification rate of 49% for ITS2 versus 42% for psbA-trnH. However, the use of both markers has made it possible to almost double the capacity for identifying and assigning species, reaching for some samples (HT_013-HT_015) to identify 100% of the declared species on the label. Only 12 species have been identified by both markers. It has been noticed that some species are identified only by one marker rather than the other. Specifically, the species *Eschscholzia californica, Mentha sp., Vitis vinifera, Illicium verum and Rhamnus frangula* species were identified only with the *psbA-trnH* marker, while the species *Althea officinalis, Epilobium angustifolium, Foeniculum vulgare, Malva sp., Pimpinella anisum, Solidago virgaurea, Taraxacum officinale, Urtica dioica* and *Zingiber officinale* were identified only with the ITS2 marker. This result appears to be reproducible

between different samples from different companies. Furthermore, it was also confirmed by the fact that *Althea officinalis* and *Solidago virgaurea* species were present in the mock created by us in the laboratory and were never recognized by the *psbA-trnH* marker. In order to exclude the absence of ingredient identification due to a lack in the database, for each species the presence of sequences deposited in the NCBI database (https://www.ncbi.nlm.nih.gov/) was verified.

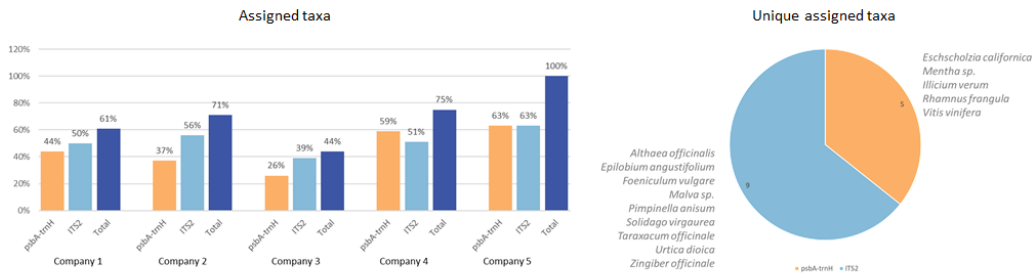
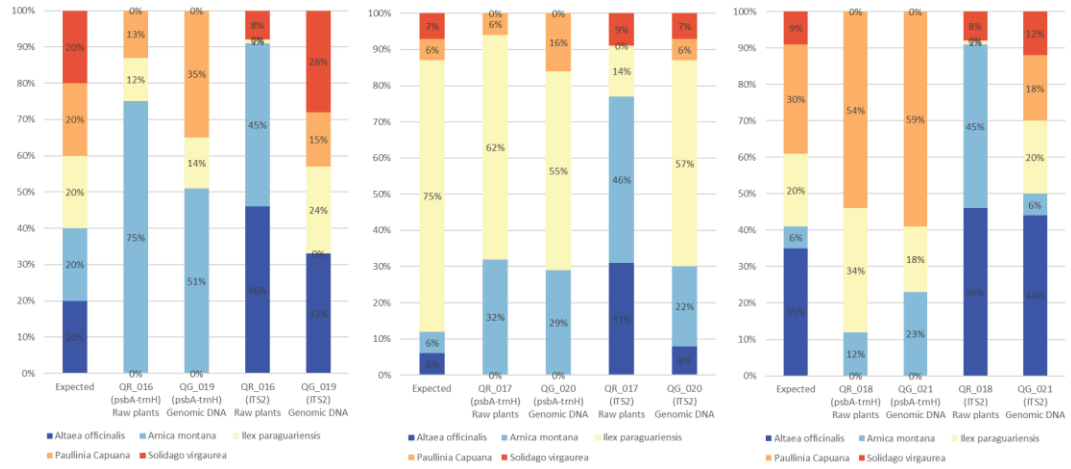


Figure 2. On the left) In the graphic are shown in orange the percentage of assigned taxa declared in the label for the barcode marker psbA-trnH, in light-blue percentage of assigned taxa declared in the label for the barcode marker ITS2 and in blue the total percentage using both markers for all the companies. On the right) The pie chart shows the unique assigned taxa for the markers ITS2 in light blue (n=9) and for psbA-trnH in orange (n=5).

DNA metabarcoding for mock mixture quantification

The following species were used for mock mixture creation: *Althaea officinalis* (roots), *Arnica montana* (flowers), *Ilex paraguariensis* (leaves), *Paullinia capuana* (seeds) and *Solidago virgaurea* (aerial parts). As shown in Figure 3, by analysing both DNA barcode regions (i.e., *psbA-trnH* and ITS2), we were able to identify and correctly assign each plant used for the mock mixture preparation. Nevertheless,

through the analysis of the barcode region *psbA-trnH* we were not able to identify the species *Althaea officinalis* and *Solidago virgaurea*, both in the biomass mixture and in the genomic DNA mixture. Similarly, with the analysis of ITS2 barcode region we did not find the species *Paullinia cupana* in the mock mixture created from raw plant. Probably this was due to the low yield of genomic DNA (see Supplementary Table 1). Because *Paullinia cupana* was found in the gDNA mixture, this result confirms the bias due to the DNA extraction phase. To verify if a mock mixture created from gDNA is more effective in quantification than the raw plant's mock mixture, a PCoA analysis with Bray-Curtis distance was performed. As shown in Figure 3, gDNA samples (red points) are closer to the real samples (in blue), compared to the mock mixture obtained starting from the raw plants (in green), considering their composition obtained both with psbA-trnH and ITS2. Furthermore, a linear mixed-effect model was performed to compare the distance between the gDNA mocks mixture and the corresponding real samples (Fig. 3) with that occurring between the mocks mixture and the real samples. The distance of the gDNA group from real samples is significantly lower compared with that of raw plant mock mixture ($p < 0.001$), confirming that gDNA samples are significantly closer to the corresponding real samples.



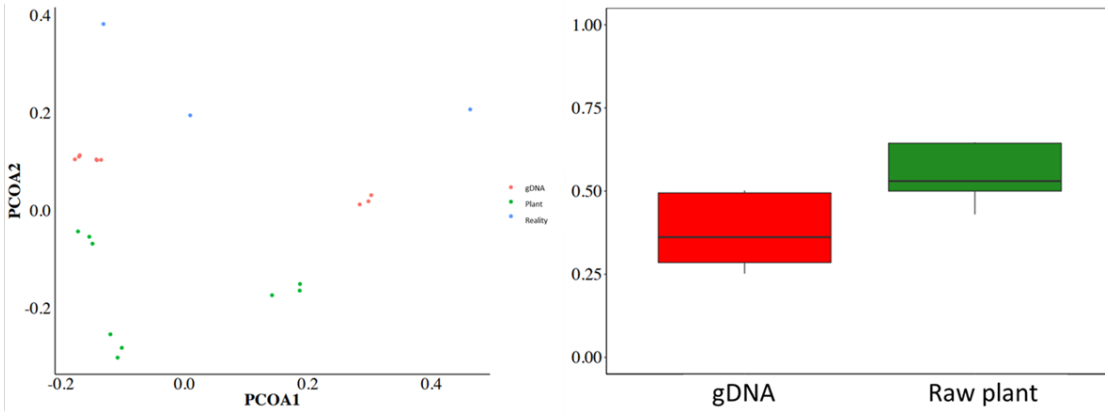


Figure 3. The figure above is a graph showing the expected percentages and the percentages obtained after HTS sequencing for the *psbA-trnH* and ITS2 barcode markers both for the mock mixture created starting from raw plants and genomic DNA. The figures below show the results of the ordination analyses. In the PCoA graph, the blue dots represent the real samples, the green ones indicate the mock mixture created starting from the raw plants and with the red ones samples from the mock mixture of genomic DNA are reported. On the right, the boxplots show the Bray-Curtis distance between the gDNA mocks mixture and the corresponding real samples (in red) and the distance between the mocks mixture starting from the plants and the real samples (in green).

Discussion

A multi-marker approach

In this study it was chosen to proceed with a multimarker approach. Previous studies have used a multimarker approach for mesozooplankton DNA analysis (Stefanni et al., 2018), for dietary analysis of animal (da Silva et al., 2018) and for traditional medicine (Arulandhua et al., 2019). As plants have low intraspecific variability, the use of more than one DNA barcoding marker increases the chance of identifying ingredients at the species level. Our analyses confirm our hypothesis, for example, in the sample HT_007: only by using psbA-trnH we were able to assign the genera Glycyrrhiza sp., but with the addition of the ITS2 marker we were able to arrive at the species level of *Glycyrrhiza glabra*. Additionally, our result showed that *psbA-trnH* and ITS2 markers have a complementary output, with 12 shared taxa, but 5 and 9 unique taxa for *psbA-trnH* and ITS2, respectively. A similar result was shown by Arulandhua and colleagues (Arulandhua et al., 2019) that demonstrate that a multi-marker approach increases the resolution and the quality results.

Since the assignment of different species for the two markers also occurred in the mock created in the laboratory, in which we are sure that the declared species are present, one of the problems could be related to the absence of sequences in the database. However, the presence of these sequences in the NCBI database was checked, and they are present for almost all the species under analysis and for both markers. For species where the sequence deposited in the database was not present, as in the case of *Paullinia cupana*, we used sequences from our private databases. Nevertheless, further studies are needed to verify whether this phenomenon is reproducible.

Since using the ITS2 marker analysis we were able to identify a greater number of taxa (60 vs 35) and given its extensive use in the literature it is probably the best marker. This is in line with Chen et al. (2010) research, who proposed that ITS2 can be a universal barcode to identify plants at species level, especially for botanicals plants. However, this result could be due to the type of sample matrix or to the taxa present in the sample analysed. Future studies could expand the number of taxa and analyze different parts of the plant for each taxa, to evaluate whether the matrix can actually affect the amplification preferences of one marker rather than another.

As regards the contaminants detected in the products, since many of them are unique to that product, we can deduce that it is possible that contamination occurred along the supply chain of the manufacturing company. For this reason, HTS analysis can be a method for controlling not only raw materials but also any contamination, which can be both a quality and food safety problem.

Quantitative ability of High-Throughput Sequencing

In the scientific literature, there is a debate about the ability of HTS to provide quantitative identification. Lamb and colleagues (Lamb et al., 2019) interpreted the relative features abundance as the relative abundance of biomass (Lamb et al., 2019). The main obstacle concerns all the biases that occur during the analysis such as DNA extraction, PCR and bioinformatics analysis (Piñol et al., 2019). In our previous study (Bruno et al., 2019), we noticed that amplicon DNA metabarcoding efficacy could be biased by the PCR amplification step using “universal” markers and the occurrence of bias during PCR amplification may cause the inaccurate estimation of quantities. Also Krehenwinkel and colleagues found bias in differential amplification due to priming efficiency during PCR. They suggest using degenerate primers and/or

target amplicons with high priming site conservation (Krehenwinkel et al., 2017). Our analysis suggests that DNA metabarcoding has a relative quantitative ability, even if there are some biases in the identification of all species. Moreover, starting from a mixture of gDNA a more accurate quantification compared to the raw plant mixture is obtained, thus demonstrating the presence of a bias in the DNA extraction phase. Nevertheless, using two DNA barcode regions allows overcoming one of the identified biases.

In future works, it may be useful to test different parts of the same plant, to verify whether extraction bias can affect the amplification of DNA by one marker rather than the other, both quantitatively and qualitatively. Furthermore, it would be advisable to test different processing levels of a food product, to evaluate the integrity of the DNA and assess whether it can affect the amplification capacity.

Conclusions

The authenticity of a herbal product is a major concern for consumers, producers, processors and food authorities. The substitution of raw materials with others of lower value can reduce and sometimes eliminate the therapeutic efficacy of the herbal product, thus causing a loss of consumer confidence. Furthermore, botanicals fraud can represent potential health risks to consumers (Grazina et al., 2020). In this context, biomolecular analysis has become more important year after year (Galimberti et al., 2019). High-Throughput DNA Sequencing Technologies (HTS) offered the opportunity to analyse multiple DNA amplicons by sequencing them in parallel (Haynes et al., 2019). The application of HTS allowed us to detect and identify the plant composition of herbal commercial teas. Although HTS technology has some critical aspects such as the quality of the extracted DNA (Grazina et al., 2020) or the relative ability to quantify all the ingredients, this study shows the value of

the application of HTS analysis for a quality control tool and routine monitoring analysis, from the characterization of the raw ingredients to the final processed products. Furthermore, the use of a multimarker approach has allowed identifying a greater number of species within a sample and it is therefore advisable, for future work, to use more than one marker to increase the identification rate. In conclusion, this tool can be applied to a wider range of botanicals to improve the traceability of all products. HTS analysis has such a sensitivity that it can find even small quantities of plants that can be potentially poisonous or harmful to health, so this tool has great potential in quality and safety control in the field of herbal teas. For this reason, it is desirable to implement this analysis by EFSA or other control agencies. As these agencies take long periods to implement analysis, it may be appropriate for companies to start using these tools for preventive control of their supply chain and their products before it becomes mandatory.

Supplementary

ID LAB	ng/ul
HT_001	[5-7]
HT_002	[4-11]
HT_003	[4-5]
HT_004	[2-8]
HT_005	[2-5]

HT_006	[2-4]
HT_007	[10-14]
HT_008	[10-22]
HT_009	[3-6]
HT_010	[3-5]
HT_011	[4-5]
HT_012	[7-27]
HT_013	[2-4]
HT_014	[5-11]
HT_015	[11-15]
QR_016	[6-9]
QR_017	[3-5]
QR_018	[6-8]
QG_019	[31]
QG_020	[23]
QG_021	[36]

Supplementary table 1. In the table are indicated the concentration of genomic DNA (ng/ul)

Primer name	5'-3'	Reference
S2_F	ATG CGA TAC TTG GTG TGA AT	Chen et al., 2010
S2_R	GGA CGC TTC TCC AGA CTA CAA T	
psbA	GTT ATG CAT GAA CGT AAT GCT C	Newmaster and Subramanyam, 2017
trnH	GCG CGC ATG GTG GAT TCA CAA TCC	

Supplementary table 2. In the table are indicated the primer sequences for psbA-trnH and ITS2 amplification. To every primer was added the Illumina adaptor.

Reference

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of molecular biology, 215(3), 403-410.

Arulandhu, Alfred J., et al. "The application of multi-locus DNA metabarcoding in traditional medicines." Journal of Food Composition and Analysis 79 (2019): 87-94.

Anthoons, B., Karamichali, I., Schrøder-Nielsen, A., Drouzas, A. D., de Boer, H., & Madesis, P. (2020). Metabarcoding reveals low fidelity and presence of toxic species in short chain-of-commercialization of herbal products. Journal of Food Composition and Analysis, 103767.

Biella, P., Tommasi, N., Akter, A., Guzzetti, L., Klecka, J., Sandionigi, A., ... & Galimberti, A. (2019). Foraging strategies are maintained despite workforce reduction: A multidisciplinary survey on the pollen collected by a social pollinator. PloS one, 14(11), e0224037.

Bolson, M., Smidt, E. D. C., Brotto, M. L., & Silva-Pereira, V. (2015). ITS and trnH-psbA as efficient DNA barcodes to identify threatened commercial woody angiosperms from southern Brazilian Atlantic rainforests. PloS one, 10(12), e0143049.

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C., Al-Ghalith, G. A., ... & Bai, Y. (2018). QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science (No. e27295v2). PeerJ Preprints.

Cornara, L.; Smeriglio, A.; Frigerio, J.; Labra, M.; Di Gristina, E.; Denaro, M.; Mora, E.; Trombetta, D. The problem of misidentification between edible and poisonous wild plants: Reports from the Mediterranean area. Food and Chemical Toxicology 2018, 119, 112–121.

Elbrecht, V.; Leese, F. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—Sequence relationships with an innovative metabarcoding protocol. PLoS ONE 2015, 10, e0130324. [CrossRef]

da Silva, Luís P., et al. "Advancing the integration of multi-marker metabarcoding data in dietary analysis of trophic generalists." Molecular ecology resources 19.6 (2019): 1420-1432.

Frigerio, J.; Pellesi, R.; Mezzasalma, V.; De Mattia, F.; Galimberti, A.; Lambertini, F.; Suman, M.; Zanardi, S.; Leporati, A.; Labra, M. Development of a DNA Barcoding-Like Approach to Detect Mustard Allergens in Wheat Flours. Genes 2019, 10, 234.

Garzo, C. F., Gómez, P. P., Barrasa, A. B., Martínez, R. A., Ramírez, R. F., & Ramón, F. R. (2002). Cases of neurological symptoms associated with star anise consumption used as a carminative. Anales espanoles de pediatria, 57(4), 290-294.

Grazina, L., Amaral, J. S., & Mafra, I. (2020). Botanical origin authentication of dietary supplements by DNA-based approaches. Comprehensive Reviews in Food Science and Food Safety, 19(3), 1080-1109.

Haynes, E., Jimenez, E., Pardo, M. A., & Helyar, S. J. (2019). The future of NGS (Next Generation Sequencing) analysis in testing food authenticity. Food Control, 101, 134-143. https://doi.org/10.1016/j.foodcont.2019.02.010

Herbal supplement sales in US increase 8.5% in 2017, topping $8 billion Herbal-Gram, 119 (2018), pp. 62-71

https://www.efsa.europa.eu/it/efsajournal/pub/4908

Lu, Y., Jiao, L., He, T., Zhang, Y., Jiang, X., & Yin, Y. (2020). An optimized DNA extraction protocol for wood DNA barcoding of Pterocarpus erinaceus. IAWA Jour-nal, 1(aop), 1-16.

Lu, Z., Rubinsky, M., Babajanian, S., Zhang, Y., Chang, P., & Swanson, G. (2018). Visualization of DNA in highly processed botanical materials. Food Chemistry, 245, 1042–1051. https://doi.org/10.1016/j.foodchem.2017.11.067

Marieschi, M., Torelli, A., Poli, F., Sacchetti, G., & Bruni, R. (2009). RAPD-based method for the quality control of Mediterranean oregano and its contribution to pharmacognostic techniques. Journal of Agricultural and Food Chemistry, 57(5), 1835-1840.

Mishra, P., Kumar, A., Nagireddy, A., Shukla, A. K., & Sundaresan, V. (2017). Evaluation of single and multilocus DNA barcodes towards species delineation in complex tree genus Terminalia. PloS one, 12(8), e0182836.

Mosa, K. A., Soliman, S., El-Keblawy, A., Ali, M. A., Hassan, H. A., Tamim, A. A., & Al-Ali, M. M. (2018). Using DNA barcoding to detect adulteration in different herb-al plant-based products in the United Arab Emirates: proof of concept and valida-tion. Recent Patents on Food, Nutrition & Agriculture, 9(1), 55-64.

Stefanni, Sergio, et al. "Multi-marker metabarcoding approach to study mesozooplankton at basin scale." Scientific reports 8.1 (2018): 1-13.

Steven G, Newmaster, & Subramanyam, R. (2009). Testing plant barcoding in a sister species complex of pantropical Acacia (Mimosoideae, Faba-ceae). Molecular ecology resources, 9, 172-180.

Nithaniyal, S., Vassou, S. L., Poovitha, S., Raju, B., & Parani, M. (2017). Identification of species adulteration in traded medicinal plant raw drugs using DNA barcoding. Genome, 60(2), 139-146.

Paranaiba, R. T., Carvalho, C. B., Paiva, R. S., Trindade, B. R., Barros, M. G., Souza, E. P., ... & Silveira, D. (2020). DNA from wood-A simple approach facing a challenging matrix-A preliminary study. Forensic Science International, 314, 110371.

Piñol, J., Senar, M. A., & Symondson, W. O. (2019). The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. Molecular ecology, 28(2), 407-419.

Raclariu, A. C., Paltinean, R., Vlase, L., Labarre, A., Manzanilla, V., Ichim, M. C., ... & de Boer, H. (2017). Comparative authentication of Hypericum perforatum herbal

products using DNA metabarcoding, TLC and HPLC-MS. Scientific reports, 7(1), 1-12.

Richardson, R. T., Lin, C. H., Sponsler, D. B., Quijia, J. O., Goodell, K., & Johnson, R. M. (2015). Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. Applications in Plant Sciences, 3(1), 1400066.

Rognes, T., Flouri, T., Nichols, B., & Quince, C. Mah e, F.(2016). VSEARCH: A versatile open source tool for metagenomics

Shokralla, S.; Spall, J.L.; Gibson, J.F.; Hajibabaei,M. Next-generation sequencing technologies for environmental DNA research. Mol. Ecol. 2012, 21, 1794–1805.

Smith et al., 2018 T. Smith, K. Kawa, V. Eckl, C. Morton, R. Stredney

Wiedenfeld, H. (2011). Plants containing pyrrolizidine alkaloids: toxicity and problems. Food Additives & Contaminants: Part A, 28(3), 282-292.

Due to the desire to patent the ready-to-use kits, the results of the project have not been published

Ready-to-use kit for food authentication: the cases of truffle

Introduction

Truffle is a product with high added value, and processed products such as sauces and gravies can be cases of fraud or adulteration (Segelke et al., 2020). Truffle is the common name given to underground mushrooms, belonging to the genus Tuber and consequently to the Tuberceae family. The species of truffles are different, including for example, *Tuber magnatum* (white truffle), *Tuber melanosporum, Tuber brumale, Tuber aestivum, Tuber albidum pico, Tuber uncinatum, Tuber mesentericum* etc. The *T. magnatum* (white truffle) is considered the most valuable and has a high economic value of about 220 € / 100 gr. To morphologically distinguish truffles from each other or to identify possible fraud, an expert morphologist in the field is required, which often results in a waste of money and time. The use of rapid DNA-based analyses could give cheaper and more accurate results. The purpose of this project is to develop a rapid and reliable analyses for the evaluation of quality and safety in the agri-food sector. The main goal is to develop diagnostic kits to be internalized and used in the quality control of companies without the need for dedicated equipment or qualified personnel and reducing costs and time. These kits consist in the development of a rapid extraction method that can replace the standard extraction kits and in the replacement of the classic PCR with the LAMP technique.

Material and Methods

Samples collection

A total of seven fresh truffles of different species (see table 1) were collected after morphological identification by an expert.

Sample	Species
DIF 65	*Tuber aestivum*
DIF 66	*Tuber albidum pico*
DIF 83	*Tuber uncinatum*
DIF 123	*Tuber magnatum*
DIF 124	*Tuber melanosporum*
DIF 125	*Tuber mesentericum*
DIF 163	*Tuber brumale*

Table 1. List of truffles collected for this project

DNA Barcoding analysis

In order to confirm the species declared, all samples were tested by DNA barcoding. Genomic DNA was extracted from the samples using the Plant Genomic DNA Extraction Kit (Fisher Molecular Biology), according to the manufacturer's

instructions. The amplification of the DNA barcode region was carried out using the PCR Mix Plus kit (A&A BIOTECHNOLOGY). The reaction mix contains a total volume of 25 µl of which: 12.5 µl of PCR Mix Plus which contains 0.1 U / µl Taq DNA polymerase, 4 mM MgCl2, 0.5 mM of each dNTPs, anti-DNase; 1 µl of each primer at a concentration of 10 mM, 7.5 µl of sterile H2O and finally 3 µl of 30-50 ng of genomic DNA. For amplification, an "Eppendorf Mastercycler ep gradient" thermal cycler was used according to the following thermal program: Preliminary denaturation at 95 ° C for 5 ', 35 cycles of denaturation at 95 ° C for 45 ', annealing at a temperature dependent on the primer pair for 45 , extension to 72 ° C for 60 '' and a final extension at 72 ° C for 7 '.

The primers used are described in the following table.

Regione marker	**Sequenza 5'-3'**	**TA (temp di annealing)**
ITS KYO	GATGAAGAACGYAGYRAA	60 °C
ITS u4	RGTTTCTTTTCCTCCGCTTA	60 °C

Table 2. Primer used for DNA barcoding analysis

The amplifiers were visualized into a 1.5% agarose gel stained with ethidium bromide. Purified amplicons were bidirectionally sequenced using an ABI 3730XL automated sequencing machine at Eurofins Genomics (Ebersberg, Germany). The 3' and 5' terminal portions of each sequence were clipped to generate consensus sequences for each sample. For all the tested samples, the reliability of DNA barcoding identification was assessed by adopting a standard comparison approach against a GenBank database with BLASTn. Each sequence was taxonomically assigned to the plant species with the nearest matches (maximum identity >99% and query coverage of 100%).

Methods of rapid DNA extraction

One of the goals of this work is to develop a rapid DNA extraction method. Several quick methods have been tested outlined below. Only two truffles (DIF 065 and DIF 83) were tested for all methods.

- Method 1: Rapid extraction with commercial saline lysis buffer
  A small amount of truffle is added to each 2 mL tube (a tip of about 3 mm). Add 50 µl of commercial lysis buffer and incubate for 20' at 65 ° C.
- Method 2: Rapid extraction with 50 mM NaOH caustic soda. A small amount of truffle (a tip of about 3 mm) and 25 µl of 50mM NaOH is added to each 2 mL tube. After stirring by vortexing the tubes are left at room temperature for 10 '. After incubation, 25 µl of Tris-HCl are added.
- Method 3: Rapid extraction using a microwave oven (Ramya et at., 2018). Add 200 µl of physiological water to a small amount of truffle (a tip of about 3 mm). After stirring by vortexing the 2 mL tubes are left in the microwave for 1'30 "at 750 W. Add another 100 µl of physiological water and repeat the procedure as above.
- Method 4: Extraction with CTAB swab. The extraction method described by Doyle & Doyle (1990) is based on the use of Hexadecyltrimethylammonium Bromide (CTAB), a detergent that in the presence of high saline concentrations binds to DNA forming a stable and water-soluble complex. Small quantities of white truffle are added to 5 2 mL tubes. Mechanical shredding is carried out using a tip. 200 µl of the CTAB solution (CTAB 2%, NaCl 1.4 M, EDTA 20 mM, TRIS-HCl 100 mM pH 8) are taken and placed in each tube. These samples are then incubated at 65 ° C in 5 different times: 15 ', 30' 1 h,

1h and 30 'and 2 h. After the rapid extraction procedure, the DNA of the 5 samples were quantified using the Qubit fluorometer.

- Method 5: Buffer A + Buffer B (Edwards et al., 1991). Buffer A is prepared which contains: 200 mM TRIS-HCl, 250 mM NaCl, 25 mM EDTA, 0.5% SDS, pH 7.5 and a Buffer B which contains: 10 mM TRIS-HCl, 1 mM EDTA, pH 8. 9 parts of Buffer B and 1 part of Buffer A are placed in a 2 mL tube. The truffle is divided into 5 different 2 mL tubes, in which an increasing amount of buffer A + B has been placed, specifically 20 / 50/75/100/200 µl.
- Method 6: Buffer A + Buffer B + PVP. This method consists of using the buffers of method 5 with the addition of Polyvinylpyrrolidone (PVP). Polyvinylpyrrolidone is a water-soluble polymer of formula (C6H9NO) n, consisting of monomers of 1-vinyl-2-pyrrolidone. PVP is added into the extraction buffer mixture to form hydrogen bonds with polyphenolic compounds that can be separated by DNA (Maliyakal et al., 1992). To prepare the rapid extraction mixture, 1% of PVP, 9 parts of Buffer B and one part of Buffer A are added. For example, a solution with 10 mg of PVP, 900 µl of Buffer B and 100 µl of Buffer A is prepared. The mixture is homogenized for 5 '' in vortex. 50 mg of truffle and 200 µl of the extraction mixture are placed in a new 2 mL test tube. The DNA of each sample is then diluted 1:10.

The DNA of all samples was amplified by PCR under the conditions illustrated previously. After testing rapid extraction methods, the DNA of the samples was quantified using the Qubit fluorometer

Bioinformatic analysis for primer design for LAMP

To carry out the DNA amplification, four specific primers are required, two external (F3 and B3) and two internal (BIP and FIP) which recognize a total of six target DNA

regions and a thermostable DNA polymerase obtained from *Bacillus stearothermophilus* (Bst) which it maintains the polymerase activity in the 5'-3 'direction but lacks the exonuclease activity in the 5'-3' direction. The primers were designed using the online Primer Explorer program (http://primerexplorer.jp/elamp4.0.0/index.html) which, depending on the target nucleotide sequence entered by the user, generates a series of effective primer combinations for the amplification with LAMP technique of the sequence itself. Then the different combinations of primers were analysed to select the one (s) that show the greatest specificity towards the target. The ITS2 nuclear region was chose as a possible candidate to distinguish the white truffle from other truffles, and the region that codes for β-tubulin, being very conserved, for the design of universal primers that amplify all the species of the genus Tuber.

LAMP technique

The mix of the original LAMP reaction has a volume of 25 µl total and is made up as follows: 12.5 µl of Warm Start Colorimetric LAMP 2x Master Mix (New England Biolab) which contains phenol red, which serves as an indicator dye sensitive to pH change (from an initial alkaline pH to a final acid pH) capable of giving a color change (from red to orange / yellow), 4 µl 10 mM of BIP and FIP primers, 0.5 µl 10 mM of primers B3 and F3, 0.5 µl of sterile Milliq water and 3 µl of 30-50 ng of genomic DNA. All analysis was made under a biological hood to limit cross-contamination events. The amplification can be seen with the naked eye, through the color change from pink to yellow. To verify that the color change was due to the actual amplification of the DNA, the amplicons were visualized on agarose gel. A negative sample containing milliq water in place of DNA was analyzed for each amplification reaction.

Results

DNA barcoding

Amplification by PCR (ITS2 region) produced only one band on agarose gel. After Sanger sequencing, the sequences obtained from the electropherogram were compared using the BLAST algorithm and showed 100% identity with the species indicated in Table 3. The obtained DNA barcoding results are showed in the table below.

Samples ID	**Declared species**	**Detected species**
DIF 65	*Tuber aestivum**	*Tuber aestivum**
DIF 66	*Tuber albidum pico*	*Tuber albidum pico*
DIF 83	*Tuber uncinatum**	*Tuber uncinatum**
DIF 123	*Tuber magnatum*	*Tuber magnatum*
DIF 124	*Tuber melanosporum*	*Tuber melanosporum*
DIF 125	*Tuber mesentericum*	*Tuber mesentericum*
DIF 163	*Tuber brumale*	*Tuber brumale*

Table 3: Name of the detected species of the truffle samples analyzed with the DNA barcoding approach Note *: The analyzed regions of *Tuber aestivum* and *T. uncinatum* have the same nucleotide sequence therefore they belong to the same species. The soil and climate conditions can induce *T. aestivum* to bear fruit under different conditions, which in turn could affect the flavor, aroma and morphology of the truffle, thus giving rise to two

morphotypes *(T. aestivum* and *T. uncinatum*) (Paolocci et al., 2004). A greater discriminating factor between the two morphotypes derives from the fact that they are harvested in different seasons: *T. aestivum* from 1st June to 31st August and *T. uncinatum* from 21st September to 31st December (data taken from the National Center for Truffle Studies https: //www.tuber .it /).

Methods of rapid DNA extraction

Two samples of truffles (DIF 065 and 083) were selected to test the rapid extraction of DNA using three different methods: lysis buffer, NaOH and microwave. The DIF 123 sample was chosen to test the rapid extraction using the CTAB buffer performed at an incubation at 65 ° C in 5 different times: 15 ', 30', 1 h, 1.5 h and 2 h. Finally, the rapid extraction method using Buffer A + B + PVP (Method 6 of paragraph 2.3) and by kit were used on all the tubers found for this study. PCR and agarose gel electrophoresis were performed on all selected samples and the results obtained are shown in Table 4.

Sample ID	**Species**	**Extraction method**	**DNA quantification (ng/µl)**	**Positivity on agarose gel**	**Sanger identification**
DIF 65	*Tuber aestivum*	Lysis buffer	11.2	x	/

DIF 65	*Tuber aestivum*	NaOH	3.62	x	/
		Microwave	0.42	/	/
DIF 83	*Tuber uncinatum*	Lysis buffer	17.9	x	/
		NaOH	7.47	x	/
		Microwave	3.23	/	/
DIF 123	*Tuber magnatum*	CTAB 15’	0.18	/	/
		CTAB 30’	0.24	/	/
		CTAB 1h	0.21	/	/
		CTAB 1.5 h	0.21	/	/
		CTAB 2 h	0.27	/	/
DIF 65	*Tuber aestivum*	Buffer A+ B+ PVP	2.3	x	x

DIF 66	*Tuber albidum pico*	Buffer A+ B+ PVP	14.7	x	x
DIF 83	*Tuber uncinatum*	Buffer A+ B+ PVP	3.2	x	x
DIF 123	*Tuber magnatum*	Buffer A+ B+ PVP	0.7	x	x
DIF 124	*Tuber melano-sporum*	Buffer A+ B+ PVP	0.2	x	x
DIF 125	*Tuber mesenter-icum*	Buffer A+ B+ PVP	0.5	x	x
DIF 163	*Tuber brumale*	Buffer A+ B+ PVP	4.5	x	x
DIF 65	*Tuber aestivum*	Commercial kit	96.4	x	x
DIF 66	*Tuber albidum pico*	Commercial kit	20.9	x	x
DIF 83	*Tuber uncinatum*	Commercial kit	26.9	x	x
DIF 123	*Tuber magnatum*	Commercial kit	29.8	x	x

DIF 124	*Tuber melano-sporum*	Commercial kit	47.8	x	x
DIF 125	*Tuber mesen-terium*	Commercial kit	25.4	x	x
DIF 163	*Tuber brumale*	Commercial kit	28.7	x	x

Table 4. In the table are represented the sample ID, the species, the extraction method used, the DNA quantification (ng/µl), the positivity showed on agarose gel and finally Sanger identification.

Lysis buffer: For both samples chosen this method was unsuccessful, as Sanger sequencing produced an unreadable electropherogram.

Sodium hydroxide: This method was discarded because similar results to the lysis buffer method were obtained.

Microwave: For both selected samples it was not possible to obtain well-defined DNA bands by agarose gel electrophoresis.

The CTAB buffer: This extraction method was eliminated as results similar to the microwave method were obtained.

Buffer A + B + PVP: Success is achieved with the extraction method through Buffer A + B + PVP compared to other methods regarding the identification of the species. For all the truffle samples analysed the presence of a well-defined band in agarose gel electrophoresis was visualized and the electropherogram obtained by the Sanger sequencing method was readable. It was possible to identify the species by

the BLAST platform. The declared species was confirmed for all the samples analysed. This extraction method was chosen as the best for our work for a few reasons: It is very fast (once the buffers are prepared it only takes a few minutes to perform the extraction), it has a low cost (no expensive reagents or tools are used), it does not require specialized personnel and finally the presence of PVP retains polyphenols (which can inhibit PCR) which are separated from DNA.

Couples of primer design

Universal primers which amplify all the species belonging to the genus Tuber were designed. These primers were designed on the gene that codes for Beta tubulin. The expected result is the amplification of all the samples.

Primer name	**Sequence 5'-3'**	**Incubation temperature**	**Incubation time**
BIP Tuber	AGC TAA GGA CGT TTT GGA GCG TAC CGG AAC GAC CGA CAA T	65 °C	1 h and 30'
FIP Tuber	TTG GTC GAC AAG TTC AGC ACC CTT CGT CTT TGG ACA ATC CGG		
B3 Tuber	CCA TAC CAG CAC CAG TG		
F3 Tuber	GAA ACC TCT TCA GGC CAG AT		

Table 5. In the table are represented the universal couples of primers designed for LAMP for Tuber sp. amplification.

All species of Tuber spp. Have been amplified as it can be possible to see in the figures below. The order of the samples: DIF 65, DIF 66, DIF 83, DIF 123, DIF 124 and the negative control.

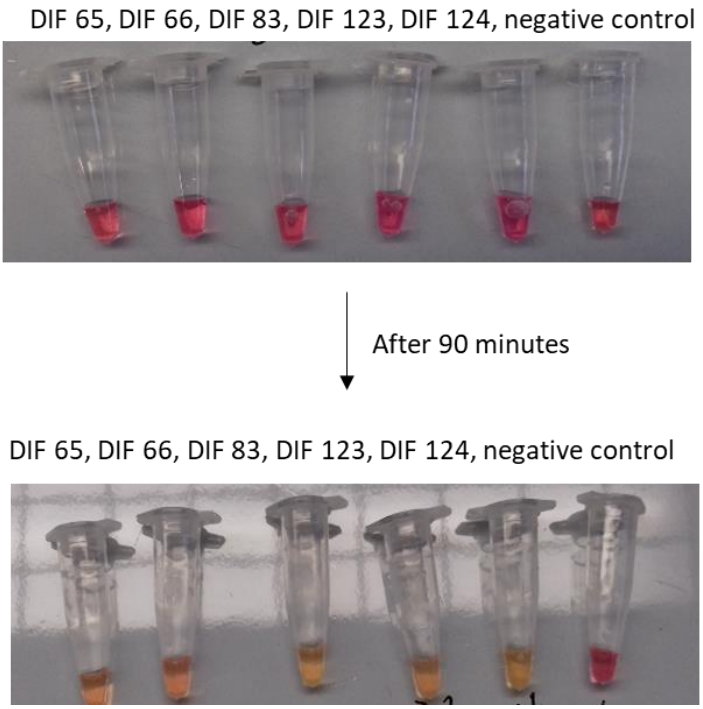


Figure 1. Demonstrative figure before and after the amplification of Tuber spp. After 90 minutes in incubation at 65 ° C all the samples were amplified, except for the negative control.

To test the specificity of the primers towards the Tuber genus and not towards other species of fungi or plants, we did an experiment on different samples including fungi and plants present in our databases. The samples are renamed as follows: FEM 917, FEM 918, FEM 466, FEM 504, DIF 83 (*T. uncinatum* positive control) and the negative control. The expected result is the amplification of the DIF 83 sample.

Samples analyzed:

ID sample	**Species**
FEM 917	*Colletotrichum coccodes*
FEM 918	*Aspergillus spp.*
FEM 466	*Cetraria islandica*
FEM 504	*Betula pendula*
DIF 83	*Tuber uncinatum*

Figure 6. List of samples analysed for test the specificity of primers for Tuber sp.

The order of samples is: FEM 917, FEM 918, FEM 466, FEM 504, DIF 83 and negative control.

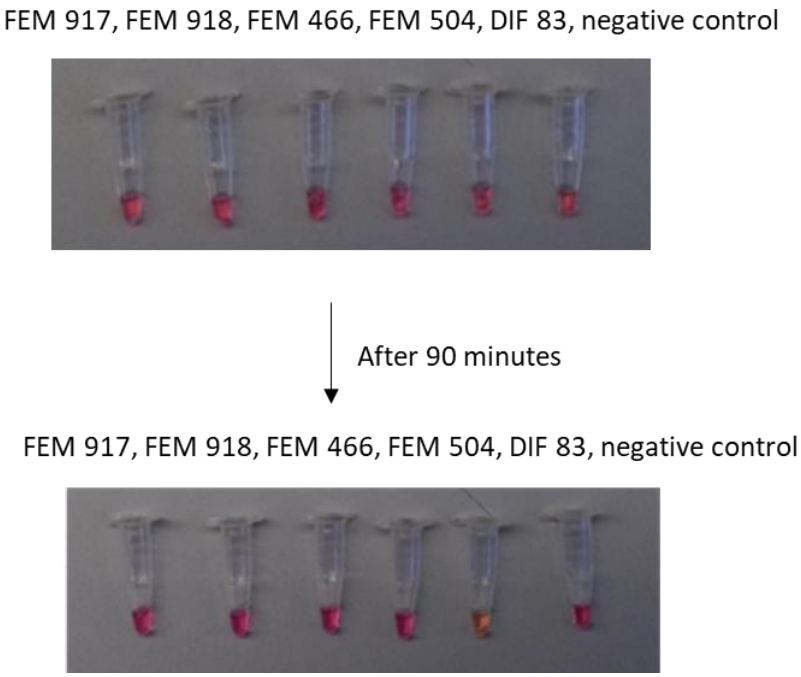


Figure 2. Demonstrative figures before and after amplification. As can be seen in the figures, the result confirms the specificity of the primers, as after 90 minutes in incubation at 65 ° C only the Dif 83 sample (T. uncinatum) was amplified.

In order to develop a method for identifying the white truffle (Tuber ma-gnatum), four primers were designed to carry out the LAMP amplification reaction in a species-specific way.

Primer name	**Sequence 5'-3'**	**Incubation temperature**	**Incubation time**
BIP Tmag	GCG AGA GCC AAG AGA TCC GTT GCC CAT CAG GCC TGA CAA AC	65 °C	45'
FIP Tmag	ATG AAG AAC GCA GCG AAA TGC GGC GCA ATG TGC GTT CAA AG		
B3 Tmag	ATG CCC TGA GGA ATG CCA		
F3 Tmag	TCC TAC CAG CAG TCT GAG AA		

Table 6. In the table are represented the species-specific couples of primers designed for LAMP for *Tuber magnatum* amplification.

All the truffle species in analysis were tested. In order to increase the specificity of the reaction, 1% of DMSO was added. After 45 minutes in incubation at 65 ° C the only amplified sample was *Tuber magnatum* (sample DIF 123). Prolonging the reaction seems to have amplified the DIF 83 and DIF 124 samples. To ensure that the amplification reaction stops, simply bring the sample to room temperature.

The order of the samples: DIF 65, DIF 66, DIF 83, DIF 123, DIF 124 and the negative control.

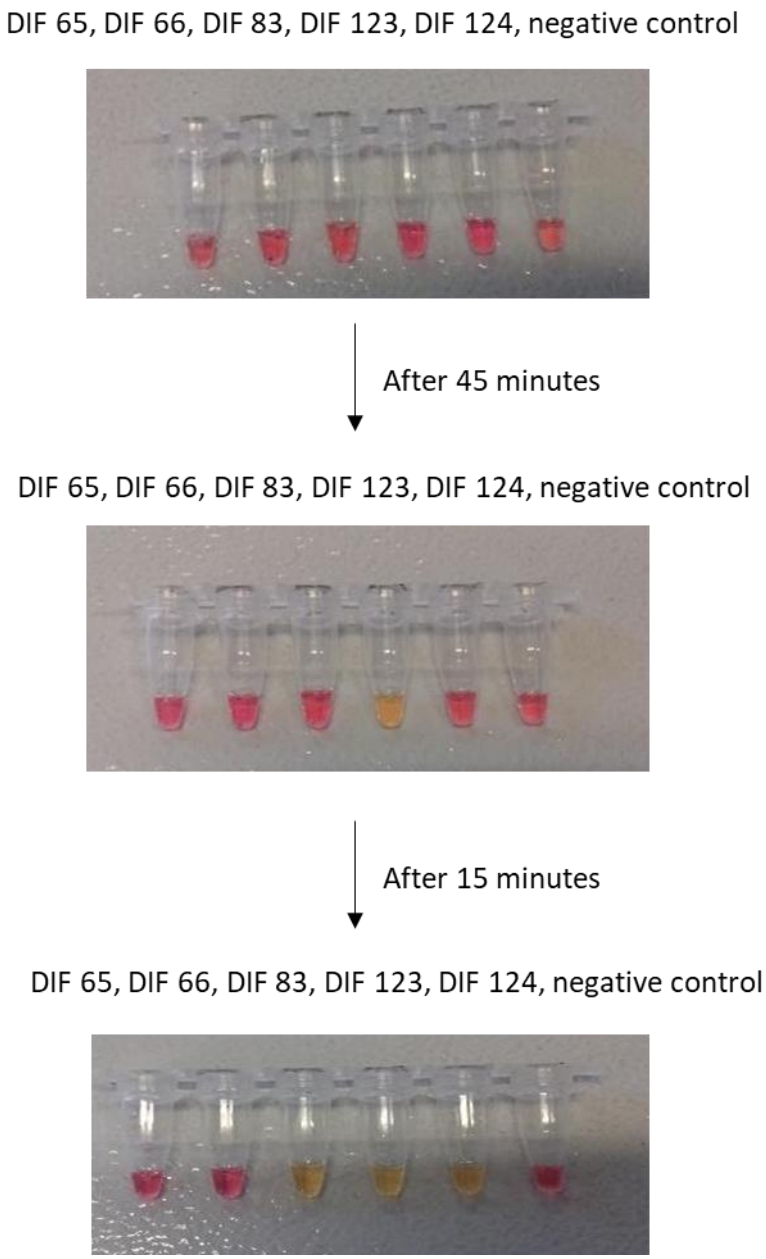


Figure 3. Demonstrative figures before and after amplification. After 45 minutes there is a color change only in the desired sample (DIF 83) but after another 15' also on other samples (DIF 83 and DIF 124) change color.

The primer set specific for Tuber sp. and *T. magnatum* were also tested with samples extracted with the rapid method. The results obtained are the same as those shown previously.

Discussion and conclusion

The aim of this project was the development of rapid and reliable analyses for the assessment of quality and safety in the agri-food sector. The main objective was to develop diagnostic kits to be used in the quality control of companies. More specifically, this work focused on identification of high commercial value truffles. Truffles was collected and morphologically identified by an expert, but it was necessary to verify the species using DNA barcoding. This approach was successful in all the steps and showed a 100% identity of the sequences obtained with those of the expected species.

In order to obtain a rapid and low-cost kit, one of the objectives of this work was to develop a rapid DNA extraction method that can replace commercial extraction kits. In the preliminary phase of the study, several rapid methods were tested on truffle samples. Rapid methods such as rapid extraction with commercial saline lysis buffer, caustic soda, microwaves or the CTAB buffer were discarded as none of them were successful in identifying the species of truffles analysed. PCR performed on truffle samples extracted with lysis buffer or NaOH produced visible bands on agarose gel, however Sanger sequencing did not yield a readable electropherogram. Regarding the microwave extraction method or CTAB buffer, the PCR performed on the DNA samples of the analysed truffles did not produce a well-defined DNA band in agarose gel electrophoresis. These extraction methods have produced reduced amounts of non-pure DNA and these characteristics may have inhibited

amplification by PCR and produce an electropherogram that cannot be read after sequencing. The only rapid extraction method that has shown success in identifying the species was Buffer A + B + PVP. For all the truffle samples analysed the presence of a well-defined band in agarose gel electrophoresis was visualized and the electropherogram obtained from the Sanger sequencing was readable, and it was possible to identify the species using the BLAST platform. The declared species was confirmed for all the samples analysed. This method was, therefore, chosen as the best as it showed several advantages: the extraction takes a few, no reagent is used or expensive tool and finally does not require specialized personnel. A comparison was made between extraction using the standard Plant Genomic DNA kit (Fisher Molecular Biology) and rapid extraction using Buffer A + B + PVP. Extraction using the standard kit produced a greater amount of DNA than rapid extraction, however, with both methods it was possible to identify the species. The rapid method was therefore chosen as the method of choice for the development of a quick and inexpensive kit.

Subsequently, an analysis based on the LAMP technique was developed for the identification of white truffle. Regarding the couples of primers design, different combinations were analysed to select the one (s) that show greater specificity towards the target. To identify the different truffle species, the primer design for the LAMP was performed on the standard barcode regions. The ITS2 nuclear region was chosen as a possible candidate to distinguish the white truffle (*Tuber magnatum*) from other truffles, and the region that codes for β-tubulin, being very conserved, for the design of universal primers that amplify all species of the genus Tuber sp..

In order to set up the method, initially the analyses were performed under a biological hood to limit cross-contamination events. The LAMP was developed to discriminate the species of Tuber spp from other species of fungi or plants and to

identify the white truffle. The primer sets for Tuber sp. and *T. magnatum* were also tested with samples extracted with the rapid method and the results obtained were the same.

Since the main objective of this study was to develop diagnostic kits to be used in the quality control of companies without requiring specialized personnel, most of the analyses were carried out outside the biological hood.

In conclusion this project allowed us to validate a rapid method of DNA extraction and to develop rapid PCRs (LAMP) for the identification of white truffles. To develop a commercial kit, however, it will be necessary to validate the test with numerous tests and on processed products such as creams and sauces.

4. Discussion and Conclusion

4.1. DNA based analyses: new frontiers of food traceability

Food authenticity is a major concern for consumers, producers, processors and food authorities as incorrect food labelling and other types of fraudulent practices have been shown to adversely affect confidence in final consumers (Barnett et al., 2016). In the last decade, the use of molecular tools for the authentication of food products has drastically increased (Lo and Shaw et al., 2018). The need for better traceability is mainly due to current legislation and the desire to have better and more uniform control of the entire food chain from the field to the market (Barcaccia et al., 2016). In this context, the biomolecular analysis of DNA barcoding has become more important year after year (Galimberti et al., 2019). Although food authentication using DNA barcoding is well supported and validated when directed to single raw material identification (De Mattia et al., 2011; Hellberg et al., 2017; Galimberti et al., 2014) the characterization of complex food matrices and processed foods is still a challenge. These food matrices are mixtures of different species that form elements such as vegetable soups, sauces and meat mixes and the degradation of DNA due to the industrial processing of food raw materials (e.g. physical and chemical treatments) is a problem to be addressed. To analyse these complex matrices, the DNA barcode approach was combined with High-Throughput DNA Sequencing technologies (HTS) which offered the opportunity to analyse multiple DNA amplicons by sequencing them in parallel (Haynes et al., 2019). The goal of the PhD project was to evaluate different DNA-based analyses to offer companies full product traceability and ensure a safe and high-quality product.

4.2. DNA mini-barcoding: a tool for highly processed products identification

Although authentication using DNA is more reliable than a protein-based method because DNA is more stable during processing (Lo and Shaw, 2017), several studies show that industrial processing treatments can lead to DNA degradation and fragmentation (Lu et al., 2018; Pecoraro et al., 2020). For this reason, DNA barcoding is not suitable to identify highly processed products such as canned food, soup and food supplements. Conversely, DNA mini-barcoding, using a smaller region of DNA compared with the standard ones, may help overcome the limits associated with DNA barcoding (Gao et al., 2019).

The first aim of this thesis project was to identify the typology of products in the botanicals field that is suitable for DNA barcoding and mini-barcoding analysis. We collected a total of 90 samples between raw materials, intermediate products and final products such as food supplements and botanical extracts. We found out that, concerning the raw material samples, our results are consistent with the assumptions of Newmaster and co-workers (2013) who suggested that a DNA barcoding approach could be successfully applied to check the identity of commercial herbal products and to find cases of contamination or substitution. As we expected, the efficacy of DNA extraction and amplification decreased when we analysed the commercial phytoextracts and their intermediates of industrial processing. The phytoextracts obtained through hydroalcoholic treatment, with the lower percentage of ethanol (<40%), and aqueous processing, at the lowest temperature, contained more DNA than the other samples. We generally assumed that the high concentration of ethanol used in the industrial processing steps lead to DNA precipitation. For this reason, both the DNA extraction and DNA barcoding authentication failed when applied to the successive intermediate products of industrial processing and DNA is

no longer available in the final botanicals. In this case, we conclude that a DNA-based approach is not adoptable to achieve a reliable traceability from the initial plant raw material to the final product for this kind of industrial production. In all the tested cases, the DNA mini-barcode locus rbcL1-B was most easily amplified and sequenced than the other two DNA barcoding markers. This suggests that the DNA obtained from phytoextracts are richer in small DNA fragments (80-200 bp). Our data support the ability of DNA mini-barcode makers to provide a reliable tracing of the intermediate products of industrial production. Anyway, it is important to highlight that some industrial processes need high temperatures and the use of solvents (such as a high concentration of ethanol) can induce DNA degradation and make this molecular tool less effective.

4.3. DNA metabarcoding: a pragmatic analysis for supply chain traceability

Many studies showed that DNA barcoding has a high potential for species differentiation in food (Delpiani et al., 2020; Xing et al., 2020; Barbosa et al., 2020; Kane et al., 2016; D'Amato et al., 2013; Hellberg et al., 2017) however this is not suitable for complex multispecies food products. DNA metabarcoding allows to identify multiple genomes in a single sample, and it is currently used for food authentication (Prosser and Hebert, 2017; Utzeri et al, 2018). Haynes and colleagues (Haynes et al., 2019) recommend this approach for improving quality control along the food supply chain. Although huge developments in HTS technology and the use of cheaper benchtop instruments have drastically reduced sequencing costs compared to a few years ago, DNA metabarcoding studies are still scarce to detect food adulteration (Dobrovolny, 2019). For this reason, the second aim of this thesis project was to test the HTS technique on processed food products. Highly processed food products are one of the current challenges of food traceability as food undergoes DNA degradation during industrial processing. Therefore, DNA metabarcoding

can be a valuable tool for tracing food origin and food quality control and can provide new perspectives to identify the composition of processed foods (Bruno et al., 2019).

In order to test the metabarcoding ability of DNA to check all ingredients declared on the label, different types of insect-based novel food products were analysed in this thesis project. On plant components, the ITS2 region was efficient in providing information on species composition, including those taxa that were likely related to the insect breeding substrate (plant diet). In several processed products, we obtained most of the reads belonging to the predicted highly abundant species. DNA tests are related to the quality and quantity of DNA, so highly transformed insect products can be very difficult to analyse. The main limitation of this analysis lies in the sensitivity of detecting less abundant species due to primer bias. This problem has already been discussed previously in the literature by Bista and colleagues (Bista et al., 2018), who stated that whole-genome sequencing without PCR could help avoid this effect. In conclusion, the proposed molecular approach offers a universal diagnostic system to identify the composition of the breeding substrate and to verify compliance with current and future regulations.

In order to improve the identification of ingredients in a food compound and exclude false positives, a coupled multi-marker approach could be useful. For this purpose, we analysed the commercial herbal teas collected by different companies using two markers: *psbA-trnH* and ITS2. The results showed that the *psbA-trnH* marker had a recognition rate of the species declared on the label of 42%, while ITS2 was 49%. By combining the species identified by both markers, the recognition rate reaches 70%. However, this percentage is not indicative of the identification success rate, as we are unsure of the composition of commercial products. The identification success rate of samples created in the laboratory is different. In this case

the success rate of the *psbA-trnH* marker was 70% and of the ITS2 87%, with a total of 100% in the case of using both markers. What was highlighted by this study is that some species were recognized only by one marker and not by the other and vice versa. The peculiarity of this result is that the same species present in different samples had the same behaviour. This may indicate a bias in the PCR, which prefers one species over another. This study shows the ability of HTS technique as a traceability system and the importance of using two markers to increase the species identification.

In recent years, the DNA metabarcoding approach has been proposed as a semi-quantitative tool (Bruno et al., 2019). This application is often limited by both biological and technical biases affecting the number of reads. A bias in the PCR phase can lead to an inaccurate estimation of quantities, this has been partially demonstrated in plants (Elbrecht and Leese, 2015). This bias results in a final mix of amplified DNA that does not always reflect the original proportion of each species, limiting the quantitative potential of DNA metabarcoding (Galimberti et al., 2019). With the aim of evaluating the accuracy of the quantification, samples with different concentrations of five plant species were created in the laboratory, both from raw matrices and from gDNA. The results are consistent with the theory presented by Elbrecht and Leese, showing that the quantification of species is inaccurate. Our results clearly show that the mocks created with genomic DNA are statistically closer to the predicted percentage. This result could be caused by the biases of the extraction phase (including the different starting matrix). Different matrices have a different yield of DNA. Starting samples such as seeds or roots have a lower extraction efficiency than flowers and leaves. This study demonstrated the presence of bias in the extraction phase, due probably to the different DNA yield of the analysed

samples. We showed that this tool does not have a high quantification accuracy, so it is recommended to evaluate only the relative quantification capacity.

In order to test the ability of HTS technique to identify the pathogen and use microbiome as a tool for supply chain traceability, we investigated the microbiome composition of insect-based commercial food products. In this context, High-Throughput DNA Sequencing (HTS) techniques can give insight into the carryover of insect microbiota into final food products. This study aimed to analyse the microbiota variability of different categories of some commercial insect-based food products, including commercial raw materials and processed food products, purchased through e-commerce from different companies.

Overall, our results showed that insect-based food products cluster based on their microbial signature. Even in the case of processed foods in which more than one constituent could interfere with its microbial contribution, we have identified a core microbiota and unique taxa that can be used as biomarkers.

The identification of a microbial signature for traceability purposes has also been suggested by forensic scientists as a natural consequence of the application of HTS technologies (Bishop, 2019): with the globalization of trade, food traceability is a hot topic and identification of a microbial signature in these products can provide a deeper insight into the "food ecosystem"(Parente et al., 2019).

Regarding food safety, some potentially harmful bacteria have been found, but only with viability tests would we be able to define the dangerousness of a product.

In conclusion, the application of HTS has allowed to detect and identify the diversity of microbial communities in some raw and processed food products. The HTS approach is a valuable tool to protect food safety as routine monitoring analysis and characterization of raw materials up to final processed products.

4.4. Globalization and Covid-19 outbreaks: how they had changed the food industry?

People migrate to urban cities and no longer live in places where food is grown. As the physical distance from farm to plate increased, food supply chains have lengthened, and the consumption of processed and packaged foods has increased. Food is produced, processed and distributed to the world population thanks to globalization, the help of technological progress and the availability of transport. For example, unlike in the past when food was difficult to store, modern food is hygienically packaged and lasts longer. This means that people can consume packaged food without any worries regarding the health implications of poor sanitation. In general, the transformation of the food industry and supply chain has greatly facilitated people's lifestyle by allowing them to get the food they want, differently from the past where people had to live within the place farm or go to farmers' markets to acquire food. The transformation of the food industry adapts to changing lifestyles, demographics, rising incomes and education levels, which trigger an increase in consumer demands for quality, variety and food safety. Thus, the diversification of people across different demographic locations has led the food industry to find ways to produce, process and distribute valuable and hygienically packaged foods that meet consumer needs. The food industry has adapted to globalization by making food availability an easy task. However, globalization has created security gaps in the food supply chain and, as a result, recalls are increasing. The technology of food production, storage, transportation and supply chain has evolved to fill these gaps. But the supply chain took some time to identify all its vulnerabilities and to keep up with the tools introduced (Ukessays, 2018). Although the globalization of the food supply chain has improved many aspects of consumers' lives, diseases such as obesity are increasingly being promoted due to fast foods that are cheap and available

but full of unhealthy fats and sugar (Inglis and Gimlin, 2009). However, persistent and widespread hunger and malnutrition remain a huge challenge in many parts of the world. The lengthening of food chains and changes in food patterns necessarily require greater use of resources, energy and emissions. These items threaten the sustainability of food systems and the ability to meet food needs around the world. Although the future implications of climate change on agriculture are difficult to predict, the impacts are imagined to be of varying degrees and different nature in each ecological zone, region and production system (FAO, 2017).

The transport and stretched supply chains have been one of the causes of the rapid spread of Covid-19 around the world. In this context, Covid-19 had a direct impact on access to food. These impacts were also due to interruptions in availability, shifts in consumer demand towards cheaper and less nutritious foods, and instability in food prices. The pandemic is affecting all four pillars of food security: Is the supply of food adequate (Availability)? Can people access to the food they need (Access)? Can people obtain all the intake of nutrients that they need (Stability)? And finally, can people access to food whenever they want (Utilization)? (Laborde et al., 2020).

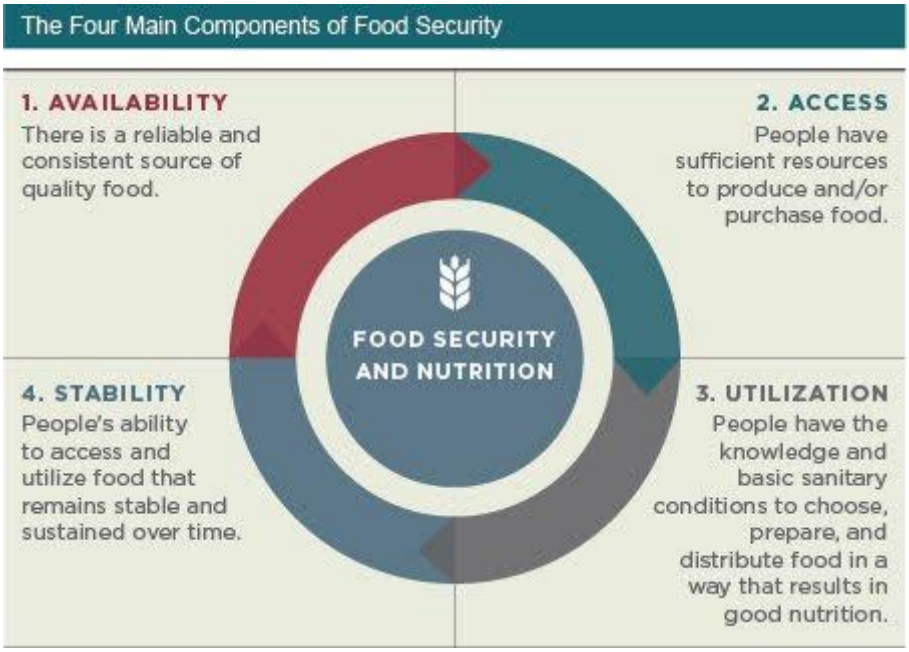


Figure 18. The four main components of Food Security: Availability, Access, Stability and Utilization (Laborde et al., 2020).

Despite this, the main threats of Covid-19 refer to food security and it is important that this global health crisis does not also become a global food crisis (Laborde et al, 2020). Currently, global poverty and food insecurity due to the pandemic have been estimated at 90-150 million people, who could fall into poverty by increasing global poverty by 15-24%, especially in sub-Saharan Africa and South Africa. This factor has a large impact on food security and nutrition, because people in poverty do not have enough money to buy food with the right amount of nutrient, preferring cheaper and low-quality food. This factor will have a major impact on children that would tend to be affected by undernutrition (Gerszon Mahler et al., 2020).

The Covid-19 pandemic crisis has affected all types of food supply chains, starting from fresh vegetables and fruit, baked goods, cereals to perishable products (Ivanov and Dolgui 2020). Much of the logistical supply activities have been

disrupted, and therefore food shortages were inevitable. In addition, some areas declared red or containment areas due to the high infection were another issue of facility failure. Due to labour shortages and the uncertainty of vehicle availability, difficulties have been developed in achieving the desired purpose of the supply chain (Singh et al., 2020).

Consumers reacted by stockpiling products and anticipated food shortages (Hall et al., 2020). They have drastically changed their buying behaviour towards online delivery, hampering the supply chain's ability to satisfy the requests (Dunkley, 2020; Smith, 2020).

Poverty is increasing and therefore more and more consumers are struggling to afford food, due to the sudden loss of jobs and the shift in demand towards food pantries (Charles, 2020). Within the supply chain, sudden changes in health demand and regulations have caused profound disruptions such as the unavailability of agricultural workers, the collapse of the restaurant sector and the change in working conditions (Cagle, 2020; Corkery and Yaffe-Bellany, 2020).

The Covid-19 crisis has provided an opportunity for transformation in the food supply chain. Much of what is known about the food crisis comes from media reports and has been divided into four themes (see figure 19): changing consumer behaviours, labour mismatch, changing health and safety behaviours and finally demand and supply mismatch.

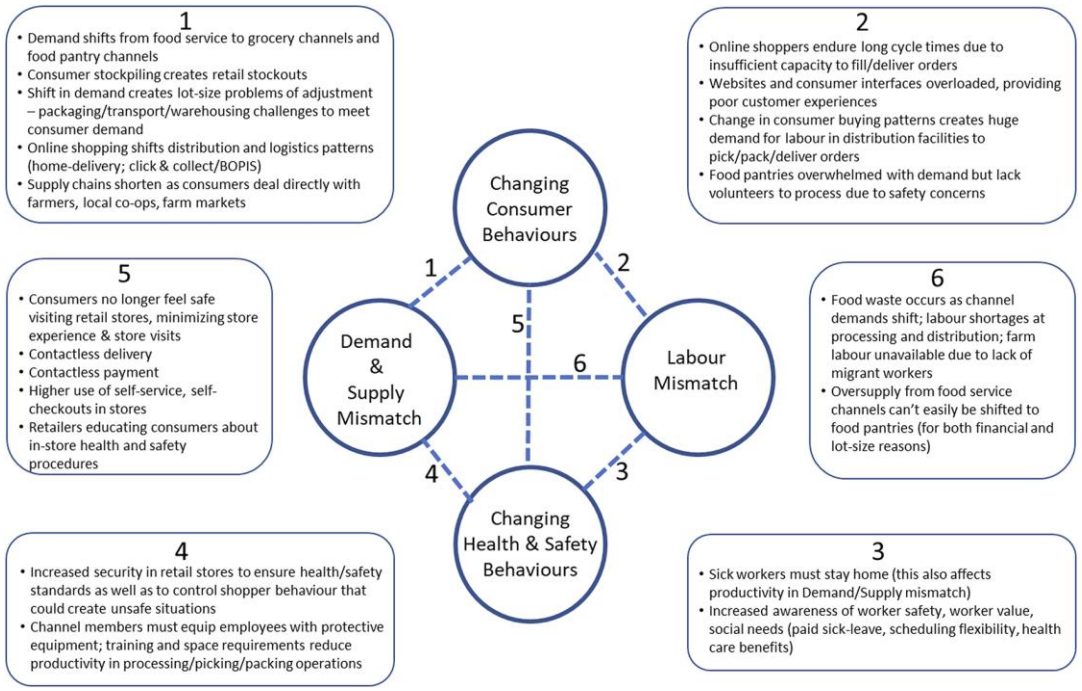


Figure 19. Supply chain themes identified during the coronavirus disease of 2019 pandemic (Diane et al., 2020).

The new normal of food supply chains has yet to be achieved, but it can be imagined by evaluating these key issues in relation to the current situation (Mollenkopf et al., 2020).

4.5. New frontiers in supply chain control

As the authenticity and safety of food products are increasing attention, food industries and control laboratories must use all available techniques to verify products and control suppliers. Food safety and authenticity regulations are increasingly stringent, so the importance of routine control techniques are becoming more and more important.

HTS is a candidate to become the technology of choice for food analysis and food fraud protection, particularly for highly complex food matrices.

The costs of HTS platforms are decreasing, allowing public laboratories to use these technologies. However, HTS currently requires highly specialized equipment and skilled personnel to analyse data, and future developments are likely to aim at ease of application and data extraction, allowing a widespread application for less specialized personnel.

In the next future, as HTS will become an established technique, both methodologies and analysis pipelines should be harmonized across testing laboratories (Endrullat et al., 2016, Gargis et al., 2016). Technological advances and increased competition will continue to push the field towards lower costs (Goodwin et al., 2016), higher productivity and more user-friendly analysis options (Haynes et al., 2019). HTS analysis has the weakness of being a long analysis (sequencing service requires 1-2 months to provide the results) and these times are not compliant to the quality control required by companies. In this context, Nanopore MinION could be successful in the next years. This tool allows to sequence multiple genomes in a few hours and does not require large machinery, but simply a tool as large as a USB stick. Nanopore MinION has already been tested in extreme environments such as deserts, glaciers and even space (Edwards et al., 2019; Castro-Wallace et al., 2016). Although the scientific literature has not explored much of the application of Nanopore MinION in the food field, it could be the future of safety and quality control. Voorhuijzen-Harink and colleagues tested the Nanopore MinION on fish analysis and assessed that the MinION sequencing is as accurate as the Illumina MiSeq platform. In conclusion, the high accuracy, practicability, low cost and time-safer of the

MinION sequencer make MinION-based DNA metabarcoding promising for fast analysis of complex food and feed mixtures. (Voorhuijzen-Harink et al., 2019).

In addition to this tool, cost-effective and rapid technologies have been developing in recent years that can speed up the quality control of the supply chain. LAMP technology is currently used in food safety quality control, to identify, for example, the presence of pathogens such as *Salmonella sp.* (Vichaibun and Kanchanaphum, 2020). Vichaibun and Kanchanaphum also tested the quantitative capacity of LAMP technology (qLAMP) evaluating that both LAMP and qLAMP are more dominants than PCR and qPCR in many aspects. They have more specificity, they are more rapid, they have more sensitivity and finally they are less complex. This methodology has increasingly broad applications in the food field, since it is used both to detect allergens (Allgöwer et al., 2020; Guevara et al., 2016), bacterial contamination (Wang et al., 2017) but also for quality control (Mao et al., 2020; Li et al., 2017). The FDA has also developed the LAMP technique to identify the presence of *Salmonella sp*. in pet food (Yang et al., 2018). In the future, new platforms for food detection will be more and more common. Some companies are already developing and selling portable kits for food safety based on the LAMP technique (Hu et al., 2017).

To conclude, quality and safety control of products and supply chain traceability are necessary and required by both companies and consumers. The main objective is to obtain simple, fast and efficient technologies that require less specialized personnel. Although it is difficoult to think a complete replacement of analysis laboratories, the trend of the future is to develop targeted and more immediate kits and analyses, perhaps making it possible to analyse food even by the consumer.

5. Other publications

The problem of misidentification between edible and poisonous wild plants: Reports from the Mediterranean area

Laura Cornara[a], Antonella Smeriglio[b], **Jessica Frigerio**[c] , Massimo Labra[d] , Emilio Di Gristina[e] , Marcella Denaro[b], E.Mora[a], Domenico Trombetta[b]

[a] Department of Earth, Environment and Life Sciences, University of Genova, Italy

[b] Department of Chemical, Biological, Pharmaceutical and Environmental Sciences, University of Messina, Italy

[c] FEM2 Ambiente Srl, Università di Milano-Bicocca, Italy

[d] Department of Biotechnology and Bioscience, University of Milano-Bicocca, Italy

[e] Department of Biological, Chemical and Pharmaceutical Sciences and Technologies (STEBICEF), Section of Botany and Plant Ecology, University of Palermo, Via Archirafi 38, 90123 Palermo, Italy

Food Chem Toxicol. 2018 Sep;119:112-121. doi: 10.1016/j.fct.2018.04.066.

Development of a DNA Barcoding-Like Approach to Detect Mustard Allergens in Wheat Flours

Jessica Frigerio [1], Roberta Pellesi [3], Valerio Mezzasalma [1], Fabrizio De Mattia [1], Andrea Galimberti [2], Francesca Lambertini [3], Michele Suman [3], Sandro Zanardi [3], Andrea Leporati [3] and Massimo Labra [2,*]

[1] FEM2-Ambiente, Piazza della Scienza 2, I-20126 Milano (Italy); jessica.frigerio@fem2ambiente.com, valerio.mezzasalma@fem2ambiente.com, fabrizio.demattia@fem2ambiente.com.

[2] Zooplantlab, Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, I-20126 Milano (Italy); andrea.galimberti@unimib.it, massimo.labra@unimib.it.

[3] BARILLA G. e R. FRATELLI Spa, Via Mantova, 166 - 43122 Parma (Italy); roberta.pellesi@barilla.com, francesca.lambertini@barilla.com, michele.suman@barilla.com, sandro.zanardi@barilla.com, andrea.leporati@barilla.com.

***** Correspondence: massimo.labra@unimib.it; Tel.: +39-02-6448-3472

Genes 2019, 10(3), 234; https://doi.org/10.3390/genes10030234

Food Tracking Perspective: DNA Metabarcoding to Identify Plant Composition in Complex and Processed Food Products

Antonia Bruno 1, Anna Sandionigi 1*, Giulia Agostinetto 1, Lorenzo Bernabovi 1,†, **Jessica Frigerio** 2, Maurizio Casiraghi 1 and Massimo Labra

1 Zooplantlab, Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, I-20126 Milano, Italy

2 FEM2-Ambiente, Piazza della Scienza 2, I-20126 Milano, Italy

*Author to whom correspondence should be addressed.

† This paper is dedicated to the memory of our wonderful colleague, Dr. Lorenzo Bernabovi, who recently passed away.

Genes (Basel). 2019 Mar; 10(3): 248 doi: 10.3390/genes10030248

Comparative and Functional Screening of Three Species Traditionally used as Antidepressants: *Valeriana officinalis* L., *Valeriana jatamansi Jones* ex Roxb. and *Nardostachys jatamansi* (D.Don) DC.

by Laura Cornara 1, Gabriele Ambu 1, Domenico Trombetta 2, Marcella Denaro 2, Susanna Alloisio 3,4, **Jessica Frigerio** 5, Massimo Labra 6, Govinda Ghimire 7, Marco Valussi 8 and Antonella Smeriglio 2*

1 Department of Earth, Environment and Life Sciences, University of Genova, 16132 Genova, Italy

2 Department of Chemical, Biological, Pharmaceutical and Environmental Sciences, University of Messina, Via Giovanni Palatucci, 98168 Messina, Italy

3 ETT Spa, via Sestri 37, 16154 Genova, Italy

4 Institute of Biophysics-CNR, 16149 Genova, Italy

5 FEM2 Ambiente Srl, Piazza della Scienza 2, 20126 Milan, Italy

6 Department of Biotechnology and Bioscience, University of Milano-Bicocca, Piazza della Scienza 2, 20126 Milan, Italy

7 Nepal Herbs and Herbal Products Association, Kathmandu 44600, Nepal

8 European Herbal and Traditional Medicine Practitioners Association (EHTPA), Norwich 13815, UK

*Author to whom correspondence should be addressed.

Plants 2020, 9(8), 994; https://doi.org/10.3390/plants9080994

Application of DNA mini-barcoding and infrared spectroscopy for the authentication of the Italian product "Bottarga"

Jessica Frigerio a,b, Claudio Marchesi c, Chiara Magoni b, Francesco Saliu d*, Davide Ballabio d, Viviana Consonni d, Tommaso Gorini a, Fabrizio De Mattia a, Paolo Galli d, Massimo Labra b

aFEM2-Ambiente, Piazza della Scienza 2, I-20126 Milano, Italy

b Biotechnology and Bioscience Department, University of Milano-Bicocca, Piazza della Scienza 2, 20126 Milano - Italy

c Department of Mechanical and Industrial Engineering, University of Brescia, Via Branze, 38 - 25123 Brescia

d Department of Earth and Environmental Science, University of Milano-Bicocca, Piazza della Scienza 1, 20126 Milano - Italy

Corresponding author * Tel +39-0264482813.

E-mail address: francesco.saliu@unimib.it (F. Saliu).

https://doi.org/10.1016/j.lwt.2020.110603

6. References

Abdelmajeed, N. A., Danial, E. N., & Ayad, H. S. (2013). The effect of environmental stress on qualitative and quantitative essential oil of aromatic and medicinal plants. Archives Des Sciences, 66(4), 100-120. Retrieved online on https://www.researchgate.net/publication/258959867_The_effect_of_environmental_stress_on_qualitative_and_quantitative_essential_oil_of_aromatic_and_medicinal_plants

Alhassan, A., Li, Z., Poole, C. B., & Carlow, C. K. (2015). Expanding the MDx toolbox for filarial diagnosis and surveillance. Trends in parasitology, 31(8), 391-400. https://doi.org/10.1016/j.pt.2015.04.006

Barbuto, M., Galimberti, A., Ferri, E., Labra, M., Malandra, R., Galli, P., & Casiraghi, M. (2010). DNA barcoding reveals fraudulent substitutions in shark seafood products: The Italian case of "palombo" (Mustelus spp.). Food Research International, 43(1), 376–381. https://doi.org/10.1016/j.foodres.2009.10.009

Barcaccia, G., Lucchin, M., & Cassandro, M. (2015). DNA Barcoding as a Molecular Tool to Track Down Mislabeling and Food Piracy. Diversity, 8(4), 2. https://doi.org/10.3390/d8010002

Barrere, V., Everstine, K., Théolier, J., & Godefroy, S. Food fraud vulnerability assessment: Towards a global consensus on procedures to man-age and mitigate food fraud. Trends in Food Science and Technology, Vol. 100, pp. 131–137. https://doi.org/10.1016/j.tifs.2020.04.002

Behjati, S., & Tarpey, P. S. (2013). What is next generation sequencing? Archives of Disease in Childhood: Education and Practice Edition, 98(6), 236–238. https://doi.org/10.1136/archdischild-2013-304340

Bertolini, F., Ghionda, M. C., D'Alessandro, E., Geraci, C., Chiofalo, V., & Fontanesi, L. (2015). A Next Generation Semiconductor Based Sequencing Approach for the Identification of Meat Species in DNA Mixtures. PLOS ONE, 10(4), e0121701. https://doi.org/10.1371/journal.pone.0121701

Bhardwaj, S., & Kaushik, M. (2018). Blockchain—technology to drive the future. Smart Innovation, Systems and Technologies, 78, 263–271. https://doi.org/10.1007/978-981-10-5547-8_28

Bishop, A. H. (2019). The signatures of microorganisms and of human and environmental biomes can now be used to provide evidence in legal cases. FEMS Mi-crobiology Letters, 366(3). https://doi.org/10.1093/femsle/fnz021

Bruni, I., Galimberti, A., Caridi, L., Scaccabarozzi, D., De Mattia, F., Casiraghi, M., & Labra, M. (2015). A DNA barcoding approach to identify plant species in multiflower honey. Food Chemistry. https://doi.org/10.1016/j.food-chem.2014.08.060

Calegari, B. B., Avila, E. F., Reis, R. E., & Alho, C. S. (2020). DNA barcode authentica-tion reveals highly fraudulent Cod commerce in Porto Alegre, Brazil. Forensic Science International: Reports. https://doi.org/10.1016/j.fsir.2020.100072

Caruso, M., Federici, C. T., & Roose, M. L. (2008). EST-SSR markers for asparagus genetic diversity evaluation and cultivar identification. Molecular Breeding, 21(2), 195–204. https://doi.org/10.1007/s11032-007-9120-z

Casiraghi, M., Labra, M., Ferri, E., Galimberti, A., & De Mattia, F. (2010). DNA barcoding: a six-question tour to improve users' awareness about the method. Briefings in bioinformatics, 11(4), 440-453. https://doi.org/10.1093/bib/bbq003

Castro-Wallace, S. L., Chiu, C. Y., John, K. K., Stahl, S. E., Rubins, K. H., McIntyre, A. B. R., … Burton, A. S. (2017). Nanopore DNA Sequencing and Genome Assembly on the International Space Station. Scientific Reports. https://doi.org/10.1038/s41598-017-18364-0

Laborde, D., Martin, W., Swinnen, J., & Vos, R. (2020). COVID-19 risks to global food security. Science, 369(6503), 500-502. https://doi.org/ 10.1126/science.abc4765

Cauhan, K., Nayeem, M., Jawed, I., & Tareen, F. R. (2020). The Pharma Innovation Journal 2020; 9(5): 01-04 Food safety and COVID-19: Limitations of HACCP and the way forward. https://doi.org/10.22271/tpi.2020.v9.i5a.4616

Charlebois, S., Schwab, A., Henn, R., & Huck, C. W. (2016). Food fraud: An exploratory study for measuring consumer perception towards mislabeled food products and influence on self-authentication intentions. Trends in Food Science and Technology, Vol. 50, pp. 211–218. https://doi.org/10.1016/j.tifs.2016.02.003

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., … Leon, C. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS ONE. https://doi.org/10.1371/journal.pone.0008613

Mollenkopf, D. A., Ozanne, L. K., & Stolze, H. J. (2020). A transformative supply chain response to COVID-19. Journal of Service Management. https://doi/10.1108/JOSM-05-2020-0143

Cornara, L., Smeriglio, A., Frigerio, J., Labra, M., Di Gristina, E., Denaro, M., … Trombetta, D. (2018). The problem of misidentification between edible and poisonous wild plants: Reports from the Mediterranean area. Food and Chemical Toxicology. https://doi.org/10.1016/j.fct.2018.04.066

Cultivar Identification in Olive Based on RAPD Markers in: Journal of the American Society for Horticultural Science Volume 126 Issue 6 (2001). Retrieved November 30, 2020, from https://journals.ashs.org/jashs/view/journals/jashs/126/6/article-p668.xml

Dai, T., Hu, T., Yang, X., Shen, D., Jiao, B., Tian, W., & Xu, Y. (2019). A recombinase polymerase amplificationlateral flow dipstick assay for rapid detection of the quarantine citrus pathogen in China, Phytophthora hibernalis. PeerJ, 2019(11), e8083. https://doi.org/10.7717/peerj.8083

Zhao, M., Wang, B., Xiang, L., Xiong, C., Shi, Y., Wu, L., ... & Sun, W. (2019). A novel onsite and visual molecular technique to authenticate saffron (Crocus

sativus) and its adulterants based on recombinase polymerase amplification. Food Control, 100, 117-121. https://doi.org/10.1016/j.foodcont.2019.01.011

De Battisti, C., Marciano, S., Magnabosco, C., Busato, S., Arcangeli, G., & Cattoli, G. (2014). Pyrosequencing as a tool for rapid fish species identification and commercial fraud detection. Journal of Agricultural and Food Chemistry, 62(1), 198–205. https://doi.org/10.1021/jf403545m

De Castro, O., Comparone, M., Di Maio, A., Del Guacchio, E., Menale, B., Troisi, J., … Guida, M. (2017). What is in your cup of tea? DNA Verity Test to characterize black and green commercial teas. PLOS ONE, 12(5), e0178262. https://doi.org/10.1371/journal.pone.0178262

De Mattia, F., Bruni, I., Galimberti, A., Cattaneo, F., Casiraghi, M., & Labra, M. (2011). A comparative study of different DNA barcoding markers for the identification of some members of Lamiacaea. Food Research International, 44(3), 693–702. https://doi.org/10.1016/j.foodres.2010.12.032

Deconinck, D., Volckaert, F. A. M., Hostens, K., Panicz, R., Eljasik, P., Faria, M., … Derycke, S. (2020). A high-quality genetic reference database for European commercial fishes reveals substitution fraud of processed Atlantic cod (Gadus morhua) and common sole (Solea solea) at different steps in the Belgian supply chain. Food and Chemical Toxicology. https://doi.org/10.1016/j.fct.2020.111417

Derz, W., Pavlovic, M., Huber, I., Schalch, B., & Gerdes, L. (2020). Food fraud in the Alps? — Detection of chamois (Rupicapra rupicapra) in firm raw sausages, ham, and meat via qualitative duplex real-time PCR. Food Control, 107764. https://doi.org/10.1016/j.foodcont.2020.107764

Böhme, K., Calo-Mata, P., Barros-Velázquez, J., & Ortea, I. (2019). Review of recent DNA-based methods for main food-authentication topics. Journal of agricultural and food chemistry, 67(14), 3854-3864. https://doi.org/10.1021/acs.jafc.8b07016

Di Pinto, A., Bottaro, M., Bonerba, E., Bozzo, G., Ceci, E., Marchetti, P., … Tantillo, G. (2015). Occurrence of mislabeling in meat products using DNA-based assay. Journal of Food Science and Technology, 52(4), 2479–2484. https://doi.org/10.1007/s13197-014-1552-y

Dobrovolny, S., Blaschitz, M., Weinmaier, T., Pechatschek, J., Cichna-Markl, M., Indra, A., … Hochegger, R. (2019). Development of a DNA metabarcoding method for the identification of fifteen mammalian and six poultry species in food. Food Chemistry, 272, 354–361. https://doi.org/10.1016/j.foodchem.2018.08.032

Druml, B., & Cichna-Markl, M. (2014). High resolution melting (HRM) analysis of DNA - Its role and potential in food analysis. Food Chemistry, Vol. 158, pp. 245–254. https://doi.org/10.1016/j.foodchem.2014.02.111

Druml, B., Mayer, W., Cichna-Markl, M., & Hochegger, R. (2015). Development and validation of a TaqMan real-time PCR assay for the identification and quantifi-cation of roe deer (Capreolus capreolus) in food to detect food adulteration. Food Chemistry, 178, 319–326. https://doi.org/10.1016/j.foodchem.2015.01.003

Duarte-Guevara, C., Swaminathan, V. V., Reddy, B., Huang, J. C., Liu, Y. S., & Bashir, R. (2016). On-chip electrical detection of parallel loop-mediated isothermal amplification with DG-BioFETs for the detection of foodborne bacterial path-ogens. RSC Advances. https://doi.org/10.1039/c6ra19685c

Elbrecht, V., & Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. PLoS ONE. https://doi.org/10.1371/journal.pone.0130324

Morin and Lees, 2018. FI Handbook on Food Authenticity Issues and Related Analytical Techniques. https://doi.org/10.32741/fihb

Ellis, D. I., Muhamadali, H., Haughey, S. A., Elliott, C. T., & Goodacre, R. (2015). Point-and-shoot: Rapid quantitative detection methods for on-site food fraud

analysis-moving out of the laboratory and into the food supply chain. Analytical Methods, Vol. 7, pp. 9401–9414. https://doi.org/10.1039/c5ay02048d

Everstine, K., Spink, J., & Kennedy, S. (2013). Economically motivated adulteration (EMA) of food: Common characteristics of EMA incidents. Journal of Food Protection, 76(4), 723–735. https://doi.org/10.4315/0362-028X.JFP-12-399

Facing up to food fraud in a pandemic. (2020). Food Science and Technology. https://doi.org/10.1002/fsat.3403_4.x

Fao. (n.d.). The future of food and agriculture: Trends and challenges. Retrieved from www.fao.org/publications

Food Fraud and the Food, Drug, and Cosmetic Act: Bridging a Disconnect - Food and Drug Law Institute (FDLI). (n.d.). Retrieved November 30, 2020, from https://www.fdli.org/2019/07/food-fraud-and-the-food-drug-and-cosmetic-act-bridging-a-disconnect/

Galal-Khallaf, A., Osman, A. G. M., Carleos, C. E., Garcia-Vazquez, E., & Borrell, Y. J. (2016). A case study for assessing fish traceability in Egyptian aquafeed formulations using pyrosequencing and metabarcoding. Fisheries Research. https://doi.org/10.1016/j.fishres.2015.09.009

Galimberti, A., Labra, M., Sandionigi, A., Bruno, A., Mezzasalma, V., & De Mattia, F. (2014). DNA Barcoding for Minor Crops and Food Traceability. Advances in Agriculture. https://doi.org/10.1155/2014/831875

Barcaccia, G., Lucchin, M., & Cassandro, M. (2015). DNA Barcoding as a Molecular Tool to Track Down Mislabeling and Food Piracy. Diversity, 8(4), 2. https://doi.org/10.3390/d8010002

Barnett, J., Begen, F., Howes, S., Regan, A., McConnon, A., Marcu, A., et al. (2016). Consumers' confidence, reflections and response strategies following the horsemeat incident. Food Control, 59, 721–730. https://doi.org/10.1016/j.foodcont.2015.06. 021

Galimberti, A., Casiraghi, M., Bruni, I., Guzzetti, L., Cortis, P., Berterame, N. M., & Labra, M. (2019). From DNA barcoding to personalized nutrition: the evolution of food traceability. Current Opinion in Food Science, 28, 41-48. https://doi.org/10.1016/j.cofs.2019.07.008

Bruno, A.; Sandionigi, A.; Agostinetto, G.; Bernabovi, L.; Frigerio, J.; Casiraghi, M.; Labra, M. Food Tracking Perspective: DNA Metabarcoding to Identify Plant Composition in Complex and Processed Food Products. Genes 2019, 10, 248. https://doi.org/10.3390/genes10030248

Bista, I.; Carvalho, G.R.; Tang, M.;Walsh, K.; Zhou, X.; Hajibabaei, M.; Shokralla, S.; Seymour, M.; Bradley, D.; Liu, S.; et al. (2018). Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples. Mol. Ecol. Resour. 18, 1020–1034. https://doi.org/10.1111/1755-0998.12888

Galvez, J. F., Mejuto, J. C., & Simal-Gandara, J. (2018). Future challenges on the use of blockchain for food traceability analysis. TrAC - Trends in Analytical Chemistry, Vol. 107, pp. 222–232. https://doi.org/10.1016/j.trac.2018.08.011

Gao, Z., Liu, Y., Wang, X., Wei, X., & Han, J. (2019). DNA Mini-Barcoding: A Derived Barcoding Method for Herbal Molecular Identification. Frontiers in Plant Science, Vol. 10, p. 987. https://doi.org/10.3389/fpls.2019.00987

Gee, J., & Button, M. (n.d.). The latest data from around the world.

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. Nature Reviews Genetics. https://doi.org/10.1038/nrg.2016.49

Guntzburger, Y., Théolier, J., Barrere, V., Peignier, I., Godefroy, S., & de Marcellis-Warin, N. (2020). Food industry perceptions and actions towards food fraud: Insights from a pan-Canadian study. Food Control, 113, 107182. https://doi.org/10.1016/j.foodcont.2020.107182

Hall, M. C., Prayag, G., Fieger, P., & Dyason, D. (2020). Beyond panic buying: consumption displacement and COVID-19. Journal of Service Management. https://doi.org/10.1108/JOSM-05-2020-0151

Haynes, E., Jimenez, E., Pardo, M. A., & Helyar, S. J. (2019). The future of NGS (Next Generation Sequencing) analysis in testing food authenticity. Food Control, Vol. 101, pp. 134–143. https://doi.org/10.1016/j.foodcont.2019.02.010

Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences. https://doi.org/10.1098/rspb.2002.2218

Hegewald, H., Wensch-Dorendorf, M., Sieling, K., & Christen, O. (2018). Impacts of break crops and crop rotations on oilseed rape productivity: A review. European Journal of Agronomy, 101, 63–77. https://doi.org/10.1016/j.eja.2018.08.003

European Parliament, 2014. Food crisis, fraud in the food chain and the control thereof European Parliament resolution of 14 January 2014 on the food crisis, fraud in the food chain and the control thereof (2013/2091(INI)). Retrieved online (11/02/2020) on https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52014IP0011(01)&from=EN.

Delpiani, G., Delpiani, S. M., Antoni, M. D., Ale, M. C., Fischer, L., Lucifora, L. O., & de Astarloa, J. D. (2020). Are we sure we eat what we buy? Fish mislabelling in Buenos Aires province, the largest sea food market in Argentina. Fisheries Research, 221, 105373. https://doi.org/10.1016/j.fishres.2019.105373

D'Amato, M. E., Alechine, E., Cloete, K. W., Davison, S., & Corach, D. (2013). Where is the game? Wild meat products authentication in South Africa: a case study. Investigative genetics, 4(1). https://doi.org/1-13. 10.1186/2041-2223-4-6

Bertolini, F., Ghionda, M. C., D'Alessandro, E., Geraci, C., Chiofalo, V., & Fontanesi, L. (2015). A next generation semiconductor based sequencing approach for

the identification of meat species in DNA mixtures. PloS one, 10(4), e0121701. https://doi.org/10.1371/journal.pone.0121701

Barbosa, A. J., Sampaio, I., da Silva, E. M., Alcântara, J. V. L., & Santos, S. (2020). Molecular authentication by DNA barcoding and multiplex PCR assay reveals mislabeling and commercial fraud of the Acoupa weakfish (Cynoscion acoupa), an economically important sciaenid marketed in Brazil. Food Control, 117, 107351. https://doi.org/10.1016/j.foodcont.2020.107351

Hellberg, R. S., Hernandez, B. C., & Hernandez, E. L. (2017). Identification of meat and poultry species in food products using DNA barcoding. Food Control. https://doi.org/10.1016/j.foodcont.2017.04.025

Hong, E., Lee, S. Y., Jeong, J. Y., Park, J. M., Kim, B. H., Kwon, K., & Chun, H. S. (2017). Modern analytical methods for the detection of food fraud and adulteration by food category. Journal of the Science of Food and Agriculture, Vol. 97, pp. 3877–3896. https://doi.org/10.1002/jsfa.8364

Hu, L., Ma, L. M., Zheng, S., He, X., Wang, H., Brown, E. W., … Zhang, G. (2017). Evaluation of 3M Molecular Detection System and ANSR Pathogen Detection System for rapid detection of Salmonella from egg products. Poultry Science. https://doi.org/10.3382/ps/pew399

Inglis, D., & Gimlin, D. (2009). The Globalization of Food. Retrieved from https://abdn.pure.elsevier.com/en/publications/the-globalization-of-food

Ivanov, D., & Dolgui, A. (2020). Viability of intertwined supply networks: extending the supply chain resilience angles towards survivability. A position paper motivated by COVID-19 outbreak. International Journal of Production Research. https://doi.org/10.1080/00207543.2020.1750727

Johnson, R. (2014). Food Fraud and “Economically Motivated Adulteration” of Food and Food Ingredients. Retrieved from www.crs.govR43358

Kamath, R. (2018). Food Traceability on Blockchain: Walmart's Pork and Mango Pilots with IBM. The Journal of the British Blockchain Association, 1(1), 1–12. https://doi.org/10.31585/jbba-1-1-(10)2018

Kane, D. E., & Hellberg, R. S. (2016). Identification of species in ground meat products sold on the U.S. commercial market using DNA-based methods. Food Control. https://doi.org/10.1016/j.foodcont.2015.05.020

Kappel, K., Haase, I., Käppel, C., Sotelo, C. G., & Schröder, U. (2017). Species identification in mixed tuna samples with next-generation sequencing targeting two short cytochrome b gene fragments. Food Chemistry, 234, 212–219. https://doi.org/10.1016/j.foodchem.2017.04.178

Khan, I. A., & Smillie, T. (2012). Implementing a "quality by design" approach to as-sure the safety and integrity of botanical dietary supplements. Journal of Na-tural Products. https://doi.org/10.1021/np300434j

Koh, J. C. O., Barbulescu, D. M., Norton, S., Redden, B., Salisbury, P. A., Kaur, S., … Slater, A. T. (2017). A multiplex PCR for rapid identification of Brassica species in the triangle of U. Plant Methods, 13(1), 49. https://doi.org/10.1186/s13007-017-0200-8

Kumar, S., Kahlon, T., & Chaudhary, S. (2011). A rapid screening for adulterants in olive oil using DNA barcodes. Food Chemistry. https://doi.org/10.1016/j.foodchem.2011.01.094

Li, P., Niu, W., Li, H., Lei, H., Liu, W., Zhao, X., … Bai, C. (2015). Rapid detection of Acinetobacter baumannii and molecular epidemiology of carbapenem-resistant A. baumannii in two comprehensive hospitals of Beijing, China. Frontiers in Microbiology, 6(SEP), 997. https://doi.org/10.3389/fmicb.2015.00997

Li, Y., Fan, P., Zhou, S., & Zhang, L. (2017). Loop-mediated isothermal amplification (LAMP): a novel rapid detection platform for pathogens. Microbial pathogenesis, 107, 54-61. https://doi.org/10.1016/j.micpath.2017.03.016

Liou, P., Banda, A., Isaacs, R. B., & Hellberg, R. S. (2020). Labeling compliance and species authentication of fish fillets sold at grocery stores in Southern California. Food Control, 112, 107137. https://doi.org/10.1016/j.foodcont.2020.107137

Liu, Y., Wang, X. Y., Wei, X. M., Gao, Z. T., & Han, J. P. (2018). Rapid Authentication of Ginkgo biloba Herbal Products Using the Recombinase Polymerase Amplification Assay. Scientific Reports, 8(1). https://doi.org/10.1038/s41598-018-26402-8

Lo, Y. T., & Shaw, P. C. (2018). DNA-based techniques for authentication of processed food and food supplements. Food Chemistry, Vol. 240, pp. 767–774. https://doi.org/10.1016/j.foodchem.2017.08.022

Gargis, A. S., Kalman, L., Berry, M. W., Bick, D. P., Dimmock, D. P., Hambuch, T., et al. (2012). Assuring the quality of next-generation sequencing in clinical laboratory practice. Nature Biotechnology, 30(11), https://doi.org/10.1038/nbt.2403 doi:10. 1038/nbt.2403

Vichaibun, V., & Kanchanaphum, P. (2020). Quantitative LAMP and PCR Detection of Salmonella in Chicken Samples Collected from Local Markets around Pathum Thani Province, Thailand. International Journal of Food Science. https://doi.org/10.1155/2020/8833173

Allgöwer, S. M., Hartmann, C. A., & Holzhauser, T. (2020). The development of highly specific and sensitive primers for the detection of potentially allergenic soybean (Glycine max) using loop-mediated isothermal amplification combined with lateral flow dipstick (LAMP-LFD). Foods. https://doi.org/10.3390/foods9040423

Edwards, A., Debbonaire, A., Nicholls, S., Rassner, S., Sattler, B., Cook, J., … Hodson, A. (2016). In-field metagenome and 16S rRNA gene amplicon nanopore sequencing robustly characterize glacier microbiota. BioRxiv. https://doi.org/10.1101/073965

Endrullat, C., Glökler, J., Franke, P., & Frohme, M. (2016). Standardization and quality management in next-generation sequencing. Applied and Translational Genomics. https://doi.org/10.1016/j.atg.2016.06.001

Lobato, I. M., & O'Sullivan, C. K. (2018). Recombinase polymerase amplification: Basics, applications and recent advances. TrAC - Trends in Analytical Chemistry. https://doi.org/10.1016/j.trac.2017.10.015

Lu, Y., Shen, Y., Warren, W., & Walter, R. (2016). Next Generation Sequencing in Aquatic Models. In Next Generation Sequencing - Advances, Applications and Challenges. https://doi.org/10.5772/61657

Shen Y. Y., Chen X., Murphy R. W. (2013). Assessing DNA barcoding as a tool for species identification and data quality control. PLoS ONE 8:e57125. 10.1371/journal.pone.0057125

Lu, Z., Rubinsky, M., Babajanian, S., Zhang, Y., Chang, P., & Swanson, G. (2018). Visualization of DNA in highly processed botanical materials. Food Chemistry, 245, 1042–1051. https://doi.org/10.1016/j.foodchem.2017.11.067

Machado, T. B., Ricciardi, L., & Beatriz P P Oliveira, M. (2020). Blockchain technology for the management of food sciences researches. Trends in Food Science and Technology, Vol. 102, pp. 261–270. https://doi.org/10.1016/j.tifs.2020.03.043

Manning, L. (2016). Food fraud: policy and food chain. Current Opinion in Food Science, Vol. 10, pp. 16–21. https://doi.org/10.1016/j.cofs.2016.07.001

Manning, L., & Soon, J. M. (2016). Food Safety, Food Fraud, and Food Defense: A Fast Evolving Literature. Journal of Food Science, 81(4), R823–R834. https://doi.org/10.1111/1750-3841.13256

Manning, L., & Soon, J. M. (2019). Food fraud vulnerability assessment: Reliable data sources and effective assessment approaches. Trends in Food Science and Technology, Vol. 91, pp. 159–168. https://doi.org/10.1016/j.tifs.2019.07.007

Points and Manning, 2020. Facing up to food fraud in a pandemic. Institute of Food Science and Technology. https://doi.org/10.1002/fsat.3403_4.x

Mao, R., Xie, K., Zhao, M. et al. (2020). Development and Evaluation of a Loop-Mediated Isothermal Amplification (LAMP) Assay for Rapid Detection of Pistachio (Pis-tacia vera) in Food Samples. Food Anal. Methods 13, 658–666. https://doi.org/10.1007/s12161-019-

Mezzasalma, V., Ganopoulos, I., Galimberti, A., Cornara, L., Ferri, E., & Labra, M. (2017). Poisonous or non-poisonous plants? DNA-based tools and applications for accurate identification. International Journal of Legal Medi-cine, Vol. 131, pp. 1–19. https://doi.org/10.1007/s00414-016-1460-y

The EU Food Fraud Network and the Administrative Assistance and Cooperation System. 2019. Publications Office of the European Union.

Mollenkopf, D. A., Ozanne, L. K., & Stolze, H. J. (2020). A transformative supply chain response to COVID-19. Journal of Service Management. https://doi.org/10.1108/JOSM-05-2020-0143

Moreno, F. J., & Clemente, A. (2008). 2S Albumin Storage Proteins: What Makes them Food Allergens? The Open Biochemistry Journal, 2, 16–28. https://doi.org/10.2174/1874091X00802010016

Mosa, K. A., Soliman, S., El-Keblawy, A., Ali, M. A., Hassan, H. A., Tamim, A. A. B., & Al-Ali, M. M. (2018). Using DNA Barcoding to Detect Adulteration in Different Herbal Plant- Based Products in the United Arab Emirates: Proof of Concept and Validation. Recent Patents on Food, Nutrition & Agriculture. https://doi.org/10.2174/2212798410666180409101714

Mottese, A. F., Fede, M. R., Caridi, F., Sabatino, G., Marcianò, G., Calabrese, G., … Dugo, G. (2020). Chemometrics and innovative multidimensional data analysis (MDA) based on multi-element screening to protect the Italian porcino (Boletus sect. Boletus) from fraud. Food Control, 110, 107004. https://doi.org/10.1016/j.foodcont.2019.107004

Moyer, D. C., DeVries, J. W., & Spink, J. (2017). The economics of a food fraud incident – Case studies and examples including Melamine in Wheat Gluten. Food Control, 71, 358–364. https://doi.org/10.1016/j.foodcont.2016.07.015

Bouzembrak, Y., Steen, B., Neslo, R., Linge, J., Mojtahed, V., & Marvin, H. J. P. (2018). Development of food fraud media monitoring system based on text mining. Food Control, 93, 283-296. https://doi.org/10.1016/j.foodcont.2018.06.003

Muñoz-Colmenero, M., Martínez, J. L., Roca, A., & Garcia-Vazquez, E. (2016). Detection of Different DNA Animal Species in Commercial Candy Products. Journal of Food Science, 81(3), T801–T809. https://doi.org/10.1111/1750-3841.13225

Newmaster, S. G., Grguric, M., Shanmughanandhan, D., Ramalingam, S., & Ragupathy, S. (2013). DNA barcoding detects contamination and substitution in North American herbal products. BMC Medicine, 11(1), 222. https://doi.org/10.1186/1741-7015-11-222

Steven G, Newmaster & Subramanyam, R. (2009). Testing plant barcoding in a sister species complex of pantropical Acacia (Mimosoideae, Fabaceae). Molecular ecology resources, 9, 172-180.

Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., & Hase, T. (2000). Loop-mediated isothermal amplification of DNA. Nucleic Acids Research, 28(12), e63. https://doi.org/10.1093/nar/28.12.e63

Olaimat, A. N., Shahbaz, H. M., Fatima, N., Munir, S., & Holley, R. A. (2020). Food Safety During and After the Era of COVID-19 Pandemic. Frontiers in Microbiology, 11, 1854. https://doi.org/10.3389/fmicb.2020.01854

Paolocci, F., Rubini, A., Riccioni, C., Topini, F., & Arcioni, S. (2004). Tuber aestivum and Tuber uncinatum: Two morphotypes or two species? FEMS Microbiology Letters, 235(1), 109–115. https://doi.org/10.1016/j.femsle.2004.04.029

Pardo, M. Á., & Jiménez, E. (2020). DNA barcoding revealing seafood mislabeling in food services from Spain. Journal of Food Composition and Analysis. https://doi.org/10.1016/j.jfca.2020.103521

Parente, E., De Filippis, F., Ercolini, D., Ricciardi, A., & Zotta, T. (2019). Advancing integration of data on food microbiome studies: FoodMicrobionet 3.1, a major up-grade of the FoodMicrobionet database. International Journal of Food Microbiolo-gy, 305, 108249. https://doi.org/10.1016/j.ijfoodmicro.2019.108249

Parvathy, V. A., Swetha, V. P., Sheeja, T. E., Leela, N. K., Chempakam, B., & Sas-ikumar, B. (2014). DNA Barcoding to Detect Chilli Adulteration in Traded Black Pepper Powder. Food Biotechnology. https://doi.org/10.1080/08905436.2013.870078

Parveen, I., Gafner, S., Techen, N., Murch, S. J., & Khan, I. A. (2016). DNA barcoding for the identification of botanicals in herbal medicine and dietary supplements: Strengths and limitations. Planta Medica. https://doi.org/10.1055/s-0042-111208

Pawar, R. S., Handy, S. M., Cheng, R., Shyong, N., & Grundel, E. (2017). Assessment of the Authenticity of Herbal Dietary Supplements: Comparison of Chemical and DNA Barcoding Methods. Planta Medica. https://doi.org/10.1055/s-0043-107881

Pecoraro, C., Crobe, V., Ferrari, A., Piattoni, F., Sandionigi, A., Andrews, A. J., … Tinti, F. (2020). Canning processes reduce the DNA-based traceability of commercial tropical tunas. Foods. https://doi.org/10.3390/foods9101372

Piepenburg, O., Williams, C. H., Stemple, D. L., & Armes, N. A. (2006). DNA Detection Using Recombination Proteins. PLoS Biology, 4(7), e204. https://doi.org/10.1371/journal.pbio.0040204

Prosser, S. W. J., & Hebert, P. D. N. (2017). Rapid identification of the botanical and entomological sources of honey using DNA metabarcoding. Food Chemis-try, 214, 183–191. https://doi.org/10.1016/j.foodchem.2016.07.077

Raclariu, A. C., Ţebrencu, C. E., Ichim, M. C., Ciupercă, O. T., Brysting, A. K., & de Boer, H. (2018). What's in the box? Authentication of Echinacea herbal products using DNA metabarcoding and HPTLC. Phytomedicine, 44, 32–38. https://doi.org/10.1016/j.phymed.2018.03.058

Raclariu, A. C., Paltinean, R., Vlase, L., Labarre, A., Manzanilla, V., Ichim, M. C., ... & de Boer, H. (2017). Comparative authentication of Hypericum perforatum herbal products using DNA metabarcoding, TLC and HPLC-MS. Scientific reports, 7(1), 1-12.

Ichim, M. C. (2019). The DNA-based authentication of commercial herbal products reveals their globally widespread adulteration. Frontiers in pharmacology, 10, 1227. https://doi.org/10.3389/fphar.2019.01227

Ribani, A., Schiavo, G., Utzeri, V. J., Bertolini, F., Geraci, C., Bovo, S., & Fontanesi, L. (2018). Application of next generation semiconductor based sequencing for species identification in dairy products. Food Chemistry, 246, 90–98. https://doi.org/10.1016/j.foodchem.2017.11.006

Robson, K., Dean, M., Haughey, S., & Elliott, C. (2021). A comprehensive review of food fraud terminologies and food fraud mitigation guides. Food Control, Vol. 120, p. 107516. https://doi.org/10.1016/j.foodcont.2020.107516

Cruse, C. (2019). Food Fraud and the Food, Drug, and Cosmetic Act: Bridging a Disconnect. Food & Drug LJ, 74, 322.

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. PeerJ. https://doi.org/10.7717/peerj.2584

Sarkis, J. (2020). Supply chain sustainability: learning from the COVID-19 pandemic. International Journal of Operations and Production Management. https://doi.org/10.1108/IJOPM-08-2020-0568

Sasaki, Y., & Nagumo, S. (2007). Rapid identification of Curcuma longa and C. aromatica by LAMP. Biological and Pharmaceutical Bulletin, 30(11), 2229–2230. https://doi.org/10.1248/bpb.30.2229

Sasaki, Y., Komatsu, K., & Nagumo, S. (2008). Rapid Detection of Panax ginseng by Loop-Mediated Isothermal Amplification and Its Application to Authentication of Ginseng. Biological & Pharmaceutical Bulletin, 31(9), 1806–1808. https://doi.org/10.1248/bpb.31.1806

Segelke, T., Schelm, S., Ahlers, C., & Fischer, M. (2020). Food Authentication: Truffle (Tuber spp.) Species Differentiation by FT-NIR and Chemometrics. Foods, 9(7), 922. https://doi.org/10.3390/foods9070922

Sheu, S.-C., Tsou, P.-C., Lien, Y.-Y., & Lee, M.-S. (2018). Development of loop-mediated isothermal amplification (LAMP) assays for the rapid detection of allergic peanut in processed food. Food Chemistry, 257, 67–74. https://doi.org/10.1016/j.foodchem.2018.02.124

Shokralla, S., Hellberg, R. S., Handy, S. M., King, I., & Hajibabaei, M. (2015). A DNA Mini-Barcoding System for Authentication of Processed Fish Products. Scientific Reports. https://doi.org/10.1038/srep15894

Singh, S., Kumar, R., Panchal, R., & Tiwari, M. K. (2020). Impact of COVID-19 on logistics systems and disruptions in food supply chain. International Journal of Production Research, 1–16. https://doi.org/10.1080/00207543.2020.1792000

Song, H., Lu, B., Ye, C., Li, J., Zhu, Z., & Zheng, L. (2020). Fraud vulnerability quantitative assessment of Wuchang rice industrial chain in China based on AHP-EWM and ANN methods. Food Research International, 109805. https://doi.org/10.1016/j.foodres.2020.109805

SONG, M., DONG, G. Q., ZHANG, Y. Q., LIU, X., & SUN, W. (2017). Identification of processed Chinese medicinal materials using DNA mini-barcoding. Chinese Journal of Natural Medicines, 15(7), 481–486. https://doi.org/10.1016/S1875-5364(17)30073-0

Spink, J. W. (2019). The current state of food fraud prevention: overview and re-quirements to address 'How to Start?' and 'How Much is Enough?' Cur-rent Opinion in Food Science, Vol. 27, pp. 130–138. https://doi.org/10.1016/j.cofs.2019.06.001

Spink, J., & Moyer, D. C. (2011). Defining the Public Health Threat of Food Fraud. Journal of Food Science, 76(9), R157–R163. https://doi.org/10.1111/j.1750-3841.2011.02417.x

Spink, J., Bedard, B., Keogh, J., Moyer, D. C., Scimeca, J., & Vasan, A. (2019). Inter-national Survey of Food Fraud and Related Terminology: Preliminary Results and Discussion. Journal of Food Science, 84(10), 2705–2718. https://doi.org/10.1111/1750-3841.14705

Spink, J., Moyer, D. C., Park, H., Wu, Y., Fersht, V., Shao, B., … Edelev, D. (2015). In-troducing Food Fraud including translation and interpretation to Russian, Ko-rean, and Chinese languages. Food Chemistry, 189, 102–107. https://doi.org/10.1016/j.foodchem.2014.09.106

Spink, J., Ortega, D. L., Chen, C., & Wu, F. (2017). Food fraud prevention shifts the food risk focus to vulnerability. Trends in Food Science and Tech-nology, Vol. 62, pp. 215–220. https://doi.org/10.1016/j.tifs.2017.02.012

Srirama, R., Gurumurthy, B. R., Senthilkumar, U., Ravikanth, G., Shaanker, R. U., & Shivanna, M. B. (2014). Are mini DNA-barcodes sufficiently informative to re-solve species identities? An in silico analysis using phyllanthus. Journal of Ge-netics, 93(3), 823–829. https://doi.org/10.1007/s12041-014-0432-6

Sumar, S., & Ismail, H. (1995). Adulteration of foods -past and present. Nutrition & Food Science, 95(4), 11–15. https://doi.org/10.1108/00346659510088663

The EU Food Fraud Network and the Administrative Assistance and Cooperation System Health and Food Safety. (n.d.). https://doi.org/10.2875/326318

Théolier, J., Barrere, V., Charlebois, S., & Benrejeb Godefroy, S.(2020). Risk analy-sis approach applied to consumers' behaviour toward fraud in food products.

Trends in Food Science and Technology. https://doi.org/10.1016/j.tifs.2020.11.017

Thomas, A. C., Deagle, B. E., Eveson, J. P., Harsch, C. H., & Trites, A. W. (2016). Quantitative DNA metabarcoding: Improved estimates of species proportional biomass using correction factors derived from control material. Molecular Ecology Resources. https://doi.org/10.1111/1755-0998.12490

Ulberth, F. (2020). Tools to combat food fraud – A gap analysis. Food Chemistry, 330, 127044. https://doi.org/10.1016/j.foodchem.2020.127044

Utzeri, V. J., Ribani, A., Schiavo, G., Bertolini, F., Bovo, S., & Fontanesi, L. (2018). Application of next generation semiconductor based sequencing to detect the bo-tanical composition of monofloral, polyfloral and honeydew honey. Food Control, 86, 342-349. https://doi.org/10.1016/j.foodcont.2017.11.033

Voorhuijzen-Harink, M. M., Hagelaar, R., van Dijk, J. P., Prins, T. W., Kok, E. J., & Staats, M. (2019). Toward on-site food authentication using nanopore sequencing. Food Chemistry: X. https://doi.org/10.1016/j.fochx.2019.100035

Wang, T., Kim, S., & An, J. H. (2017). A novel CMOS image sensor system for quantitative loop-mediated isothermal amplification assays to detect food-borne pathogens. Journal of Microbiological Methods. https://doi.org/10.1016/j.mimet.2016.12.002

Xing, B., Zhang, Z., Sun, R., Wang, Y., Lin, M., & Wang, C. (2020). Mini-DNA barcoding for the identification of commercial fish sold in the markets along the Taiwan Strait. Food Control. https://doi.org/10.1016/j.foodcont.2020.107143

Xing, R. R., Wang, N., Hu, R. R., Zhang, J. K., Han, J. X., & Chen, Y. (2019). Application of next generation sequencing for species identification in meat and poultry products: A DNA metabarcoding approach. Food Control, 101, 173–179. https://doi.org/10.1016/j.foodcont.2019.02.034

Yang, Q., Domesle, K. J., & Ge, B. (2018). Loop-Mediated Isothermal Amplification for Salmonella Detection in Food and Feed: Current Applications and Future

Directions. Foodborne Pathogens and Disease. https://doi.org/10.1089/fpd.2018.2445

Zhao, M., Wang, B., Xiang, L., Xiong, C., Shi, Y., Wu, L., … Sun, W. (2019). A novel onsite and visual molecular technique to authenticate saffron (Crocus sativus) and its adulterants based on recombinase polymerase amplification. Food Control, 100, 117–121. https://doi.org/10.1016/j.foodcont.2019.01.011

Grazina, L., Amaral, J. S., & Mafra, I. (2020). Botanical origin authentication of dietary supplements by DNA-based approaches. Comprehensive Reviews in Food Science and Food Safety, Vol. 19, pp. 1080–1109. https://doi.org/10.1111/1541-4337.12551

Food Fraud Summary October 2020 | Knowledge for policy. (n.d.). Retrieved December 17, 2020, from https://knowledge4policy.ec.europa.eu/food-fraud-quality/food-fraud-summary-october-2020_en

Canadian Food Inspection Agency. Retrieved December 17, 2020, from https://www.inspection.gc.ca/food-safety-for-industry/information-for-consumers/food-safety-system/food-fraud/eng/1548444446366/1548444516192

Websites

The Implications of Food Fraud - Page 3 of 3 - Food Quality & Safety. (n.d.). Retrieved October 22, 2020, from https://www.foodqualityandsafety.com/article/the-implications-of-food-fraud/3/

Cagle, S. (2020), “A disastrous situation': mountains of food wasted as coronavirus scrambles supply chain”, The Guardian, 9 April 2020, available at: https://www.theguardian.com/world/2020/apr/09/us-coronavirus-outbreak-agriculture-food-supply-waste?CMP=share_btn_fb (accessed 27 December 2020).

Charles, D. (2020), “Food shortages? Nope, too much food in the wrong places”, The Salt, on National Public Radio, 3 April 3 2020, available at:

https://www.npr.org/sections/thesalt/2020/04/03/826006362/food-shortages-nope-too-much-food-in-the-wrong-places?utm_campaign=story-shar%E2%80%A6 (accessed 27 December 2020).

Corkery, M. and Yaffe-Bellany (2020b), "'We had to do something' trying to prevent massive food waste", New York Times, 2 March 2020, available at: https://www.nytimes.com/2020/05/02/business/coronavirus-food-waste-destroyed.html?referringSource=articleShare (accessed 27 December 2020).

Daniel Gerzon Mahler, Christoph Laknerr, Andres Castaneda Aguilar, Haoyu Wu. 2020. Retrieved on https://blogs.worldbank.org/opendata/updated-estimates-impact-covid-19-global-poverty

Dunkley, D. (2020), "Cashierless stores and online deliveries are the future of supermarket shopping", Stuff.co.nz, 30 March 2020, available at: https://www.stuff.co.nz/business/better-business/120587848/cashierless-stores-and-online-deliveries-are-the-future-of-supermarket-shopping?r%E2%80%A6 (accessed 27 December 2020).

Effect of Globalization on the Food Industry. (n.d.). Retrieved December 22, 2020, from https://www.ukessays.com/essays/nutrition/effect-globalization-food-industry-4414.php

Elliot, C. (2020), "Food fraud – one of the big winners during the COVID pandemic" available at: https://www.newfoodmagazine.com/article/115243/food-fraud/ (accessed 27 December 2020).

Everstine K. (2020) "COVID-19 and Food Fraud Risk" available at: https://foodsafetytech.com/column/covid-19-and-food-fraud-risk/ (accessed 27 December 2020).

Fresh Plaza (2020) "False fragole nel Burgenland" available at: https://www.freshplaza.de/article/9222681/falsche-erdbeeren-im-burgenland/ (accessed 27 December 2020).

Smith, J. (2020), “Warehouse hiring surge defies crashing U.S. jobs market”, Wall Street Journal, 3 April 2020, available at: https://www.wsj.com/articles/warehouse-hiring-surge-defies-crashing-u-s-jobs-market-11585939391 (accessed 27 December 2020).

Taylor P. (2020) “New horsemeat incidents as study highlights beef fraud risks” available at: https://www.securingindustry.com/food-and-beverage/new-horsemeat-incidents-as-study-highlights-beef-fraud-risks/s104/a11685/#.X-hsddhKi72 (accessed 27 December 2020).

An Update On Food Fraud | FSNS. (n.d.). Retrieved December 17, 2020, from http://fsns.com/news/an-update-on-food-fraud