DEPARTMENT OF AGRICULTURAL, FOOD AND ENVIRONMENTAL SCIENCES

DEGREE COURSE: Master's Degree in Food and Beverage Innovation and Management

DNA BARCODING FOR TRACEABILITY ON PLANT-BASED FOOD

TYPE OF DISSERTATION: Research.

Student:
MARIANA TOSTADO MONACO
S1093149

Supervisor/Mentor:
DOTT.SSA ELENA BITOCCHI

ACADEMIC YEAR 2020-2021

CONTENTS:

LIST OF TABLES: 3

LIST OF FIGURES: 3

CHAPTER 1. THESIS INTRODUCTION 5

CHAPTER 2. DNA BARCODING 7

2.1. DEFINITON 7
2.2. HISTORIC BACKGROUND 7
2.3. MOLECULARIZATION OF IDENTIFICATION PROCESS 10
2.4. STANDARIZATION OF THE PROCEDURES 13
2.5. COMPUTARIZATION 16

CHAPTER 3. DNA BARCODING TO IDENTIFY AND CERTIFY FOOD RAW MATERIAL 22

3.1. DNA IN FOOD RAW MATERIALS 23
3.2. LEGISLATION 24
3.3. COMPLEMENTARY METHODOLOGIES 26
3.4. ADVANTAGES AND LIMITATIONS OF DNA BARCODING 29

CHAPTER 4. DNA BARCODING APPLICATION ON EDIBLE PLANT-BASED PRODUCTS 30

CHAPTER 5. CASES OF STUDY 32

5.1. CASE STUDY 1- *TRNH-PSBA* FOR COFFEE TRACEABILITY 32
5.2. CASE STUDY 2- BAR-HRM FOR IDENTIFICATION OF PGI LENTIL VARIETY "EGLOUVI" 35
5.3. CASE STUDY 3- DNA METABARCODING FOR HONEY TRACEABILITY 38
5.4. CASE STUDY 4- WINE TRACEABILITY THROUGH MICROBIOME BARCODING 44
5.5. CASE STUDY 5- DNA BARCODING FOR HIGHLY PROCESSED PRODUCTS: THE AÇAÍ CASE. 47

CHAPTER 6. CONCLUSIONS 50

CHAPTER 7. BIBLIOGRAPHY 52

LIST OF TABLES:

Table 1. *Percentage success in classifying species to membership of a particular taxonomic group based upon sequence variation at COI. n, number of taxa that were calssified using each taxon 'profile'. Source: Herbert et al. 2003.* 10
Table 2*. Percentage of nucleotide sequence divergence (K2P distance) at COI among members of fivel lepidopteran families at three levels of taxonomic affinity. n, number of genera with two or more species, while at the family level it indicates the total number of species that were analysed. Source: Herbert et. al. 2003.* 10
Table 3*. Target DNA Barcoding genomic regions per phyla.* 13
Table 4. *Summary of the bioinformatics methods useable to analyse sequences. Source: Casiraghi et al., 2010.* 19
Table 5. *Summary of genetic K2P divergences (%) within different taxonomic levels. Data are from 754 sequences from 207 species and 122 genera; [a], one example in Pristiophorus and one in Plectropomus, where in each instance one sequence among multiple specimens appeared to be of a different, but congeneric, species; [b], Hydrolagus novazaelandiae. Source: Hebert et al. (2005)* 20
Table 6. *Comparison of full-lenght DNA barcode and mini-barcode. Source: Hajibabaei and McKenna 2012.* 27
Table 7. *Advantages and limits of the use of DNA Barcoding for food taceability.* 29
Table 8. *Example applications of DNA Barcoding techonolgy in plant-based food products.* 32
Table 9. *Barcode amplicon sizes of six plant species used in the Unco and Unco 2018 study.* 34
Table 10. *Development of the Mi-Seq library.* 42
Table 11. *Random Forests Models predict vineyards origin of grape musts; [a], Error rate: percentage of misclassification of out-of-bag must samples to the wrong vineyard; [b], Random rate: percentage of misclassification expected due to random error; c, ER/Random: Error rate / Random rate.* 47

LIST OF FIGURES:

Figure 1. *Traceability tools. ITs, Informatic technologies; QR code, Quick Response Code; RFDI, Radio Frequency Identification; NFC, Near Field Communication; NIR, Near Infrared spectroscopy.* 6
Figure 2. *Hierarchy of Biological Classification. Source: NCERT, 2015.* 9
Figure 3*. Mitochondrial DNA location.* 12
Figure 4. *BOLD Website (accessed on 25 November 2021); Barcode Index Number (BIN) System clusters sequences using well established algorithms to produce operational taxonomic units that closely correspond to species.* 17
Figure 5. *BOLD Sampling Points (accessed on 25 November 2021)* 17
Figure 6. *Distribution of K2P distance (>1%) for cox1 within different taxonomic categories. Source: Hebert et al. (2005)* 20
Figure 7. *Workflow from sequence to phylogenetics. Source: Kress and Erickson 2012.* 21
Figure 8. *DNA Barcoding vs DNA Metabarcoding. Source: Galimberti et. al., 2019. DNA barcoding and metabarcoding pipelines. Food products can be pure (e.g. saffron) or mixtures of different species (e.g. multifloral honey). For DNA barcoding the DNA from each food product is extracted separately, amplified by PCR at specific loci (e.g. ITS2) and Sanger sequenced, to produce barcode sequences which are assigned to a species by comparing them against a reference database (e.g. NCBI GenBank). In DNA metabarcoding, the whole food product is homogenized and DNA is extracted directly from this. PCR is performed on the DNA extract with primers usually designed for a shorter read length to comply with the chemical and technical features of High Throughput Sequencing (HTS) instruments. Once HTS of the PCR products is performed, bioinformatics pipelines are used to extract unique sequence read and to assign them to species using again a DNA barcode reference database.* 26
Figure 9. *Comparison of DNA barcode size versus proportion of species identified reveals the efficiency of mini-barcodes in resolving species. Sequence read lengths typically obtained from three commonly used next-generation sequencing technologies as well as Sanger sequencing are shown on the graph. It is clear that 454 pyrosequencing and Sanger are currently optimal technologies for mini-barcode and full-barcode recovery. Source: Meusnier et al. 2008.* 27

Figure 10. *DNA barcoding steps.* ... 33
Figure 11. *Capillary Electropherograms of mixtures. A, trnH-psbA intergenic spacer amplification profiles of coffee: soybean blends; B, trnH-psbA intergenic spacer amplification profiles of coffee: corn blends; C, trnL (UAA) intron amplification profiles of coffee: barley blends. The order of admixing ratios (coffee: adulterant) is the same for rows A, B and C: I, 70:30%; II, 90:10%; III, 95:5%; IV, 99:1%.* ... 35
Figure 12. *Steps of Bar-HRM* ... 36
Figure 13. *Plot of normalized fluorescence of mixed L. culinaris/V. sativa.* ... 37
Figure 14*. Alignment of rpoC1 sequence of L. culinaris and V. faba.* ... 37
Figure 15*. Step applied for DNA metabarcoding of honey samples.* ... 39
Figure 16*. Pie charts of the botanical composition of honeys from China (A), Chile (B), Vietnam (C), and Argentina (D)* ... 42
Figure 17*. Example pie charts of the botanical composition of chestnut honey (A), acacia honey (B), rhododendron honey (C), and a polyflower honey (D)* ... 43
Figure 18. *Map of sampling sites across Napa and Sonoma Counties. Each point represents an individual vineyard from which grapes were harvested for the fermentations monitored in the study. Points are colored by AVA designation, as indicated in the legend* ... 45
Figure 19*. DNA barcoding steps.* ... 45
Figure 20*. PCoA comparisons of bacterial weighted UniFrac distance (left two columns) and fungal Bray-Curtis dissimilarity (right two columns) in musts and wines (see column labels), categorized by vineyard (color) and AVA source (shape). (A) Chardonnay, (B) Cabernet Sauvignon. Each point represents an individual sample, and sample proximity on the plot is a function of similarity in bacterial and fungal community composition* ... 46
Figure 21*. Açai Market Products; a) açaí capsules; b) açaí powder; c) pure frozen açaí pulp; d) açaí popsicle; e) açaí sorbet* ... 48
Figure 22*. DNA barcoding steps.* ... 48
Figure 23*. PCR amplification success (a) and sequencing success rates (b) of the eight DNA barcoding regions in reference samples.* ... 49
Figure 25*. Authentication test using psbK-I as DNA barcoding in commercial products. a) a total of 50 products were tested; b) a total of 35 high-quality sequences were identified depending on the concordance between the species expected and sequence obtained.* ... 50
Figure 26*. Trend in food traceability publications. Source: Violino, 2019.* ... 50

Chapter 1. THESIS INTRODUCTION

There is an increasing need to provide high quality and safe food to consumers worldwide. Globalization has proudly put to our reach an enormous variety of food products in terms of flavor and nutrition options; but this also represents a big challenge in terms of traceability, transparency and quality assurance of those food products.

The aim of this thesis is to analyze the relevance of DNA barcoding methodology as a tool to satisfy the required tracking system that the food industry needs to recognize, trace and authenticate the plant-based food products efficiently, even industrially processed ones.

Traceability allows the tracking of the source of a food at any point in the production chain enabling the quality-control processes and cutting down the production of unsafe or poor-quality foods. Food authentication is the process through which a food is tested to verify if it complies with the description contained on its label.

DNA barcoding can genetically identify the raw materials contained in a food product, allowing to detect a food alteration or replacement (either if it was accidentally or voluntary) as well as its origin.

DNA barcoding can be the solution to not only properly characterize the food products that are already present on the market, but most importantly, to prevent the risks along the whole food chain. These risks include food fraud (replacement of a product for another), misleading information (to claim the product contains an unexpected ingredient), presence of allergens and toxicity. DNA barcoding tool has the added value that since it decodes the genetic origin of the food it can also be used as a tool for tracing a food product to its origin (geographical location, terroir, environmental and cultivation techniques effects on the product), solving taxonomic problems, supporting breeding programs, tracking GMOs, and protecting quality designation products (PDO, PGI, organics).

Food traceability is defined by the European Union General Food Law as “the ability in all production, processing and sales stages of tracing and tracking the expected food”. Therefore, a traceability system should provide information of the whole food chain, from the origin of the raw materials to the point of consumption and backwards. Traceability tools that have been used over the past years include (but are not limited to) the ones represented on **Figure 1**.

The Informatic Technologies (IT) comprise Quick Response Code (QR code), Radio Frequency Identification (RFDI) and Near Field Communication (NFC), that are basically based on physical or virtual tags that are scanned so that they link to a web-based database to retrieve its tracing information. The photochemical technology refers to Near Infrared spectroscopy (NIR), a non-destructive method for determination of the chemical compounds of the product by non-visible light absorption.

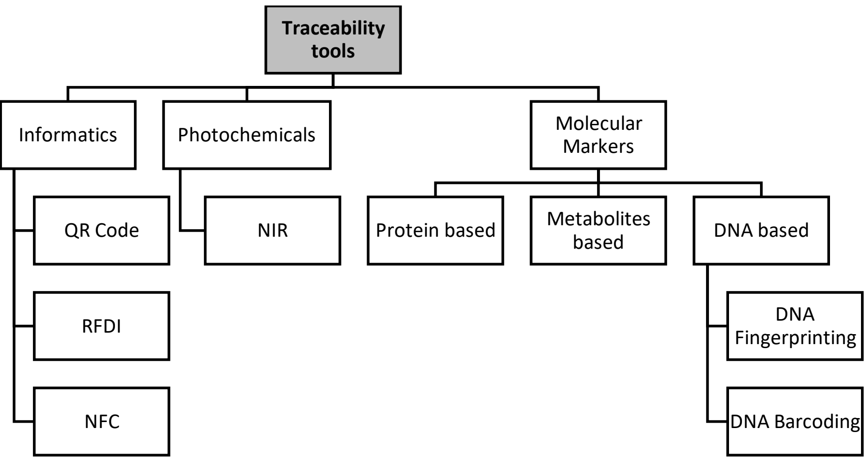


Figure 1. Traceability tools. ITs, Informatic technologies; QR code, Quick Response Code; RFDI, Radio Frequency Identification; NFC, Near Field Communication; NIR, Near Infrared spectroscopy.

Concerning molecular methods, three types of molecules can be used as a marker: i) DNA, ii) proteins and iii) metabolites. When using proteins or metabolites contained in a food product as markers for food traceability, it is important to consider that they are usually variable since they often suffer from inconsistencies due to farming system (e.g. agro-ecosystem conditions and agricultural practices) and processing methods, as they are affected by environmental conditions and industrial. This leds to variability even within not only individuals of the same species but also within the same individual (pure lines or clones for plants).

The protein-based methods include immunological assays, electrophoretic and chromatographic techniques such as High-performance liquid chromatography (HPLC) and Thin-layer chromatography (TLC), that research shows can be effective on fresh food, but not really on the processed food. (Fanelli et al. 2021).

On the other hand, DNA markers are known for being more reliable since are related to the genotype. DNA is present in nearly all the cells of a given organism and its sequence remains unchanged during all the production phases, and compared to metabolites and proteins, DNA is a much more stable and resistant molecule to industrial transformation; moreover, DNA sequence of an individual is unique, allowing to distinguish species and varieties within species (Fanelli et al. 2021). Techniques used to genotype a sample with DNA markers can be either hybridization-based (where a known sequence of the target species is highlighted by the used of a probe) or PCR based (where primers target a locus for their amplification). Molecular markers such as, for example, randomly amplified polymorphic DNA (RAPD), Amplified fragment length polymorphism (AFLP), simple sequence repeats (SSRs), and their variants (e.g., inter simple sequence repeats, ISSR, and sequence-specific amplification polymorphism, SSAP) have been successfully adopted for the characterization of crop species, as well as sequencing-based markers such as single nucleotide polymorphisms (SNPs) (Lockley and Bardsley 2000; Kumar et al. 2009; Martins-Lopes et al. 2013; Scarano et al. 2014; Corrado 2016; Danezis et al. 2016; Lo et al. 2018; Böhme et al. 2019).

DNA based traceability methods include: a) DNA fingerprinting (based on variability and polymorphism of the DNA between organisms) for molecular traceability, for which it is important to consider that many agricultural products have not yet been characterized with private markers, such as SSR or SNPs, that allow to have a reliable DNA fingerprinting system); or, b) the worldwide popular DNA based method known as “DNA barcoding”, that is a Single Region Approach, where the investigation focuses on a specific and well-known plastidial (cpDNA) or mitochondrial (mtDNA) target DNA

region sequenced by using universal markers. DNA barcoding presents two avantages over DNA fingerprint: i) it does not require an extensive knowledge of the organism's genome and ii) the fact that it uses short DNA sequences prevents the risk of DNA fragmentation that characterises highly processed food; thus, DNA barcoding is a highly attractive and reproducible traceability tool. In this regard, DNA Barcoding represents an accurate essay for traceability of plant-based food products.

Plant-based food products include edible products derived from plants, including aromatic spices, medicinal plants, tea, honey, food supplements, additives, even highly processed products such as bread, pasta or wine, and complex food matrix such as mixed products (vegetables soups, sauces, seven cereals flours, ready to eat salads, etc...).

The economic importance of a robust and efficient method for traceability relies in the fact that food piracy represents a well-known bad practice worldwide and it's valued in around 200 billion USD according to the Organization for Economic Co-operation and Development (OECD). Even if countries make an effort to state punishments by law to food piracy, there is a still latent need for a system in place to actually prevent and identify the products that are not aligned to the current food law regulations, putting into risk the health and rights of the consumers and leaving them exposed and unprotected. The utility of using a DNA barcode as a reliable trace back system has been proved to be efficient by many studies (Erickson 2008, CBOL 2009, Fanelli 2021), and nowadays is possible to also rely on laboratories that can provide this service publicly and at a relative low cost. Therefore, DNA barcoding offers a solution for check and control an important number of food chains. In this thesis we focus on the potential of the application of DNA barcoding techniques for traceability of edible plants food products.

Chapter 2. DNA barcoding

2.1. Definiton

DNA barcoding is a molecular technology based on the analysis of the variability present within a standard region of the genome, "DNA barcode" that allows the identification of biological species (Hebert et al. 2003). The name "DNA barcoding" figuratively refers to the way an infrared scanner univocally identifies a product by using the black stripes of the Universal Product Code (UPC).

2.2. Historic Background

It is necessary to first set the basis of how DNA Barcoding technology was born. DNA Barcoding founder, Paul Hebert, is a Canadian biology scientist. He and his team published a research paper on 2003 exposing for the first time the concept of DNA barcoding as a solution for taxonomy identification of life diversity, called "Biological identifications through DNA barcodes", based on the problem that whereas physicists deal with a cosmos assembled from 12 fundamental particles, biologists deals with a planet populated by millions of species, the discrimination of which is not an easy task. Indeed, as indicated by Hammond (1992), taxonomists can critically identify not much more than 0.01% of the estimated 10–15 million species living in our planet. Moreover, a huge gap of time and money that would be required to identifiy all species, unless a new

technology for taxon recognition is applied. It is evident that morphology-based identification systems are too limited, meanwhile, microgenomic identification systems, which permit life's discrimination through the analysis of a small segment of the genome, are one extremely effective approach to the diagnosis of biological diversity (Herbert et. al. 2003). This concept of using a small genome segment had previously already gained broad acceptance among those working with the least morphologically tractable groups, such as viruses, bacteria and protists, such as for example, Nanney (1982) and Pace (1997). Is important to mention that by the time Herbert was publishing his reseach, DNA identification system was not a new techonolgy, is had already been applied to higher organisms (see for instance Brown et al. 1999; Bucklin et al. 1999; Trewick 2000; Vincent et al. 2000), where genomic approaches to taxon diagnosis exploit diversity among DNA sequences to identify organisms (Kurtzman 1994; Wilson 1995). These DNA sequences can be viewed as genetic 'barcodes' that are embedded in every cell.

The "DNA barcode" term was inspired by the idea of the Universal Product Codes, used to identify retail products, they employ 10 alternate numerals at 11 positions to generate 100 billion unique identifiers. In the case of the proposed genomic barcodes, they have only four alternate nucleotides at each position, but the string of sites available for inspection is huge. The survey of just 15 of these nucleotide positions creates the possibility of 4^{15} (1 billion) codes; as has been mentioned in the last paragraph, there are only about 10-15 million of species, so these combination of numbers already opens the possibility for combinations that are 100 times the number that would be required to discriminate life if each taxon was uniquely branded. Now, this is genetics, so there would be a challenge to adress when choosing the right amount of nucleotide sequence, since in this case due to functional constraints some nucleotide positions will remain constant while some intraspecific diversity will exist at other positions. Therefore, what Hebert proposed is that the impact of functional constraints can be reduced by focusing on a protein-coding gene, given that most shifts at the third nucleotide position of codons are weakly constrained by selection because of their four-fold degeneracy. Hence, by examining any stretch of 45 nucleotides, one gains access to 15 sites weakly affected by selection and, therefore, 1 billion possible identification labels. In practice, there is no need to constrain analysis to such short stretches of DNA because sequence information is easily obtained for DNA fragments hundreds of base pairs (bp) long. This ability to inspect longer sequences is significant, given two other biological considerations that Hebert noted: first, nucleotide composition at third-position sites is often strongly biased (A–T in arthropods, C–G in chordates), reducing information content. However, even if the A–T or C–G proportion reached 1, the inspection of just 90 bp would recover the prospect of 1 billion alternatives ($2^{30} = 4^{15}$). So, at this point, an amount of 90 nucleotides length is proposed. The second constraint derives from the limited use of this potential information capacity since most nucleotide positions are constant in comparisons of closely related species. However, given a modest rate (for example 2% per Myr) of sequence change, he believed one expects about 12 diagnostic nucleotide differences in a 600 bp comparison of species with just a million-year history of reproductive isolation. Therefore, a length of around 600bp should be appropriate.

In regard to which protein- coding nucleotide sequence location Herbert would choose, it can be said that although there has never been an effort to implement a microgenomic identification system on a large scale, enough work has been done to indicate that as a key design element the target DNA segment could be mitochondrial, since mitochondrial

genome of animals is a better target for analysis than the nuclear genome because of three main reasons:

- Its composition; mithocondrial DNA is characterized by its lack of introns, its absence of recombination and its haploid mode of inheritance

- Already existing robust primers enable the routine recovery of specific segments of the mitochondrial genome

- Past phylogenetic work has often focused on mitochondrial genes encoding ribosomal (12S, 16S) DNA, but their use in broad taxonomic analyses is constrained by the prevalence of insertions and deletions (indels) that greatly complicate sequence alignments (as stated by Doyle and Gaut, 2000). The 13 protein-coding genes in the animal mitochondrial genome are better targets because indels are rare since most lead to a shift in the reading frame.

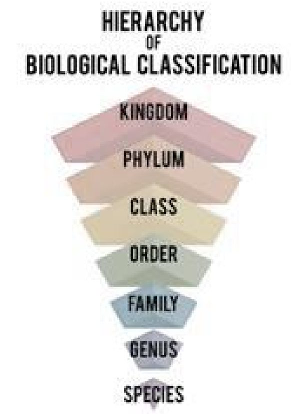


Figure 2. Hierarchy of Biological Classification. Source: NCERT, 2015.

Even if there was no compelling *a priori* reason to focus analysis on a specific gene, Hebert chose the cytochrome c oxidase (*COI*)I gene, being his experiment focused on animals. Such gne has two important advantages: first, the universal primers for this gene are very robust, enabling sequencing of such gene for most, if not all, animal phyla (as cited by Folmer et al. 1994; Zhang and Hewitt 1997). Second, *COI* appears to possess a greater range of phylogenetic signal than any other mitochondrial gene. In common with other protein-coding genes, its third-position nucleotides show a high incidence of base substitutions, leading to a rate of molecular evolution that is about three times greater than that of 12S or 16S rDNA (Knowlton and Weigt 1998). In fact, the evolution of this gene is rapid enough to allow the discrimination of not only closely allied species, but also phylogeographic groups within a single species (Cox and Hebert 2001, Wares and Cunningham 2001). By evaluating amino acid substitutions, it may be possible to assign any unidentified organism to a higher taxonomic group (phylum, order), before examining nucleotide substitutions to determine its species identity.

Nowadays, other mtDNA regions besides *COI* are also used for DNA barcoding, but it remains an informative region of around 600 bp (669 bp for the phylum analysis, 624 bp for the ordinal analysis and 617 bp for the species-level analysis) that provides an almost unique profile and more than 95% of species in test assemblages of varied animal groups have been shown to possess distinctive COI sequences (Hajibabaei et al. 2006)

The Herbert's aim was to distinguish three levels of biological classification: phylum, order and species. He proposed a "similarity" approach for the analyses at the ordinal and phylum levels by examining amino-acid divergences using Poisson corrected p-distances to reduce the impacts of homoplasy. For the species-level analysis, nucleotide-sequence divergences were calculated using the Kimura-two-parameter (K2P) model, the best model when distances are low (Nei and Kumar 2000).

Afterwards, Herbert applied Neighbour-joining (NJ) analysis (Kumar et al. 2001), to investigate the relationships among taxa in the profiles and subsequently classify 'test'

taxa, thanks to the strong track record in the analysis of large species assemblages (Kumar & Gadagkar 2000); in this way Herbert expected to appreciate the relationship at the three levels (phyla, order and species)

Herbert's results showed that analyzing the *COI* divergences gave a 100% percentage of success in classifying species, order and taxa and 96% at phyla level (**Table 1**).

Taxon	Target group	*n*	% success
Kingdom Animals	7 phyla	55	96.4
Class Hexapoda	8 orders	50	100
Order Lepidoptera	200 species	150	100

Table 1. Percentage success in classifying species to membership of a particular taxonomic group based upon sequence variation at COI. n, number of taxa that were calssified using each taxon 'profile'. Source: Herbert et al. 2003.

Table 2 shows the percentage of nucleotide sequence divergence at COI among members of fivel lepidopteran families.

Family	n	Within species	n	Within genus	n	Within family
Arctiidae	13	0.33	4	7.0	18	10.0
Geometridae	30	0.23	10	9.1	61	12.5
Noctuidae	42	0.17	12	5.8	90	10.4
Notodontidae	14	0.36	4	5.9	20	12.4
Sphingidae	8	0.17	3	6.4	11	10.5

Table 2. Percentage of nucleotide sequence divergence (K2P distance) at COI among members of fivel *lepidopteran* families at three levels of taxonomic affinity. n, number of genera with two or more species, while at the family level it indicates the total number of species that were analysed. Source: Herbert et. al. 2003.

According to the obtained results, Herbert established divergence thersholds to define wheter the samples belong to the same species or not, that are nowadays used. Hebert concluded with this consideration "*We believe that a COI database can be developed within 20 years for the 5–10 million animal species on the planet (Hammond 1992; Novotny et al. 2002) for approximately $1 billion, far less than that directed to other major science initiatives such as the Human Genome project or the International Space Station.*" (Hebert, 2003). Nowadays the technology and method has been strongly improved, costs are really decreased and it is possible to unequivocally identify organisms at the species level (Meier et al. 2016).

Hereafter the three main characteristiscs that make the DNA barcoding a very useful tool for modern taxonomy.

2.3. Molecularization of identification process

This characteristic refers to the molecularization that is the use of the variability in a molecular marker as a discriminator. Although the molecular approach at the basis of DNA barcoding is not new to science, the selection of the most suitable molecular approach depends on different aspects, including the amount of genetic variation of the analyzed species, the time needed for the analysis, the cost/effectiveness ratio, and the expertise of laboratories. Furthermore, genomic techniques require high quality DNA to work successfully because effectiveness can be negatively influenced by altered or

fragmented DNA (Hellberg and Morrisey, 2011; Meusnier et. al., 2008; Pafundo, et. al., 2007). However, to be highly species-specific, molecular approaches require access to the correct DNA sequence of the organisms, and their application is often limited to a single taxon, or to closely related taxa. Differently, DNA barcoding approach proved to be useful in solving taxonomic problems in several theoretical and practical applications (Hollingsworth et. al., 2011; Rasmussen et. al., 2009; Valentini, et. al., 2009). For this purpose, it is important to choose the right analytical software (see Chapter 5 about bioinformatics) and to select the right marker (that is not exclusively the one (i.e. *COI*) initially proposed by Herbert, but other markers can work as well as long as they present the following fundamental characteristics that an ideal DNA barcode requires:

- High taxonomic coverage: the marker has conserved flanking sites for developing universal PCR primers for the widest taxonomic application, to obtain the correct amplification of the genomic region chosen as DNA barcode in the widest panel of taxa.
- High resolution: the marker must contain significant species-level genetic variability and divergence to ensure the identification of different taxa, based on interspecific differences in DNA barcode sequences. As a general principle, DNA barcode regions should have a high interspecific, and low intraspecific variability. This is because the more overlap there is between genetic variation within species and differentiation separating sister species, the less effective DNA barcoding becomes for taxonomic identifications. The target DNA region should contain enough phylogenetic information to easily assign species to its taxonomic group (genus, family, etc.)
- Ideal lenght: the marker must have an appropriate sequence lenght to facilitate current capabilities of DNA extraction. DNA barcodes consist of a standardized short sequence of DNA between 400 and 800 bp long that, in theory, can be easily isolated and characterized for all species on the planet. It has to be short enough to allow amplification of degraded DNA.
- Robust: the marker should be extremely robust, with highly conserved sites for annealing of primers, which guarantee highly reliable DNA amplifications and sequencing. This is particularly important when using environmental DNA, for which each extract contains a mixture of many species to be identified at the same time.

Its important to consider that evolution is still happening, so when we aim to identify a species by the molecular profile is important to consider that if a shift or variability has already mutated in that species the molecular profile might not match. This risk is reduced by the fact that a single DNA barcoding marker is used every time.

Many researchers keep on validating how the DNA markers have become the most effective instrument in the analysis of the DNA of plant cultivars and animal breeds and are also used to track the raw materials in food industry processes (Kumar, Gupta, Misra, Modi, & Pandey, 2009; Mafra et al., 2008; Woolfe & Primrose, 2004). The required "DNA barcode" region can be represented by short segments of the chloroplast or mitochondrial genome (abbreviated as cpDNA and mtDNA for plant and animal foodstuffs, respectively) or even, by intergenic regions. In both cases, these target regions usually belong to the extra-nuclear genome of the species (**Figure 3**). These standard regions are short and they have shown sufficient nucleotide variability to assess the taxonomic identity of the majority of organisms as belonging to a particular individual, breed/cultivar, or species. Since the chloroplast and mitochondrial genomes are present in many copies within each cell, this technology is being more easily exploited to recover

information even in degraded DNA samples or transformed materials deriving from crop varieties and livestock species.

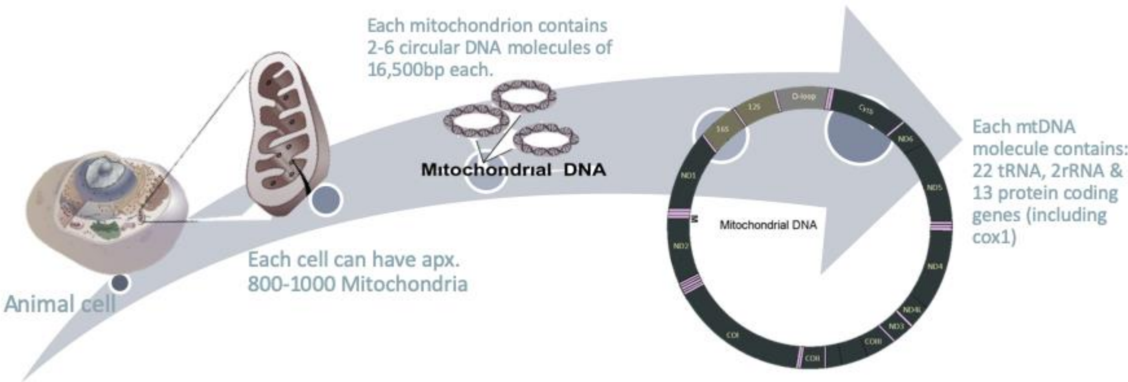


Figure 3. Mitochondrial DNA location.

It is true that mitochondrial DNA (mtDNA) was chosen as source of markers for barcoding animal species because of its features. Of the protein-coding genes present on the animal mitochondrial genome, the use of cytochrome c oxidase subunit 1 (*cox1* or *COI*) was proposed since the beginning by Hebert et al. (2003) as a standard barcode marker as previously mentioned. However, other mitochondrial genes too have been suggested as barcode markers, as reported in **Table 3**, including *cob*, which encodes for apocytochrome b, *cox2* and *cox3*, which encode for the cytochrome oxidase subunits 2 and 3, respectively, *nad1*, which encodes for NADH dehydrogenase subunit 1 and the mitochondrial 16S-rDNA gene.

Concerning chloroplast markers, establishing a standardized DNA barcoding system for plants was more challenging since their mitochondrial DNA has slower substitution rates than in metazoans, and shows intra-molecular recombination (Mower et. al., 2007), therefore limiting its resolution in identification on plants. The research for an analogous of *cox1* in terrestrial plants has focused on the plastid genome. However, limitations exist for plastid markers that usually present lower sequence variability compared to *cox1*. The proposed barcodes involved distinct combinations of seven plastid markers both from coding and non-coding regions: rpoC1+ rpoB + matK, rpoC1+matK+trnH–psbA, rbcL+trnH–psbA and atpF–H+psbKI+matK. The Consortium for the Barcode of Life (CBOL) Plant Working group recommends a core barcode consisting of portions of two plastid coding regions, rbcL+matK (Hollingsworth et. al. 2009), to be supplemented, if needed, with additional markers as required. In fact, it is well known that a two-marker plastid barcode system enables better discrimination than single barcode markers and no other multi-marker plastid barcode permits appreciably greater species resolution than the rbcL+matK combination, based on the straightforward recovery of the rbcL region and the discriminatory power of the matK region (CBOL 2009). Basically (rbcL, matK, and trnH-psbA) are used as the standard DNA barcode locus for plants. Besides, the plastidal markers, one nuclear Internal Transcribed Spacer (ITS) gene region have become the standard barcode of choice in most applications for plants and fungi (CBOL 2009, Schoch 2019), although, the nuclear ribosomal RNA gene D1/D2 region of the Large Subunit (LSU) has also been reported to be used in fungi identification. Characteristics of each marker can be found on **Table 3**.

The matK marker is one of the most rapidly evolving coding sections of the plastid genome and it is perhaps the closest plant analogue to the *cox1* animal barcode. Unfortunately, matK can be difficult to amplify using existing primer sets, particularly in non-angiosperms. In contrast, the barcode region of rbcL is easy to amplify, sequence and align with the reference sequences of most land plants and provides a useful backbone to the barcode dataset, even if it has only modest discriminatory power. Despite the high universality in terms of PCR amplification and DNA amplicon sequencing success for matK and rbcL, the analysis of these coding regions often fails due to the interspecific sharing of sequences. However, the combination of matK and rbcL with the plastid intergenic spacer region trnH-psbA increases the identification performance of DNA barcoding. It is easy to amplify this region in land plants and it shows high variability across intergenic spacers, even among closely related taxa.

Phyla	Target genome	Gene region	Coding for	Highlights	References
Animals	Mitochondrial Markers	16S-rDNA gene	Ribosomal genes	Their use in different taxonomic analyses is constrained by the prevalence of insertions and deletions (indels) that greatly complicate sequence alignments	Doyle and Gaut 2000
	Mitochondrial Markers	*cox1* (or *COI*)	Protein-coding genes	Easy to amplify, and its nucleotide substitution rate allows high power of discrimination; it is considered the universal DNA barcode for metazoans, even for closely related taxa.	Hebert, et al., 2003, 2004. Uthicke et. al 2010, Wong et. al. 2011.
		Cob, cox2, *cox3*, *nad1*, 16S-rDNA gene	Other protein-coding genes	Changes in its amino-acid sequence occur faster than those in *cox1* allowing higher taxonomic group discrimination	Nicolè et. al. 2011, 2012, 2013
Plants	Chloroplast Markers	rbcL + matK	Plastidial Genes	The CBOL Plant Working Group suggested the use of 2-locus combination of rbcL and matK as core-barcode regions, because of the easy amplification of rbcL and the high resolution of matK.	Hollingsworth et al., 2009
		trnH-psbA	Plastid intergenic spacers	It is easy to amplify and has a high genetic variability among closely related taxa. In combination of matK, rbcL and trnH- psbA increases the identification performance of DNA barcoding.	Bruni et al., 2010; Kress et al., 2010; Shaw et al., 2007
		rpoB, rpoC1	Other Plastidial genes	Most conserved and fast evolution rate (along with rbcL and matK)	Shaw et. al., 2007
		atpF-atpH and psbK-psbI	Other Intergenic spacers	Tested, because of their fast evolution rate (suggested along with trnH-psbA)	Fazekas et al., 2008, 2009.
	Nuclear Markers	ITS region	Nuclear ribosomal DNA	Indicated as a supplementary DNA barcode region.	Li et al., 2011
Fungi	Nuclear Markers	ITS region	Nuclear ribosomal RNA	Used not as a supplementary, but as an independent región (inlcuding ITS2).	Schoch et. al., 2012.
		LSU region	Nuclear ribosomal RNA	Tends to be more efficient when placing an unknown sequence in a known context than the ITS	Sonnenberb et. al., 2007

Table 3. Target DNA Barcoding genomic regions per phyla.

2.4. Standarization of the procedures

This characteristic refers to the extension of the approach to wide groups of organisms, not so closely related, for which it is possible to use the same method, that is a highly standardized procedure from sample collection to the analysis of molecular outputs. This ensures reliability.

If a strong standardization that characterizes protocols are applied worldwide for DNA barcoding, then that makes this technology particularly suitable for routine analyses required by agencies to safeguard food safety and quality during all traceability checking points.

As for all the methods, DNA barcoding can be fallacious. However, failures are mainly related to the essence of biological species, the patterns of molecular evolution, the completeness of sampling, the hybridization events and the heteroplasmy of sequences from different tissues rather than to the method itself (Hurst et Al., 2005).

Currently, molecular techniques based on DNA barcoding seem to be the most reliable and standardizable tool for authentication of food products of plant and animal origin, and researchers have developed large genetic datasets and obtained significant technical improvements in the last few years. The effective applicability of DNA barcoding is not a major problem since it is a relatively simple and cheap technology, being therefore not only sustainable if these analyses are meant to be employed as routine tests in all foodstuffs, but also affordable when they are needed just as verification in particular cases (e.g. required recall of a batch) or authentication of specific products (e.g. a PDO processed product)

The standard method used for DNA Barcoding might have some variability depending on the origin of samples (animal, insect, invertebrates, reptiles, fishes, birds, mammals, fungi, algaes, land plants, etc). The steps of collecting and storing the handled tissues and DNA samples have major effects on variability, while, subsequent steps, such as sequencing, data processing, and other back-end functions vary slightly. The procedure reported in **Figure 3** refers to DNA barcoding protocols described in Lee et al. (2012); they have worked closely with CBOL.

Figure 3. DNA Barcoding process steps.

Basically, DNA Barcoding exploits the polymorphism information content of specific DNA barcodes known to be highly variable among distinct species, that remains unchanged during all life cycle in any organism. Once the genomic DNA is extracted from the chosen matrix (e.g., animal tissue, blood, etc.) it is analyzed by single-locus PCR amplifications followed by direct sequencing of targeted DNA amplicons to obtain barcodes that allow the identification of the species.

Hereafter some important considerations for the method design (**Figure 3**):

1. *Sample collection:* collection of tissues should address the "genetic lockdown" that refers to stabilizing and securing the specimens, tissues, and DNA extracts for future genetic work as early as possible in the process chain and keeping them stable, secure, and safe from that point on. This will require different procedures in different collecting circumstances: preserving an entire specimen (or environmental sample) in such a way as to enable downstream DNA applications; or rough sorting and tissue subsampling in the field; or taking along an automated DNA extractor for on-site DNA extractions.

   The primary goal should always be preserving the integrity of the DNA and trying to maintain a high-quality voucher specimen. One without the other loses significant value.
   As soon as possible, after collection (and potentially the death of the organism), it is needed to carry out the subsampling of tissues to stop the degradation process.

   A photo documentation is also produced for groups where it is necessary; living color patterns or morphology should be captured prior to tissue subsampling, if that will decrease the value of the image. However, some methods can degrade the DNA, so care should be taken to preserve the integrity of the genetic material.

2. *DNA extraction*: can be performed with many specific protocols, guiding the choice in relation to strengths and weaknesses of the different approaches. The goal is to obtain a high-quality, high molecular weight, archival-quality DNA extract free of secondary compounds and other PCR inhibitors.

   There are several alternative methods for DNA extraction that yield a quality product from multiple sources and taxa. It is advisable to ensure, via preliminary experimentation on a few samples, that the method work prior to destroying the tissues of new taxa for all the specimens on hand.

3. *PCR for DNA barcoding*: it follows very similar procedures for most animal and plants groups. The main aspects are primers selection; groups of specific primers enhancing the success of the barcoding of different individuals are already available.

   Once obtained a successful PCR product (a clean, single-band of the target size), purification of the PCR product is carried out prior to a cycle sequencing and subsequent reaction purification.

4. *Sequencing:* the purified PCR product is then used for sequencing by using an automated sequencer.

5. *Processing and quality control of the raw sequence data*.

6. *Taxonomy analysis*: two basic steps are followed: (1) building the DNA barcode library of known species; it requires taxonomic expertise in order to select one or several individuals per species to serve as reference samples in the barcode library; (2) matching the barcode sequence of the unknown sample against the barcode library for species identification by using specific algorithms; the most used one is an algorithm that compares two DNA sequences to produce a distance measure between these two sequences. Different distance measures can be used, such as Basic Local Alignment Search Tool (BLAST) or K2P parameter or Smith-Waterman algorithm for local alignment similarity

7. *Submission to databases and repositories:* it is even possible to create and make the Project public through the BOLD platform (https://ibol.org/) by creating an account (Erickson et al., 2008).

2.5. Computarization

Computerization is the not redundant transposition of the data using informatics. Bioinformatics plays a key role in supporting and consolidating DNA barcoding, being crucial to select the right primers, to evaluate the sequence quality and to analyse data.
Hebert et al. (2003) proposed the development of a database that could serve as the basis for a Global Bioidentification System, named as "GBS". While genbanks already existent aim to a comprehensive coverage of genomic diversity, Hebert et al. (2003) proposed that the GBS database was supposed to represents a comprehensive taxonomic coverage of just a single target gene section. In 2007, Ratnasingham and Hebert, published an article describing how they launched the Consortium for the Barcode of Life (CBOL) with the cooperation of more than 120 organizations from 45 nations. The CBOL promoted the development of the GBS database. To do this, CBOL started a dialogue with the major genomics repositories (including genbank from National Center for Biotechnology Information (NCBI), and other biodiversity organizations, such as, for example, the Global Biodiversity Informaion Facility (GBIF). They established the formal guidelines that must be met for records to gain barcode designation. Thus, the GBS database was developed and called "BOLD", The Barcode of Life Data System (BOLD; http://www.boldsystems.org/). It provides an integrated bioinformatics platform that supports all phases of the analytical pathway from specimen collection to tightly validated barcode library. First, it is a repository for the specimen and sequence records that form the basic data unit of all barcode studies. Second, it is a workbench that aids the management, quality assurance and analysis of barcode data. Third, it provides a vehicle for collaboration across geographically dispersed research communities by coupling flexible security and data entry features with web-based delivery. Is possible to login even for free and access this database. Accessed on 25 November 2021, the BOLD Database counts more than 10 million barcodes collected (**Figure 4). Figure 5** shows all worldwide sampling points related to barcodes present in the BOLD repository.

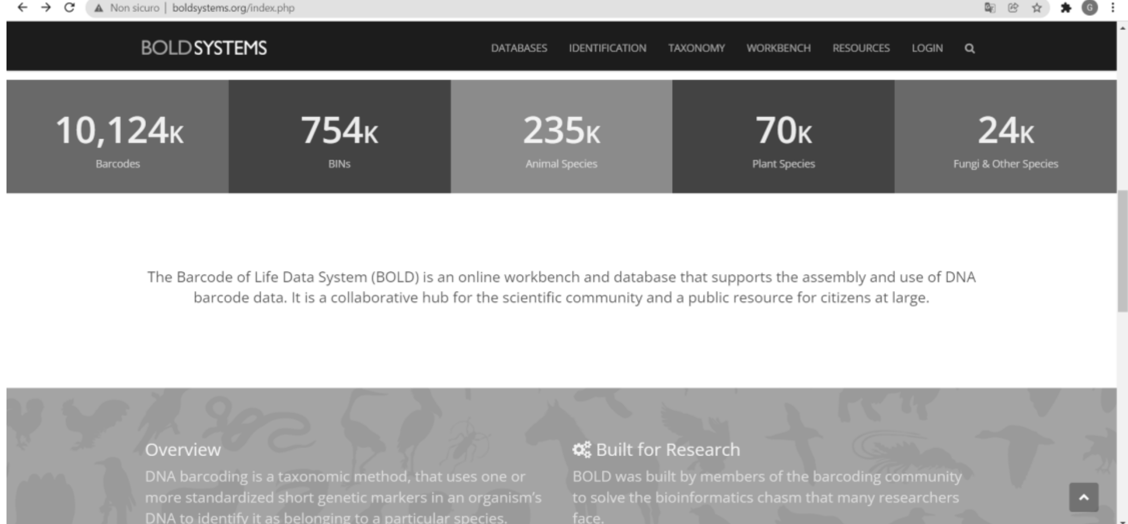


Figure 4. BOLD Website (accessed on 25 November 2021); Barcode Index Number (BIN) System clusters sequences using well established algorithms to produce operational taxonomic units that closely correspond to species.

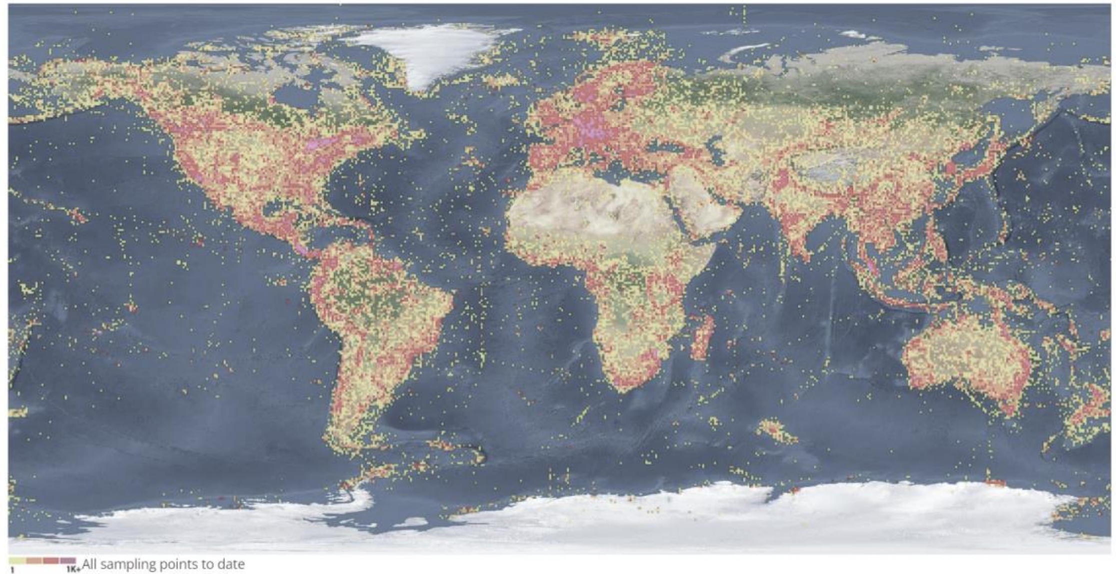


Figure 5. BOLD Sampling Points (accessed on 25 November 2021)

The BOLD Database is still in constant evolution and update. The amount of data managed by BOLD is impressive: it collects, for a large amount of deposited barcode sequences, specimen's details such as morphology, photographs, geographical distribution, collection points and more.

Computerization also refers to the bioinformatics tools to analyse sequence data. The aim of DNA barcoding analyses is simple, that is to assign each query sequence to a set of referenced (tagged-specimen) sequences. The BOLD database is commonly used as the sequence resource for comparision, but, then, there is another step, that is choosing the best, among the different bioinformatic approaches, to assign the correct taxa.

Table 4 shows the classification of the main cases a DNA barcoding user can face, and the possible bioinformatics methods that can be used, since for each of them pros and cons are present.

Is relevant to note that there is no analytical method outperforming the others, but the 'best method' is case related. In such a dynamic and fluid situation it is necessary that users get more and more acquainted with the bioinformatics of DNA barcoding. At the same time, users should learn how to properly manage data, to avoid errors and incorrect interpretation of the results.

Tipology	Method (s)	Software/tool(s)	Resources
Threshold (distance)	Similarity	Blastall-BLASTn	ftp://ftp.ncbi.nih.gov/blast/
	Similarity	BLAT	http://genome-test.cse.ucsc.edu/ kent/exe/
	Similarity	Blastall-megaBLAST	ftp://ftp.ncbi.nih.gov/blast/
	Pairwise distance	TaxI	axel.meyer@uni-konstanz.de
	Pairwise distance	TaxonDNA	http://taxondna.sf.net/
	K2P distance	MUSCLE, MEGA	maurizio.casiraghi@unimib.it
	K2P distance	BOLD-IDS	http://www.barcodinglife.org/ views/idrequest.php
	Patristic distance	MrBayes, PAUP, APE, Perl scripts	lefebure@univlyon1.fr
Phylogenetic	Neighbour Joining	MUSCLE, MEGA	marianne.elias@ed.ac.uk
	Parsimony	MUSCLE, TNT	dlittle@nybg.org
	Maximum likelihood	MUSCLE, SPR1, PHYML2	http://atgc.lirmm.fr/spr/
	Bayesian inference	SAP	http://fisher.berkeley.edu/cteg/ software/munch
	Coalescent based	-	rasmus@binf.ku.dk
	Coalescent based	-	rasmus@binf.ku.dk
	Coalescent based	COALESCENCE, FLUCTUATE, PAUP, Seq-Gen	golding@mcmaster.ca
	Coalescent based	COAL, MESQUITE	knowlesl@umich.edu
	Coalescent based	General Mixed Yule-Coalescent (GMYC) Model	monaghan@igb-berlin.de
Character based	Diagnostic	CAOS	http://www.genomecurator.org/ CAOS/CAOSindex.html
	Diagnostic	MATLAB, local perl scripts	drichardson@rsmas.miami.edu
	Diagnostic	DNA-BAR (degenbar)	http://dna.engr.uconn.edu/software/DNA-BAR/
	Diagnostic	DOME ID (local perl scripts)	dlittle@nybg.org
Combined	Yule model/coalescence	TCS, MEGA, Arlequin, PAUP, PAUPRat script, Phylip, r8s, R	http://www.imedea.uib.es/jpons/JPWPhome.htm
	BLAST/parsimony ratchet	BLAST, MUSCLE, TNT	dlittle@nybg.org
	BLAST/SPR	BLAST, MUSCLE, SPR	dlittle@nybg.org
	BLAST/Neighbour Joining	BLAST, MUSCLE, neighbour	dlittle@nybg.org
	Tree-based	ATIM: TNT, local scripts	dlittle@nybg.org

Alignment-free	Component vector	CVTree alpha 1.0	http://cvtree.cbi.pku.edu.cn
	Spectrum kernel method	Spectrum	vladimir@cs.rutgers.edu
Web tool	-	Web browser	http://www.ibarcode.org
		Web browser	http://www.dnabarcodelinker.com/
		Web browser	http://www.asianbarcode.org/
Other	-	ConFind, Python	http://www.colorado.edu/ chemistry/RGHP/software/

Table 4. Summary of the bioinformatics methods useable to analyse sequences. Source: Casiraghi et al., 2010.

The most used method is *similarity*; it means to compare obtained barcode sequences to a reference dataset. Methods based on similarity, follow a typological species concept, and discriminate entities exceeding a certain level of variability called threshold value. The main assumption is that intra-specific sequences variation does not exceed a certain distance value, otherwise they are considered as different species. In general, these methods are faster and require low knowledge on population structure or phylogenetic relationships. They can be considered the ‘first choice’ for new users approaching DNA barcoding. However, these methods imply the existence of a reference dataset, generated with the coordinated work of traditional and molecular taxonomists, to work.

To answer the question of how different the genomes need to be in order to be identified as a different species, is explained by the concept called “barcoding gap”, refering to the degree of sequence divergence between two samples above a given threshold would indicate specific distinctness, whereas divergence below such a threshold would indicate taxonomic identity among the sample. Hebert et al. (2003), firstly proposed the use of a divergence threshold following the ‘10-fold rule’, where the gap corresponds to a generic 10 times the value of intra-specific divergence. The main assumption is that DNA sequences are likely more like one another within species than between species as variation in the nucleotide composition is lower within one species than between different ones. The system aims to detect polymorphisms, so the DNA barcode sequence contains enough unique information, in terms of SNPs (single nucleotide polymorphisms) and insertion/deletions, shared among individuals of a species with slight variations, but specifically associated to one species with a unique haplotype. For assessing species identity and genetic traceability purposes, the DNA barcoding gap (the difference between intra- and inter-specific divergence values) is a necessary condition for intraspecific variation (belonging to the same species) and interspecific differentiation (belonging to different species) of the selected markers. This gap would even enable the identification of previously undescribed species.

Hebert et al. (2005) used the similarity approach to investigate the relationships among 207 fish species by computing the average within-species, genus, family, order and class Kimura two parameter (K2P) distances (**Figure 6** and **Table 5**). Their results showed that the average K2P distance of individuals within species was 0.39%, while for species within genera it was 9.93% (**Table 5**); moreover, average divergence among species within families increases to 15.5%, and among species within orders and classes it increases to 22.2% and 23.3%, respectively (**Table 5** and **Figure 6**).

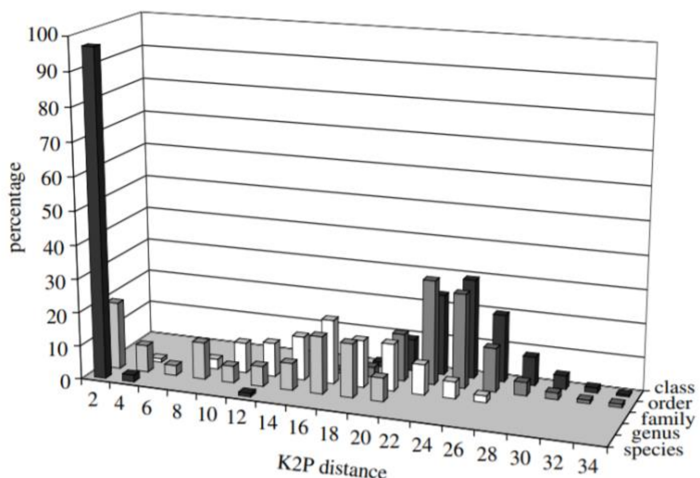


Figure 6. Distribution of K2P distance (>1%) for cox1 within different taxonomic categories. Source: Hebert et al. (2005)

Comparisons within	Number of comparisons	Minimum	Distance mean	Maximum	s.e.
species	1315	0	0.39	14.08[b]	0.031
genera	4259	0[a]	9.93	20.63	0.096
families	9479	1.39	15.46	35.72	0.049
orders	68083	9.55	22.18	37.52	0.012
classes	83265	14.33	23.27	37.39	0.009

Table 5. Summary of genetic K2P divergences (%) within different taxonomic levels. Data are from 754 sequences from 207 species and 122 genera; [a], one example in *Pristiophorus* and one in *Plectropomus*, where in each instance one sequence among multiple specimens appeared to be of a different, but congeneric, species; [b], *Hydrolagus novazaelandiae*. Source: Hebert et al. (2005)

The BOLD system uses a threshold approach, allowing a Web user to perform species identification by querying, with an appropriate sequence, the BOLD database. This tool is called Identification System Engine (BOLD-IDS). BOLD-IDS is actually based on similarity methods and distance tree reconstruction (K2P & Neighbor Joining as designed by Hebert et al. 2003). BOLD uses 1% of K2P distance as universal threshold value for metazoans discriminations. The query sequence is assigned to the species name of its nearest- neighbouring referenced sequence. BOLD-IDS is continuously upgraded and deals with the challenge to implement the better algorithms developed by the scientists.

The majority of the published works following DNA Barcoding approach also perform a simple distance matrix analysis, using a Neighbour Joining (NJ) algorithm, with a K2P correction. Since, the K2P was claimed as the best DNA substitution model for low genetic distances, as reported by Nei and Kumar (2000).

However, some literature argues that pure distance-based methods could not be the most appropriate for species identification (Ferguson 2002), due to several aspects, such as 'enchainment on the percent divergence' (Little and Stevenson 2007), the lack of a strong biological support (Meyer and Paulay 2005), the loss of character information (De Salle

2006), the deep influence by incomplete taxonomic sampling both at species and intra-specific level (Meyer and Paulay 2005) and by the chosen parameters for sequence alignment (Prendini 2005).

Similarity methods, such as BLAST, BLAT or FASTA, were largely used to infer similarity between a query sequence and barcode reference sequences. These methods deal with unaligned sequences in the reference database and use partial pairwise alignment or nearly exact matches of short strings (motifs) and are typically very fast in giving answers. In case of users dealing with very large datasets and if no higher precision is needed, similarity-based methods help to rapidly analyse the datasets. Little and Stevenson (2007) compared the performances of clustering, diagnostic and combined methods against similarity-based method on two gymnosperm datasets regarding a coding gene (matK) and a non-coding gene (ITS2). The results showed that the better values of accuracy to genus and species-level identification were reached with BLAST for the coding gene.

Phylogenetic typology of methods follow a phylogenetic species concept, they can be applied to datasets relative to groups that experienced different evolutive histories (Casiraghi, 2010). In general, they are time consuming, because of high computational effort, and are directed to users acknowledged on the phylogenetic reconstruction techniques. These methods have been developed and proposed for DNA barcoding data analysis to overcome the limits of threshold-based approaches (Vogler et. al.,2006). Moreover, the publication of trees as the only output of DNA barcoding-tagged papers contributed to enhance and spread criticisms on the technique. It is worth remembering that DNA barcoding is not, in a strict sense, a phylogenetic reconstruction, but the aim of DNA barcoding is to identify the taxa. Either way, these methods can be used as a complementary tool.

A workflow of data from sequence to a phylogeny tree is exemplified on **Figure 7**, reporting the softwares used at each step (Kress and Erickson 2012).

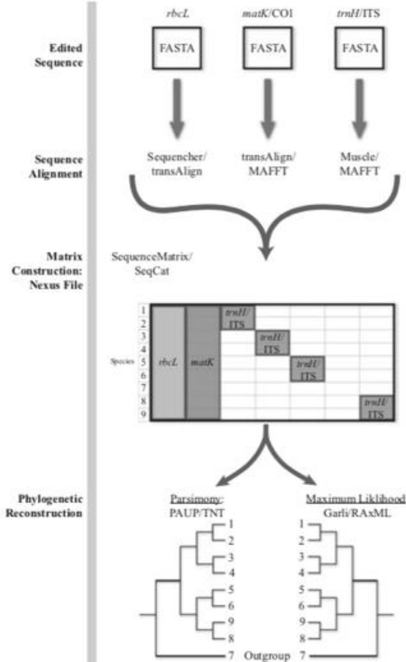


Figure 7. Workflow from sequence to phylogenetics. Source: Kress and Erickson 2012.

When species are identified through character states (meaning the presence/absence of discrete nucleotide substitutions) DNA barcoding implements character-based methods. Differently from distance-based and classical phylogenetic-based approaches, character-based methods rely only on diagnostic sites, that being a small percentage of the total characters, make the application typically faster (Casiraghi, 2010). These methods are considered consistent with the phylogenetic species concept and can also handle other sources of data, such as morphological or ecological data (Goldstain et. al., 2003). This integration leads to a synergy that has the advantage of minimizing the discrepancies between classical taxonomy and DNA barcoding. Character-based methods sidestep the distance 'nearest neighbour problem' by reconstructing hierarchical relationships (meaning that the common ancestor is inferred when two entities share derived characters). In addition, character-based methods are probably the best choice in case of datasets with few sequences for each taxonomic group (Casiraghi, 2010). Users working on organisms for which sampling is difficult and limiting, should consider these methods as a valid choice.

Chapter 3. DNA BARCODING TO IDENTIFY AND CERTIFY FOOD RAW MATERIAL

Food authenticity, traceability, origin and provenance are major concerns for consumers, industries and regulatory bodies worldwide. As demand and consumer awareness increases about food safety issues, traceability systems become an essential component of safety and quality management systems as an intrinsic part of food quality. Traceability systems are used to collect data and supply relevant information on the production chain at set control points. The more precise the traceability system, the more efficiently it can identify and resolve food safety or quality problems in a given food production chain.

New trends in agricultural practices and the recent sanitary emergencies have made producers and consumers more demanding about food authenticity, with a concurrently increasing interest for food origin, healthiness and nutraceutical properties. At present, consumers are more aware than they were years ago of ecological and environmental matters and the demand for organic food and for products obtained in an eco-sustainable system has increased, nevertheless the industrialization processes, as well as market globalization, have made it difficult for people to know the real origin and composition of their food.

Consumers have the right to access accurate and complete information regarding the products they buy and eat. In many cases, food, both raw ingredients and products are deliberately or unintentionally adulterated. This can be a consequence of either the mislabelling of a product or the substitution of one component with a similar but lower-quality or cheaper counterpart, ultimately leading to commercial fraud. Consequently, assuring the authenticity of food ingredients and products is critical for preventing not only economic fraud but also to reduce the negative impact on both consumers and industry stakeholders.

DNA barcoding might be the answer to the consumers' demand of transparency and treaceability of their food products, and it is becoming synonymous with safe and high-quality food.

Applications of DNA barcoding to food safety and food piracy issues have grown in importance due to the consumers' increasing attention to food authenticity and food safety in different products. This interest seems to be due to the consumers' lack of confidence and it is attributable to several reasons including both food safety and socio-economic changes. Bovine spongiform encephalopathy (BSE) has certainly been the most serious food safety problem of recent years, causing a drastic reduction of beef consumption in all Europe. It was then followed by the dioxin crisis and the avian influenza in the poultry sector, in fact that safety problem led to enforced food legislations.

3.1. DNA in food raw materials

DNA barcoding can be used on both fresh and raw materials for species authentication, species delimitation, and identification from different individual parts where other methods of characterization usually fail. Moreover, DNA barcoding can be applied to distinct food products and matrices deriving from single or mixed species. In this way this methodology could be used to discover voluntary or accidental replacements associated with food. (Barcaccia et. al., 2015).

Often a risk factor in identifying food mislabeling or piracy is the lack of adequate standards and the availability of high-quality repositories of reference sequences can be a critical point. In this regard, having a robust and accurate reference library and using computerization is already covered by DNA Barcoding Techonology.

DNA barcoding is a very robust diagnosis tool due to a number of features, including the evidence that DNA is inalterable, detectable in every cell, and resistant to heat treatments, and allows for individual, breed/cultivar or species identification for traceability purposes.

The experimental procedure of extracting genomic DNA and amplifying specific DNA markers is technically easy and usually does not require the destruction of the sample, which sometime needs to be safeguarded for further uses or inspections. Moreover, it allows the treatment of all kinds of biological specimens, including those non-identifiable by morphology, and it is also very fast and relatively inexpensive compared with other molecular approaches (Barcaccia et. al., 2015).

The need of traceability is present in both animal and plant foodstuff. The last 20 years have seen a considerable investment into research targeting food quality and safety methods, as consequence of the number of food scandals, which have seriously undermined consumer confidence. This has been repeatedly proven true for example in the case of fish and seafood products; they are one of the most common food commodities traded, and thus also one of those most prone to mislabelling and fraud at an international level. This is because the source (aquaculture, wild, fresh water, etc.) and identification of fish and seafood products (species, populations linked to geographic origin) are highly complex. The wide biological diversity, the removal of external features during processing steps, as well as the close phylogenetical relationships among them render the morphological identification almost impossible. This is reflected in the literature where the substitution of a high-quality seafood products with one of lower cost is frequently reported by both industry and academics (Hellberg et. al., 2011). Therefore, a big focus has been placed for DNA level identification. Not only fish has this fraud problem, but

many animal-based and also plant-based food products. Hence, the importance of applying the proper legislations for food safety and tracebility.

3.2. Legislation

Despite the implementation of strict regulations and labelling protocols by various nations, these issues remain an international concern. In fact, as a means to avoid food adulteration and food fraud, many regulations have been promulgated all over the world for food products, both general and specific for seafood products. For instance, the Council Regulation (EC-No. 104/2000) on the common organization of the markets in fishery and aquaculture products requires that seafood products must be labelled with the commercial name of the species at every step of the marketing chain.

Traceability is generally imposed by international and national control organizations for all types of foods. In recent years, all countries have updated their regulations on food traceability and labeling. The guidance document on the best practices in food traceability presents an overview of the international regulations. It presents a best practice guidance for food traceability applied to different food sectors such as bakery, dairy, meat and poultry, processed foods and seafood.

In January 2011, the US government published the Food Safety Modernization Act, FDA, (111th Congress Public Law 353) with the aim to improve capacity to prevent and detect food safety problems. In December 2014, the Chinese National People's Congress published the Second Draft of its Food Safety Law. In Japan, the legislation, allows the consumer to access the information on the source of the beef and the methods used to rear the animal. In Canada, Australia and New Zealand a trace back system based on tagging was established in 2001. In Brazil and Argentina traceability systems are in use though with different depth.

As regards to Europe, is imporatnt to mention its relevance in the food industry since the European Union (EU) is the biggest producer of foodstuffs and beverages in the world with a food- and beverage-related industries production of 16% of the total EU manufacturing turnover, corresponding to more than 950 billion Euros. Therefore, the EU has always paid great attention to food safety, especially because the agro-alimentary sector on its whole is very important for European economy.

After the creation of the Green Paper, and then the White Paper, EU implemented officially since 2005 the European General Food Law Regulation (EC 178/2002), which lays down the general principles and requirements of food law, establishes the European Food Safety Authority and lays down procedures in matters of food safety. This regulation stressed the importance of a traceability system, declaring that "the experience demonstrated how the impossibility to reconstruct the trail of a food could be a danger for the market of such product", while a traceability system able to keep all the information regarding food production can help to proceed to its recall in case of danger without damaging the entire sector. Therefore, since 2005, the regulation has become mandatory for all member countries which must define a traceability system for the whole food sector. In addition, it permits agreement among the different member countries' legislations in which several differences were present, leading to problems regarding the free exchange of food among them. The EC 178/2002 is the fundamental law regarding food safety, it has been followed by several other regulations; for animal products the

most important are 852/2004, 853/2004, 854/2004 and 882/2004, and all of them corroborate the importance of a traceability system and the need for authorities to control them.

EC 178/2002 aims to ensure three main objectives: *security* (meaning the ability to handle any problems or medical emergencies), *consumer confidence* and *control by the public authorities*. In particular, article 18 of EC 178/2002 defines scope and implementation of food traceability. It defines traceability as the ability to trace and follow food, feed and ingredients through all stages of production, processing and distribution.

As indicated in the EC 178/2002, traceability:

1. facilitates withdrawal of faulty food and feed from the market;
2. provides consumers with targeted and accurate information on specific products;
3. covers all food and feed, all food and feed business operators; and
4. affects importers who are required to be able to identify from whom the product was exported in the country of origin.

In January 2010, the EU Standing Committee on the Food Chain and Animal Health approved the new version of the guidance document.

Then, in December 2014, the EU's new "Food Information to Consumers (FIC)" regulation 1169/2011, which defines the food labeling requirements, became applicable to all pre-packaged food and drink products marketed in the EU, including those imported from outside countries. The mandatory nutrition declaration requirement, introduced by the new FIC regulation, has been applied in December 2016.

Adulteration of foods for financial gain is prohibited. It consists in omitting any valuable constituent, substituting another component wholly or partly, or adding any substance to increase weight or bulk, worsen quality or make a product appear better than it is. Most adulterants are benign, but less expensive than some constituents of the food, and the counterfeiting products are actually food frauds, economically motivated misbranding and mislabeling, fakes based on simulation processes and imitation products. When the adulterants are toxic or allergenic, serious public health consequences may result. In this case, the food mislabeling not only robs consumers of value, but it may also endanger people who have intolerance or allergies to certain foods or their components.

The most frequent incidents, based on the literature from 1980, were grouped into 11 food categories: fish and seafood, dairy products, fruit juices, oils and fats, grain products, honey and other natural sweeteners, spices and extracts, wine and other alcoholic beverages, infant formula, plant-based proteins, and other food products (Everstine, 2013). For processed food, for which the morphological characteristics of the species are removed, there is a need for inexpensive and widely available genetic testing methods.

At present, molecular techniques based on DNA barcoding seem to be the most reliable and standardizable tool of authentication for food products of plant and animal origin, and researchers have developed large genetic datasets and obtained significant technical improvements in the last few years.

3.3. Complementary methodologies

Some other terms have emerged complementing the DNA barcoding methodology along with other technologies, with the goal to make it more efficient in a more complex scenario or certain highly specific applications. Emerging terms relating to DNA Barcoding include:

DNA Metabarcoding

This method is applied when the aim is to simultaneously identify multiple species present in the same food sample by using High-Throughput DNA Sequencing (HTS) that offer the opportunity to analyze multiple DNA amplicons by sequencing them in parallel.(Galimberti, 2015) This makes metabarcoding ideal for the application on the characterization of complex food matrices and ultra-processed foods. It is called meta because it uses amplicon-based strategy. **Figure 8** shows the differences between DNA barcoding and DNA metabarcoding.

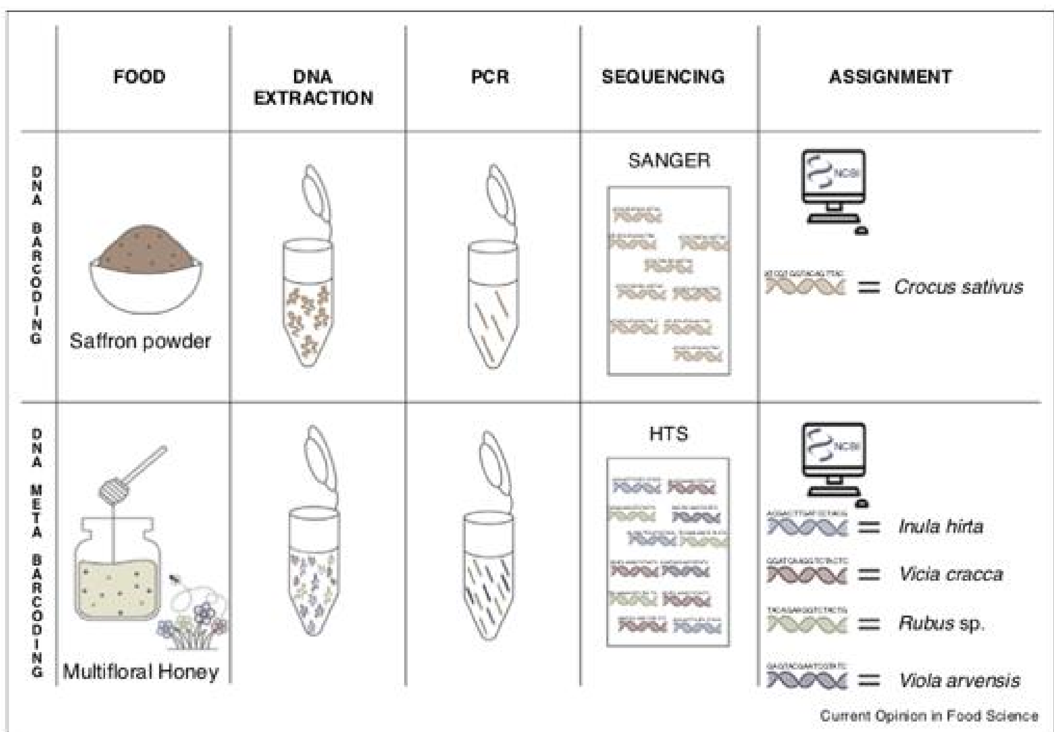


Figure 8. DNA Barcoding vs DNA Metabarcoding. Source: Galimberti et. al., 2019.
DNA barcoding and metabarcoding pipelines. Food products can be pure (e.g. saffron) or mixtures of different species (e.g. multifloral honey). For DNA barcoding the DNA from each food product is extracted separately, amplified by PCR at specific loci (e.g. ITS2) and Sanger sequenced, to produce barcode sequences which are assigned to a species by comparing them against a reference database (e.g. NCBI GenBank). In DNA metabarcoding, the whole food product is homogenized and DNA is extracted directly from this. PCR is performed on the DNA extract with primers usually designed for a shorter read length to comply with the chemical and technical features of High Throughput Sequencing (HTS) instruments. Once HTS of the PCR products is performed, bioinformatics pipelines are used to extract unique sequence read and to assign them to species using again a DNA barcode reference database.

DNA minibarcoding

DNA minibarcoding refers to the use of smaller genome portions (100-300bp), usually associated with larger DNA barcodes (Hajibabaei , 2006) citation). Conventional barcoding methodology is often limited by its failure to amplify and sequence degraded

DNA, which is often found in museum specimens and in preserved and processed biological material (food products, decayed tissues). Approaches aimed to repair DNA *in vitro* are inefficient and not cost-effective as the DNA damage and degradation in museum samples is complex and difficult to characterize. By contrast, short sequences (i.e., 100 bp) are usually stable in museum specimens. The use of a short or minimalist barcode (100–300 bp, referred herein as "mini-barcode") greatly expands the applications of DNA barcoding (Hajibabaei , 2006) (**Table 6**).

	Full-lenght barcode (650bp)	**Mini-barcode (100-300 bp)**
Specimen sequence success relative to age	>90%, 5-10 years	>90%, up to 200 years
Species resolution	95-98%	91-95%
Techonology	Sanger (ABI)	Sanger (ABI) NexGen sequencing (i.e. 454) Single pyrosequencing (PSQ)
Applicability	Barcode library construction Rutine barcoding	Museum and preserved samples Processed materials (i.e. food products, pharmaceuticals) Environmental barcoding

Table 6. Comparison of full-lenght DNA barcode and mini-barcode. Source: Hajibabaei and McKenna 2012.

Next Generation Sequencing (NGS) coupled with DNA Barcoding

Next Generation Sequencing (NGS) refers to the set of technologies used for genome-scale sequencing, like Roche 454, Illumina, Ion Torrent, and others. These technologies allow to generate more sequence reads than traditional Sanger sequencing, to pool many samples and to amplify their DNA in parallel. Short DNA regions can be amplified by using NGS technologies, allowing comprehensive and cost-effective barcoding applications (**Figure 9**)

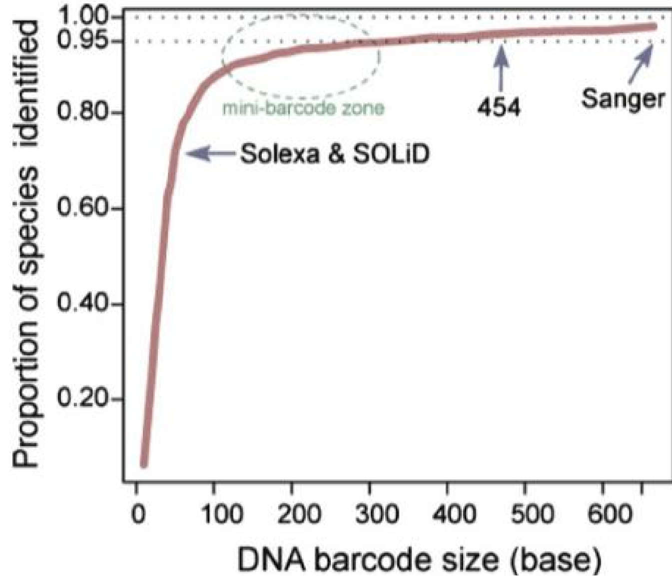


Figure 9. Comparison of DNA barcode size versus proportion of species identified reveals the efficiency of mini-barcodes in resolving species. Sequence read lengths typically obtained from three commonly used next-generation sequencing technologies as well as Sanger sequencing are shown on the graph. It is clear that 454 pyrosequencing and Sanger are currently optimal technologies for mini-barcode and full-barcode recovery. Source: Meusnier et al. 2008.

An effective tool in applying DNA barcoding to complex food matrices could be the 454-pyrosequencing methodology, which produces several hundreds of thousands of sequences per run, corresponding to the whole mix of DNA molecules extracted from the matrix, in that way it can discriminate a matrix containing a mixture of biological species. This approach allows to identify all raw materials, including contaminants, or elements occurring in traces only. Pyrosequencing was used for several DNA barcoding analyses (Hajibabaei et al., 2011; Valentini et al., 2009), including the identification of raw material of the diet of several animals (Raye et al., 2011; Soininen et al., 2009), as well as for analysing ancient DNA extracted from museum specimens (Shokralla et al., 2011).

DNA Ultra-barcoding

The limits of adopting universal barcode markers are evident at the cultivar level, where genetic variability is limited, and there are complications due to breeding events. To overcome these limits, Kane and Cronk (2008) proposed the ultra-barcoding methodology, which is based on the sequence of the whole plastidial genome, together with large portions of the nuclear genome. This combination provides enough information to evidence genetic diversity below the level of species, distinguishing hybrids from pure lines, hence it is far more sensitive than traditional DNA barcoding. Kane and Cronk (2008) evaluated the effectiveness of ultra-barcoding on cocoa (*Theobroma cacao* L.), and found several plastidial and nuclear SNPs, which were useful to identify different cultivars. This technique is promising, but it is difficult to apply on a large scale due to its high costs, and its excessive species-specificity. Furthermore, this approach is contrary to the basic DNA barcoding methodology, which requires the analysis of short and universal DNA regions only.

Bar-High Resolution Melting (HRM)

DNA barcoding is employed coupled with high resolution melting (HRM) analysis (Bar-HRM) (Jaakola et al. 2010, Madesis et al. 2012). HRM method measures the rate of double stranded DNA dissociation to single stranded DNA with increasing temperature (Reed & Wittwer, 2004). Bar-HMR consists in the amplification of a short DNA barcoding sequence and target region detection through HRM. In the last years, the Bar-HRM strategy has found a large spread in agri-food surveillance. Bosmali et al. (2017) set up a fast and cost-effective Bar-HRM method for PDO saffron authentication, other cases include tea products, and walnut milk (Ding et.al. 2020, Lagiotis et. al. 2020).

Nanotracer:

Developed by Valentini et al. (2017), *Nanotracer*is able to detect the presence of a specific species-DNA in a food sample through a colorimetric response. The proposed approach is based on an asymmetric PCR amplification of a short barcode region, yielding a single-strand amplicon that is readily hybridizable to induce a color change due to the presence of DNA-functionalized gold nanoparticles. This method offers a rapid and naked-eye authentication test, and its implementation in the agri-food sector will provide an efficient system for food surveillance in the future.

Microfluidic Enrichment (ME) Barcoding

Developed by Gostel et al. (2020), this method is a microfluidic enrichment barcoding (MEBarcoding) for high-throughput plant barcoding, a cost-effective method based on the combined use of the Fluidigm Access Array and Illumina MiSeq. They built a highly comprehensive barcode database and demonstrated that the proposed approach is

efficient in discriminating a very large number of species present in a food-borne matrix at the same time.

3.4. Advantages and limitations of DNA Barcoding

Relevant advantages and limits of DNA Barcoding in the agri-food sector for the purpose of traceability and authentification (compared to other molecular approaches) are reported in **Table 7**.

Advantages	Limitations
It requires the amplification of a very short DNA region (a few hundred base pairs)	Physical fragmentation and chemical treatment can affect the yield, integrity, and quality of DNA. To overcome this problem, several protocols for DNA extraction from processed agri-food matrices were developed with the aim to recover enough good-quality DNA for subsequent analysis (such as CTAB-based methods); these protocols were optimized to extract DNA from a specific food-borne product with the purpose of maximizing the yield while minimizing the co-extraction of enzymatic reaction inhibitors.
The widespread use of plastidial genome (cpDNA), which is more preserved during industrial processing, indeed heavily industrial treatments can severely affect nuclear DNA quality and quantity, while this occurs to a lesser extent with cpDNA due to its abundance	
DNA barcoding is more effective than use of SNPs, which being highly specific, require the knowledge of plant species putatively present in a food and access to the correct DNA sequence of interest. Therefore, their application is often limited to a single species.	Only the species for which a reference is available can be identified; therefore, database incompleteness greatly affects the reliability of analysis. To overcome this limit, it is needed to keep research ongoing to improve the barcode databases.
DNA barcoding is more effective than DNA fingerprint in plants, due to the fact it does not require the knowledge of the whole genome of an organism, being based on the exploitation of one or few genomic regions	It can only be applied to identify monophyletic species, since polyphyletic and paraphyletic species do not display a clear barcode gap (i.e., a gap between frequency distributions between intra- and interspecific distances). In these cases, a combined approach of molecular markers and DNA barcode would be the best strategy for an accurate and exhaustive authentication analysis.

Table 7. Advantages and limits of the use of DNA Barcoding for food taceability.

Chapter 4. DNA BARCODING APPLICATION ON EDIBLE PLANT-BASED PRODUCTS

Plants are an essential element in human diet, both directly (cereals are the base of the food pyramid, followed by fruits and vegetables) and indirectly (plant products are used to feed cattle). Furthermore, several plants are used as food additives and even herbal remedies (Seethapathy et al. 2015). A reliable identification of crop species, as well as their origin and traceability, are key elements in the field of food safety. Due to globalization, an increasing number of plants originating from different areas of the world are now offered to consumers, but there are not reliable, universal tools for their identification. DNA barcoding could be a reliable alternative to DNA fingerprinting approaches in plants identification, with a higher effectiveness/cost ratio.

An “ideal” traceability system would follow the “history” of a product from its origin to the moment it is used, considering all transformation and commercialization steps. Seeds, fruit, and different plant and animal parts are transformed in food with a definite shape, taste and smell through physical (for example heating, boiling, UV radiation) or chemical (for example addition of food preservatives, artificial sweeteners) treatments, that could alter the plants DNA structure. The application of DNA barcoding is usefull on plant-based products not only as a foresic application on the final product located in the market shelfs, but also along the whole process chain, from field, to processing facilities, to commercialization in order to verify the genetic identity of their raw materials that compose it and trace them to the source.

Table 8 reports some of the applications of DNA Barcoding techonolgy in plant-based food products. In many cases it can be found that the methodology has been coupled with other technologies.

Applications	Description
Juices and yogurts	DNA barcoding showed a high effectiveness in the checking of fruit-based processed products, particularly yogurts and juices. For instance, fruit juices labeled as 100% fruit represent approximately two-thirds or 10 billion liters of total EU juice sales in 2013 (European Fruit Juice Association (AIJN), 2014). The adulteration of fruit juices often consists in the addition of a cheaper fruit juice to increase the production profit. It has been demonstrated by Faria et al. (2013) the efficacy of DNA barcoding with *trnL* as a target sequence combined with HRM analysis for the complete fruit species discrimination (*i.e.*, orange, mango, peach, pear and pineapple) and their quantitative evaluation in fruit juices. Has been proven to be reliable even in highly processed products such as detecting the fruit species in yogurt (Knight et al. 2007) and fruit residues (like banana) in juices, purees, chocolates, cookies, etc. (Sakai et al., 2010).
Nut allergens	DNA barcoding can be applied to comply with the requirements of FAO and European Commission, which list allergenic species that must be declared on food labels (Directive 2003/89/EC.1).

	DNA Barcoding showed a high effectiveness in the evaluation of the presence of allergenic species, both in fresh and in processed food. Nuts are considered one of the main sources of allergens (Hubalkova and Rencova, 2011), and their presence in food (also in traces) is detectable by molecular analysis based on different markers, including DNA barcode regions (e.g., *matK*) (Yano et al., 2007). Almond (*Prunus dulcis*), commonly used in several food products (bakery, pastry, snacks) due to its pleasant flavor, is also a potential allergenic (Costa et al. 2012) Similar approaches could also be applied to food intolerance because of substances present in some genera or species, such as gluten for people with celiac disease (Maskova et. al., 2012). A proposed method by Madesis et al. (2012), was again based on the use of universal chloroplast primers for *trnL* marker amplification coupled with HRM analysis for DNA barcoding tree nut species and for quantification of their allergenic components in commercial foods. Such an approach was shown to be able not only to distinguish among the different nut species, but also to reveal a ratio of 0.01% of hazelnut contamination.
Flavouring Herbs	Spices are a numerous group of plant species characterized by aromatic oils and secondary metabolites commonly used as flavor for cooking, essences for cosmetics, and active components in medicines. De Mattia et al. (2011) proposed to use as a marker the non-coding *trnH–psbA* intergenic spacer, since it it characterizedby species and cultivar unique haplotypes providing a reliable system for their identification in all six major genera of cooking spices (*i.e.*, *Mentha*, *Ocimum*, *Origanum*, *Salvia*, *Thymus* and *Rosmarinus*). DNA barcoding has also been used to identify different aromatic species after industrial drying and shredding (De Mattia et al., 2011). DNA barcode markers were also efficiently used to identify commercial processed tea (Stoeckle et al., 2011).
Olive oil	It is a food product particularly prone to fraudulent practices since it commands a higher price than other vegetable oils. Olive oil is known to be one of the best vegetable fats in human diet and in recent years there has been increasing of interest in this kind of product due to its nutritional benefits and metabolic proprieties. Frauds have been detected through the mixing of lower-cost oils of plant origin, such as soya (*Glycine max*), canola (*Brassica napus*), maize (*Zea mays*), sunflower (*Helianthus annuus*) and even sesame (*Sesamum indicum*) oils, with olive oil. Besides with lower value vegetable oil, other frauds include fraudulent mislabelling on geographical origin and cultivar. Spaniolas et al. (2008) suggested that DNA barcoding can be succesfully applied to the main plant oils, inlcuding olive oils and allergenic oils as

	well as mixtures of them, as well as for the determination of geographical origin of the olive oil cultivars.
Honey	DNA barcoding has been proposed as molecular tool for honey traceability exploitable in distinct steps of the supply chain. *trnH́psbA* intergenic spacer was proven to be the most discriminant marker for identifying plant species in honey by Bruni et al. (2014).
Medicinal and Aromatic Plants (MAPs)	MAPs 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). The global market of herbal supplements had a value of USD 40 billion in 2017 and is expected to reach a market valuation in excess of USD 65 billion by 2025 (PErsistEncE MarkEt rEsEarch, 2017). The plants are processed immediately after being harvested. However, when the herbs undergo drying, fragmentation and pulverization processes the morphological traits cannot be used to reliably assess the botanic source. Moreover, many herbal ingredients are obtained by infusion, maceration, distillation or pressing. DNA barcoding approach is a valid molecular identification method to provide species-level resolution on MAPs as stated by Frigerio et al. (2019).
Neglected and Underutilized Species (NUS)	Campanaro et al. (2019) proposed the adoption of DNA barcoding as an effective tool to protect and promote the cultivation of NUS. These consist of a wide group of plants, diffused especially in equatorial and tropical countries, that if adopted at the global scale could substantially improve agricultural sustainability and improve conditions of human nutrition.

Table 8. Example applications of DNA Barcoding techonolgy in plant-based food products.

Chapter 5. CASES OF STUDY

This chapter reports different case studies focused on the application of DNA barcoding technology for traceability of plant-based products; their results indicate as the application of DNA barcoding is a sensitive, fast, cheap and reliable method for the identification and traceability throughout the industrial pipeline.

5.1. Case Study 1- *TrnH-psbA* for coffee traceability

The first case of study is a research published in 2018 by Uncu and Uncu, titled "Plastid *trnH-psbA* intergenic spacer serves as a PCR-based [DNA Barcode] marker to detect common grain adulterants of coffee (*Coffea arabica* L.)". The aim of the study was to verify the sensitivity of the barcode genotyping approach (by using *trnH-psbA* lenght polymorphism diversity) to authenticate botanical origin in coffee.

Problem Statement

Coffee is one of the most economically important crops, ranking second after petroleum as the most traded commodity. In addition to its high value as an export good, coffee is susceptible to fraud and adulteration of coffee with cheaper and easily accessible plant

material has long been recognized. It is reported that in Brasil, the top coffee producer and exporter country, around 25% of the brands in the coffee market are fraudulent (Lopes, 2018). Adulteration of high-quality Arabica coffee (*Coffea arabica* L.) with roasted cereal grains (e.g. corn, barley, wheat, rice) and soybean is a common practice.

Tevfik and Ozgur Uncu (2018) investigated the potential of the chloroplast *trnH-psbA* intergenic spacer to authenticate botanical origin in Arabica coffee (*C. arabica* L.). The intergenic spacer *trnH- psbA* was the barcode, as it is the most widely used noncoding barcode in molecular phylogenetics following the coding *rbcL* and *matK* sequences. It allows robust PCR amplifications from diverse plant taxa and harbors informative interspecific INDEL polymorphisms. Also, *trnL* was used as a second DNA barcode.

Interspecific length polymorphisms in the barcode sequence were used as diagnostic markers to discriminate between coffee and its common grain adulterants (corn, soybean, rice, wheat and barley).

Methods Workflow

Figure 10 shows the different samples analysed in Tevfik and Ozgur Uncu (2018) and the different steps to obtain the sequences related to their barcodes.

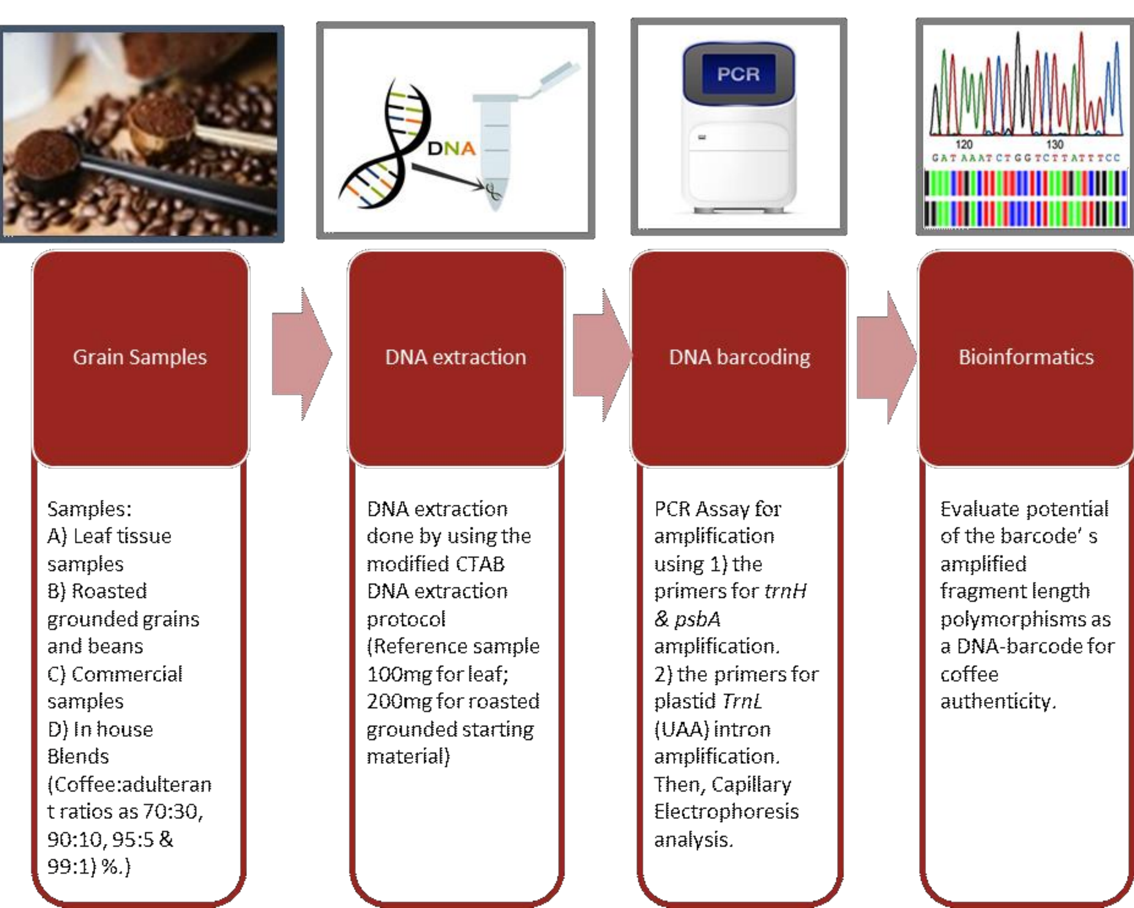


Figure 10. DNA barcoding steps.

Results

Each of the six plant species was characterized by a different *trnH-psbA* fragment size, allowing the discrimination of all the species (**Table 9**). *C. arabica* plastid marker was a single fragment of 362 bp for *trnH-psbA*. Any different peak in the amplification profile of the barcode sequence from a claimed 100% Arabica sample would indicate fraud with foreign plant material.

Comparisons within	trnH-psbA intergenic spacer	trnL (UAA) intron
Coffea arabica (coffee)	362 bp	585 bp
Glycine max (soybean)	372 bp	585 bp
Hordeum vulgare (barley)	685 bp	643 bp
Oryza sativa (rice)	690 bp	604 bp
Triticum aestivum (bread wheat)	699 bp	664 bp
Zea mays (corn)	693 bp	528 bp

Table 9. Barcode amplicon sizes of six plant species used in the Unco and Unco 2018 study.

Interspecific length polymorphisms in the *trnH-psbA* intergenic spacer were not implemented before in agrofood forensics, while the plastid *trnL* (UAA) intron was already shown to serve as a PCR-based marker of species origin in food products. The *trnL* (UAA) intron allowed the discrimination of coffee and the cereal species included in this work, but it was not useful to distinguish soybean from Arabica coffee (**Table 9**).

When different mixture of coffee with other cereal species were analysed at *trnH-psbA* it was possible to identify the different species, as shown by the capillary electropherograms displaying the barcode amplification profiles obtained from blends of Arabica coffee with soybean, corn and barley (**Figure 11**). The coffee and soybean- specific bands of 362 and 372 bp were successfully resolved, revealing the presence of soybean in all of the adulterated samples (**Figure 11A**). The same for roasted corn adulteration in Arabica coffee (**Figure 11B** and **11C**). By using characterization of trnL it was also possible to clearly identify roasted barley adulteration in Arabica (**Figure 11C**). The lower limit of adulteration detection tested with the three sets of coffee and cereals mixtures was 1 %; the PCR Capillary Elettrophoresis approach was effective and reproducible in identifying 1% adulteration in roasted, ground coffee beans in all replicate experiments (**Figure 11**).

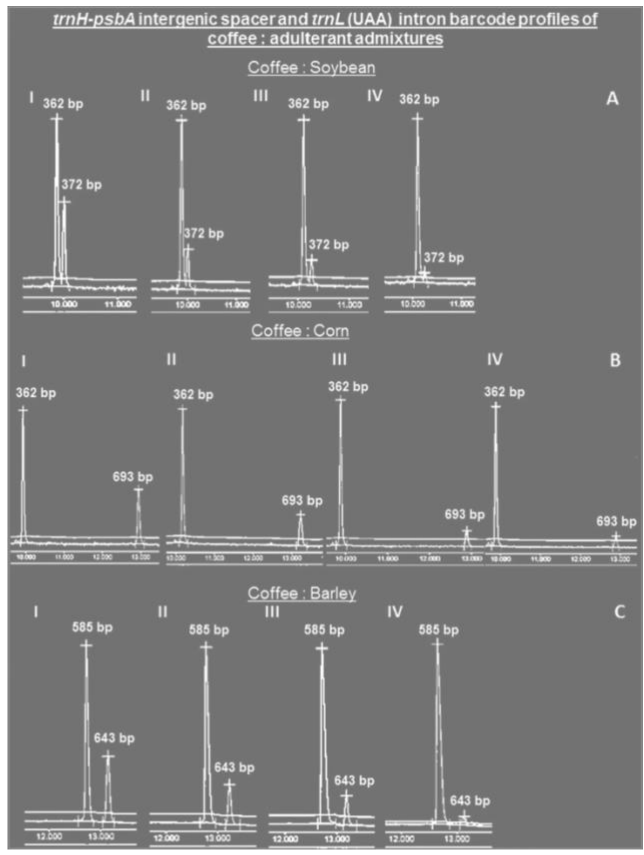


Figure 11. Capillary Electropherograms of mixtures. A, *trnH-psbA* intergenic spacer amplification profiles of coffee: soybean blends; B, *trnH-psbA* intergenic spacer amplification profiles of coffee: corn blends; C, *trnL* (UAA) intron amplification profiles of coffee: barley blends. The order of admixing ratios (coffee: adulterant) is the same for rows A, B and C: I, 70:30%; II, 90:10%; III, 95:5%; IV, 99:1%.

5.2. Case Study 2- Bar-HRM for identification of PGI lentil variety "Eglouvi"

It has been mentioned in the introduction the importance of plant-base food products for traceability even in high value products that have protected denomination of origin (PDO, PGI). This case of study was presented in Bosmali et al. (2012), which present a microsatellite and DNA-barcode regions typing combined with High Resolution Melting (HRM) analysis for the traceability and identification of the PGI lentil variety 'Eglouvi', even in mixtures.

Problem statement

A frequent problem in grain legumes is the mixing of high-quality seeds of well- known elite varieties with varieties of inferior quality (and price) or with seeds from other species, similar in color and shape but that are of poorer quality and might even be toxic for consumers. More specifically lentils are often contaminated with *Vicia* spp.

Lentil is a very important species for human nutrition as it has a high seed protein content (about 25%), it is rich in lysine and leucine and in dietary fibres (87%) and low in fat. Lentis are in fact one of the main proteins' sources for vegans. The best-known Greek lentil variety, "Eglouvi" is widely cultivated in the island of Lefkada (Western Greece),

has a good market position and potential, since it is up to be appointed as Protected Designation of Origin (PDO) EU mark. Currently, it is highly priced (10 euros per kg), offering a significant income to the local farmers.

High Resolution Melting analysis (HRM) measures the rate of double stranded DNA dissociation to single stranded DNA with increasing temperature. Coupling this with the recent development of saturating DNA dyes, allow the use of HRM for genotyping (i.e. SNP, SSR markers) and for quantification of adulterants. Therefore, when combined with DNA Barcoding results in Bar-HRM method, as previously described, which allows the authentication of plant species and PDO products and moreover the quantitation of adulterants in commercial products.

Plant Materials

In the case of study presented here, ten major Greek lentil varieties, one of them being "Eglouvi", as well as admixtures of "Eglouvi" with other varieties were used for analysis.

Methods Workflow

The steps applied to genotype the samples were showed and described in **Figure 12.** A third generation DNA intercalating dye, Syto®9, that at high concentrations can saturate all available sites within double stranded DNA was used. Fluorescence of Syto®9 provides a more accurate assessment of DNA melt status compared to SYBR Green I and can be used to monitor both the accumulation of the amplified product during PCR and the subsequent product melting by using the RotorGene 6000 software.

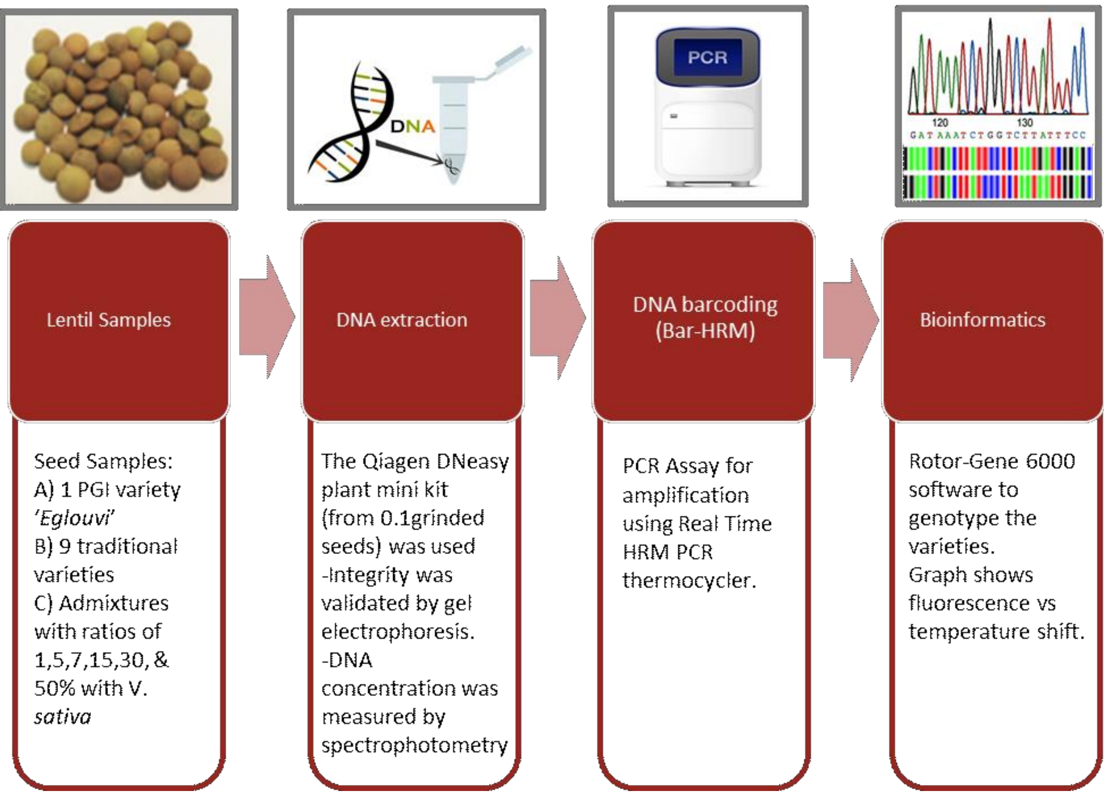


Figure 12. Steps of Bar-HRM.

Results

Different proportions of *V. sativa* adulterants in *L. culinaris* were detected via Bar-HRM analysis as it can be seen on **Figure 13** where temperature-shifted melting curves demonstrated more definite variances (amplicon dissociation reveals the actual degree of contamination resulting from adulteration).

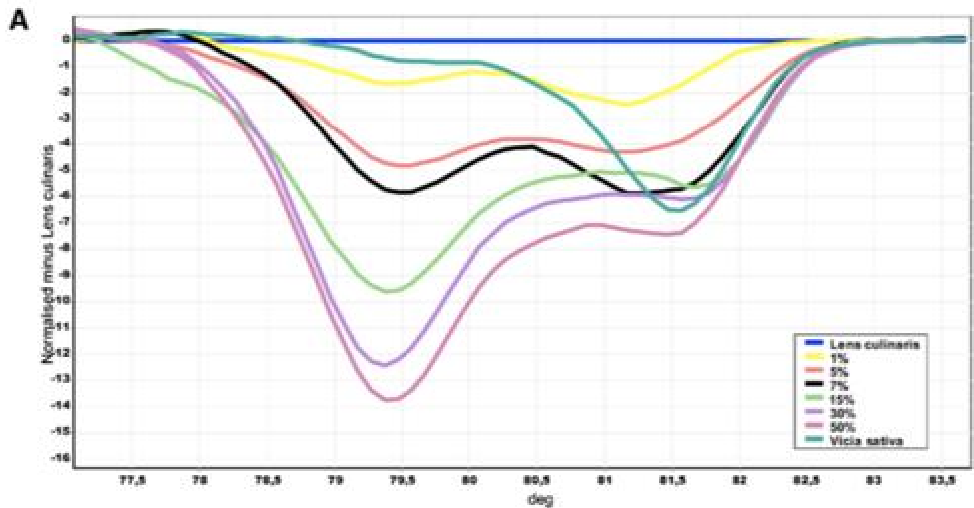


Figure 13. Plot of normalized fluorescence of mixed *L. culinaris/V. sativa.*

Figure 14 shows the sequences of *rpoC1* chloroplast DNA barcode targeting region of *L. culinaris* and *V. sativa;* differences in the DNA sequence are present, confirming the HRM results. In particular, three SNPs responsible for the differences in the observed melting curves between the different species amplicons.

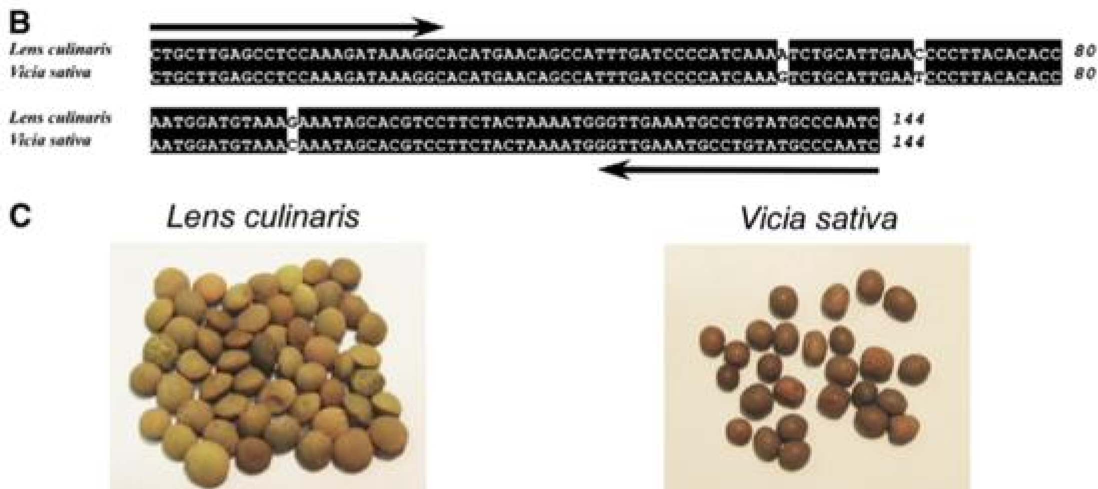


Figure 14. Alignment of *rpoC1* sequence of *L. culinaris and V. faba*.

The application of Bar-HRM on chloroplast *rpoC1* target region allowed the identification of both *V. sativa* and *L. culinaris* species, and the detection of *V. sativa* in *L. culinaris* commercial products was possible also at a ratio as low as 1:100. Hence, this

assay provided flexible, cost-effective genotyping methods that can be applied not only to the final product located in the market, but also during the industrial production chain: from ingredients reception to product packaging to prevent unintended frauds.

5.3. Case Study 3- DNA Metabarcoding for honey traceability

Beltramo et al. (2021) published an article titled “Exploring the botanical composition of polyfloral and monofloral honeys through DNA metabarcoding”. Here a summary of the main results obtained.

Problem Statement

Honey products are particularly relevant for the Italian market, that produces more than 30 different types of high-quility honey, regulated by the Council Directive 2001/110/EC of the European Union, that establishes the indication of the prevalent botanical composition and the geographical origin for honey labeling. Monofloral honey is a product obtained when bees restrict their selection prevalently to one plant, therefore monofloral honey has a higher market value due to its quality and characteristics, which is the reason it is often mislabeled making the consumer a victim of fraud.

DNA metabarcoding can be used to study complex (animal, vegetable, bacterial) matrices/populations without any a priori knowledge of their composition. The sequences generated by NGS are then analyzed by comparing them with a reference database from the same genomic region for genus/species assignment.

In this study 111 honey samples of different origin were analyzed by DNA metabarcoding of pollen to identify their botanical composition and to verify the authenticity of the declared label information of the products. To do this, were used the 80bp fragment of the chloroplastic gene *trnL* with the primers *trnL-g* and *trnL-h*, that have been already tested on restricted groups of samples (Laube et al., 2010; Utzeri et al., 2018); it was expected a succefull amplification from the extracted DNA to allow the analysis of the botanical composition of all of the honey samples.

Even tough the preferred chloroplast region to be amplified and used for barcoding is the *trnH–psbA*, Beltramo et al. (2021) used the *TrnL* as reported by Taberlet et al. (2007). Taberlet et al. (2007), known that the choice of the plant DNA barcode has to be based on the identification of a region that has to be as variable as possible; they also were aware that *TrnL* does not represent the most variable non-coding region, but it has some unique advantages such as the possibility of designing highly conserved PCR primers that amplify a very short DNA region, of no more than 100–150 bp (so basically it can be said that it is a mini-barcode region) in order allow reliable amplifications of even highly degraded DNA found in processed food or in fossil remains. In the case of honey, the DNA degradation level is expected to be high since it undergoes the conversion of flower nectar into honey inside the honey pouch of the bee, therefore a mini-barcode makes sense for this analysis.

Methods Workflow
The **Figure 15** shows the different steps of the analysis carried out in Beltramo et al. 2021, from samples preparation to species identification.

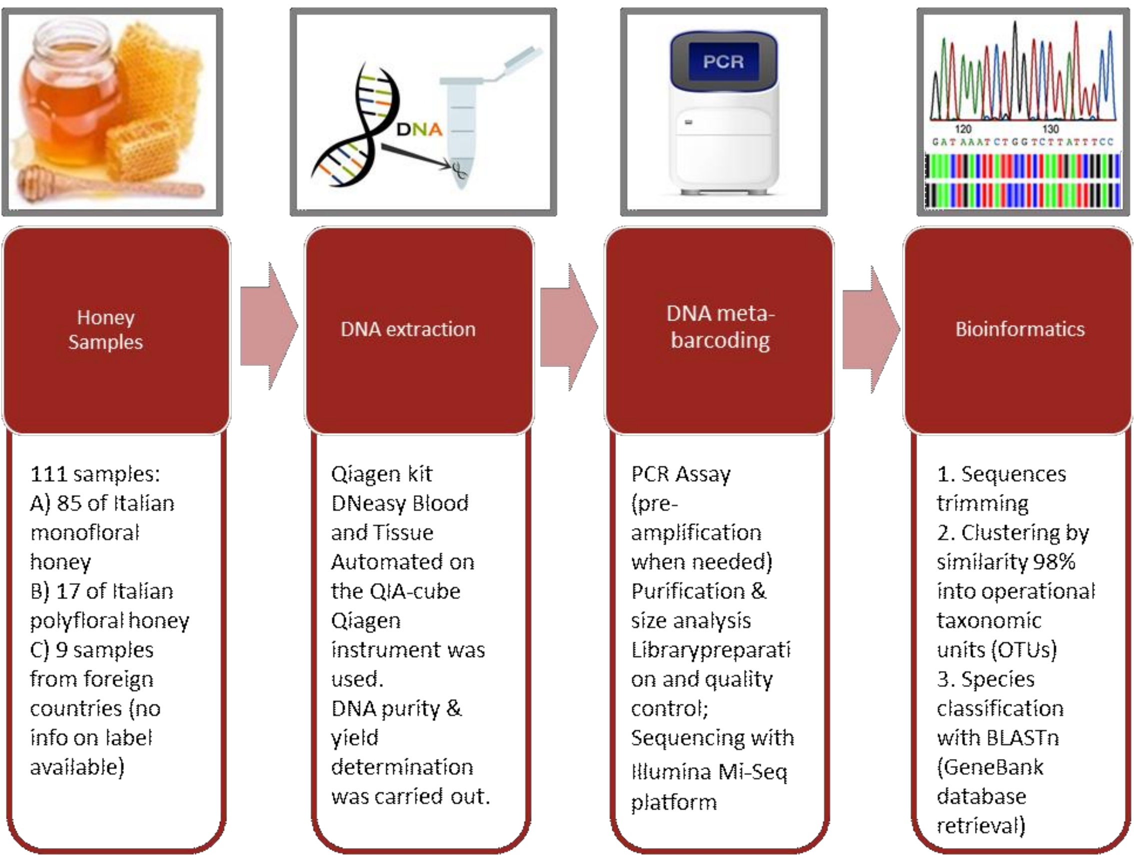


Figure 15. Step applied for DNA metabarcoding of honey samples.

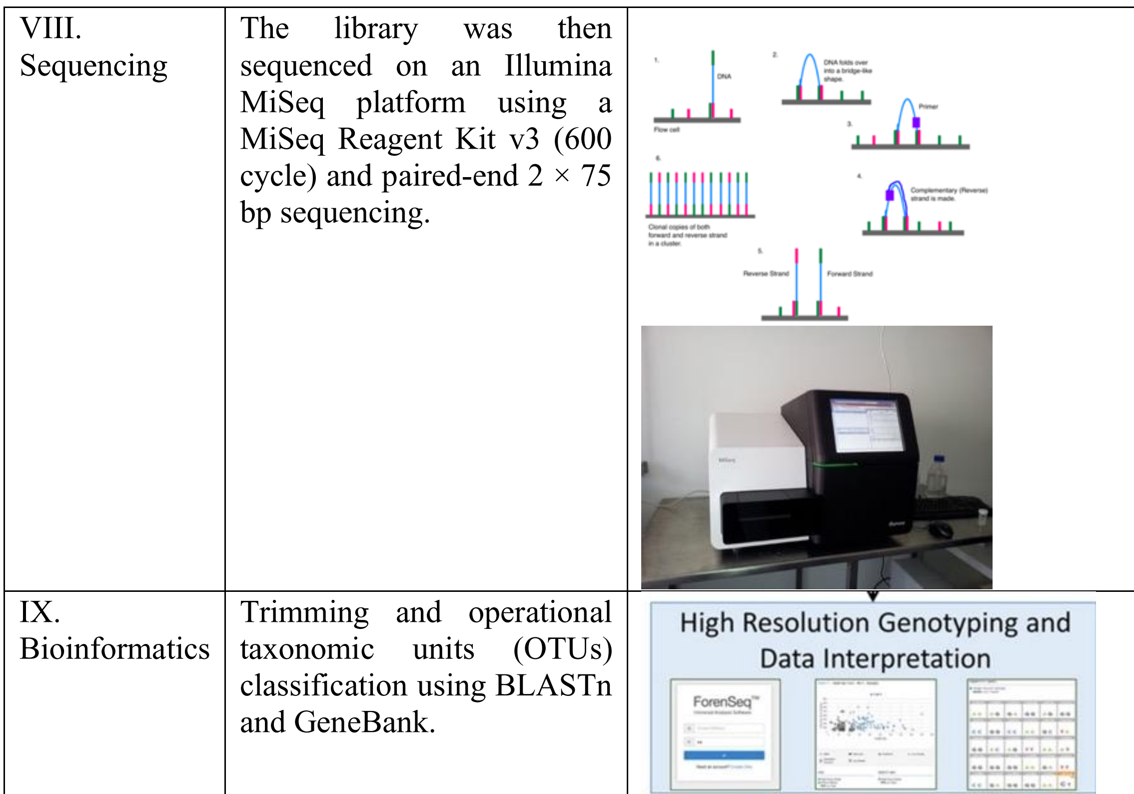
DNA was tested for the amplification of a fragment of the chloroplast gene *trnL* with the primers trnL-g and trnL-h; this primer pair provides amplicons of variable length (10–143 bp) in different taxonomical groups. Illumina adapters were ligated at the 5’ ends of the primer sequences for NGS analysis.

The NEBNext Library Quant Kit was used; its components are optimized to deliver significant improvements to qPCR-based library quantification for Illumina sequencing. The kit contains primers which target the P5 and P7 Illumina adaptor sequences, and a set of six high-quality, pre-diluted DNA standards to enable reliable quantification of diluted DNA libraries between 150–1000 bp. **Table 10** shows the steps carried out to develop the Mi-Seq library for DNA Metabarcoding:

<table>
<tr><th>Step</th><th>Description</th><th>Images</th></tr>
<tr><td rowspan="2">I. PCR for trnL fragments and illumina adapters</td><td>Pre-amplification (if needed). Some samples did not amplify easily: a pre-amplification with primers trnL-g and trnL-h <u>without adapter sequences</u> was performed to increase reaction efficiency.</td><td rowspan="2">Table 2
Sequences of primers and adapters for the DNA metabarcoding analysis of honey samples.<table><tr><th>Primer</th><th>Sequence (5′-3′)</th><th>Gene target</th></tr><tr><td>trnL-g</td><td>GGGCAATCCTGAGCCAA</td><td>trnL</td></tr><tr><td>trnL-h</td><td>CCATTGAGTCTCTGCACCTATC</td><td>trnL</td></tr><tr><td>Forward adapter</td><td>TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG</td><td></td></tr><tr><td>Reverse adapter</td><td>GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG</td><td></td></tr></table></td></tr>
<tr><td>After visualization on agarose gel, 5 µL of this pre-amplification reaction was</td></tr>
</table>

	used as a template for a second PCR using NEBNext Library Quant Master Mix & primers *trnL-g* **and** *trnL-h* **with** adapter sequences. The adapter sequences are *Illumina adapter sequences* that are added during the PCR by a technique called tagging (an enzymatic approach where the enzyme transposome binds the adapters into double strand DNA fragment)	
II. Size selection	All the PCR products were visualized on agarose gel to check the amplification of the expected product. The amplified DNA is then purified by magnetic beads (using Agencourt AMPure XP, Beckman Coulter). The way magnetic beads work is that they are in direct contact with the PCR products (that already contain adaptors to each end) and undergoe two selections: the fisrt one is for the larger size fragments, the second one removes DNA fragments smaller than the desired size; in this way only fragments with the targeted size remains.	a) b) c)
III. Index PCR	These purified fragments will be further used as a template for the index PCR; the reaction was prepared in a final volume of 50 µL using: 5 µL DNA 5 µL Nextera XT Index Primer 1 (N7xx) 5 µL Nextera XT Index Primer 2 (S5xx) 25 µL 2x KAPA HiFi HotStart ReadyMix 10 µL ultrapure H_2O	

IV. Clean up pooling and normalization	The PCR products were purified again using magnetic beads (in this case is a single step where the target library is the one attaching to the beads)	
V. Library analysis	Confirm the size distribution by diluting 1µl into buffer and analyzing on a Bioanalyzer 2100 (Agilent) using the high sensitivity DNA kit to verify library size (Qualitive validation) .	Diluted library DNA pipetted into the chip Chip reading
VI. Fragments Quantification	The amplified fragments were also quantified with the Qubit DNA HS kit on a Qubit 2.0 fluorimeter (Life Technologies) for normalization of the library at 4 nM. (Quantitative validation)	
VII. Library Quantification	The final library concentration was quantified by qPCR using the NEBNext Library Quant Kit for Illumina (New England Biolabs).	



VIII. Sequencing	The library was then sequenced on an Illumina MiSeq platform using a MiSeq Reagent Kit v3 (600 cycle) and paired-end 2 × 75 bp sequencing.	
IX. Bioinformatics	Trimming and operational taxonomic units (OTUs) classification using BLASTn and GeneBank.	

Table 10. Development of the Mi-Seq library.

Results

Honey samples for foreign countries - For these samples there was no information on the label regarding the flowers used, so no fraud can be looked for in there. **Figure 16** shows the percentage of flower origin content in four of the nine samples.

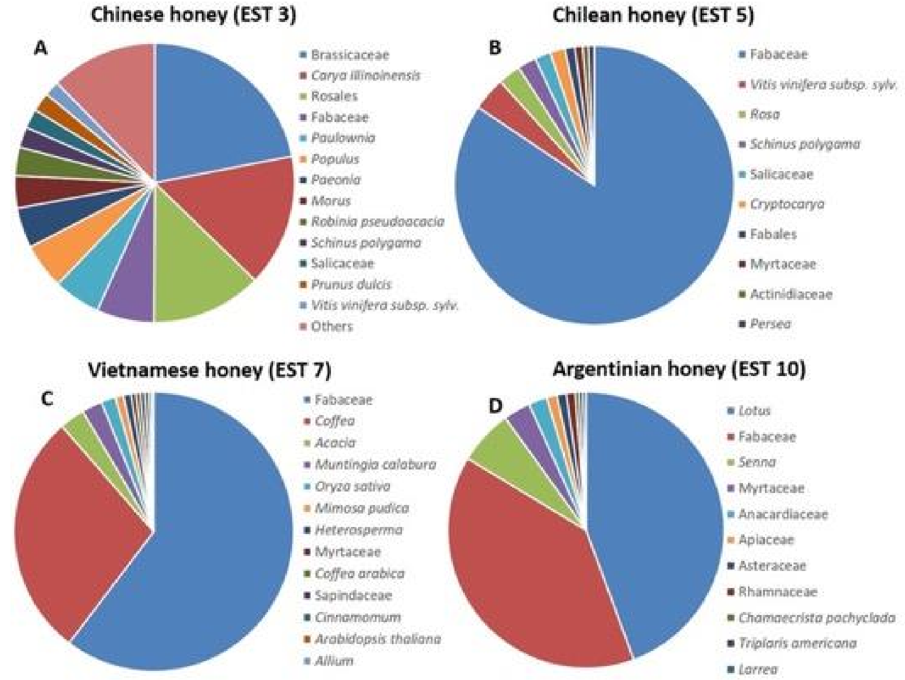


Figure 16. Pie charts of the botanical composition of honeys from China (A), Chile (B), Vietnam (C), and Argentina (D).

The list of families, genera and species identified was shown, depending on the taxonomical level of identification determined by the metabarcoding analysis.

Samples of polyfloral honey - the botanical composition of the polyfloral honeys had a high percentage (over 50%) of reads matching *Castanea sativa*, while the other plant groups were present at a frequency of around 20% or less for 13 samples. The other four polyfloral honeys had a more diverse composition.

Samples of monofloral honey - in 72 monofloral honey samples the reads matched the declared botanical origin at the species, genus or family level, The declared genus/species was not detected in 13 monofloral honeys: five of them where linden, and the remining were one of each of the maple, taraxacum, lavender, coriander, heather, acacia, strawberry tree, and rosemary honey samples. For all the other monofloral honeys, the reads matched the declared botanical origin, many of them not at a 100%, like we can see on pie chart B the Acacia honey has about 80% from the genus *Robinia* but the other 20% is a "contamination" from other flowers (**Figure 17**). Or for example the Eucalyptus Honey sample showed only 37% of its composition belonging to the Myrtaceae family (that is the same family as Eucalyptus but is not the same genus). Beside the plant taxa, origin frauds where also detected, for example, the Italian honey samples 2017–32 and 2017-125 showed 2.6% and 37.3% of reads, respectively, that matched *Dendrosenecio*, a genus of the Asteraceae family that grows in the mountain areas of Africa. There where also samples that showed origins from Asia, Chile, Mexico and Argentina. **Figure 17** shows example pie charts of the botanical composition of chestnut honey, acacia honey, rhododendron honey and a polyflower honey.

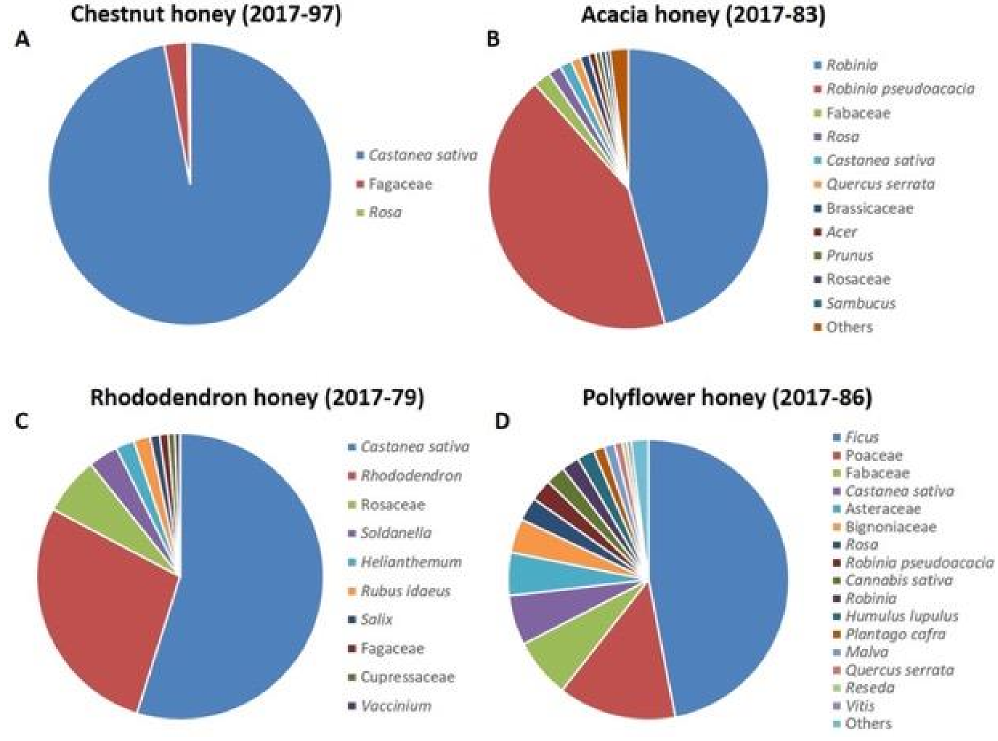


Figure 17. Example pie charts of the botanical composition of chestnut honey (A), acacia honey (B), rhododendron honey (C), and a polyflower honey (D).

Example pie charts of the botanical composition of chestnut honey (A), acacia honey (B), rhododendron honey (C), and a polyflower honey (D). The list of families, genera and

species identified was shown, depending on the taxonomical level of identification determined by the metabarcoding analysis

Conclusions

The primers *trnL-g/trnL-h* successfully amplified the target region from the extracted DNA and allowed analysis of the botanical composition of all the 111 honey samples. This short fragment has proven to be optimal to study environmental and degraded DNA. For all honey types, except heather, strawberry tree, and rosemary honeys, more than one sample was analyzed, with the identification of reads for the declared genus/family; this was not found for 13 samples. The lack of reads matching the labeled species in samples with high coverage can suggest a mislabeling event, but it must be considered that the pollen content of some types of honey, such as linden one, is very low. Finally, plant sources belonging to a species/genus/ family typical to geographical areas of the world were detected for the non-Italian and Italian honey samples. By knowing the botanical composition, speculations can be carried out about the geographical origin of honeys.

5.4. Case Study 4- Wine Traceability Through Microbiome Barcoding

An interesting approach for plant-based food products that undergoe fermentation process, such as wine, cheese, beer, or olives, is that the microbes used for its fermentation can be used to trace the area of origin. In this case study microbial diversity traces the area where the grapes are coming from.

The present case of study was published by Bokulich et al. (2016). The article is entitled "Associations among wine grape microbiome, metabolome, and fermentation behavior suggest microbial contribution to regional wine characteristics". In this study, microbial biodiversity is used to distinguish vineyards and viticultural areas.

Problem Statement

Wine production is a multi-billion-dollar global industry for which microbial control and wine chemical composition are crucial aspects of quality that need to be monitored along all the process, from ingredients reception to final bottled product. An important trait for wine is its *Terroir*, the basis of the PDO, IGP quality designations, but it has to be considered that many of the factors that contribute to terroir are nebulous. In this reseach is analyzed the postharvest microbiota (present in the grapes skin) to find out if it can exhibit distinct patterns of origin distribution at small geographical scale, meaning at scale of defining the American Viticultar Area (AVAs), by employinh high-throughput marker gene sequencing to longitudinally profile the bacterial and fungal consortia of over 200 commercial fermentations and musts of grapes grown throughout Napa and Sonoma Counties, CA.

Methods Workflow

the first step was to sample across Napa and Sonoma AVA designations, 23 vineyards in total. Since, it has been previously demonstrated by Bokulich et al. (2014) that regional, grape varietal, and climatic factors shape the bacterial and fungal communities of wine grapes across multiple growing years. This makes possible that the microbial comunity

present on the grape skin can trace the vineyard origin area. In this new study, Bokulich et al. (2016) started by mapping and sampling the different grape-growing regions of California with AVA designations, as can be seen on **Figure 18**.

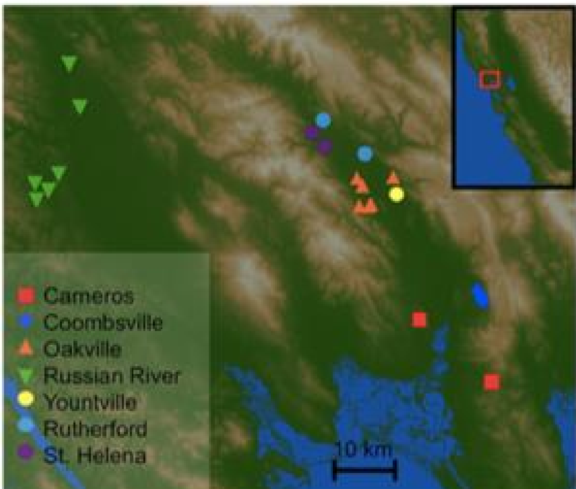


Figure 18. Map of sampling sites across Napa and Sonoma Counties. Each point represents an individual vineyard from which grapes were harvested for the fermentations monitored in the study. Points are colored by AVA designation, as indicated in the legend

DNA Barcoding was applied to the bacterial and fungal organism found on grape skin in order to validate if microbial patterns can be traced back by AVA (**Figure 19)** at different stages of fermentation (Must/Juice, Wine without MLF (white wine, meaning Chardonnay) and Wine with MLF (Red wine, meaning Cabernet Sauvignon).

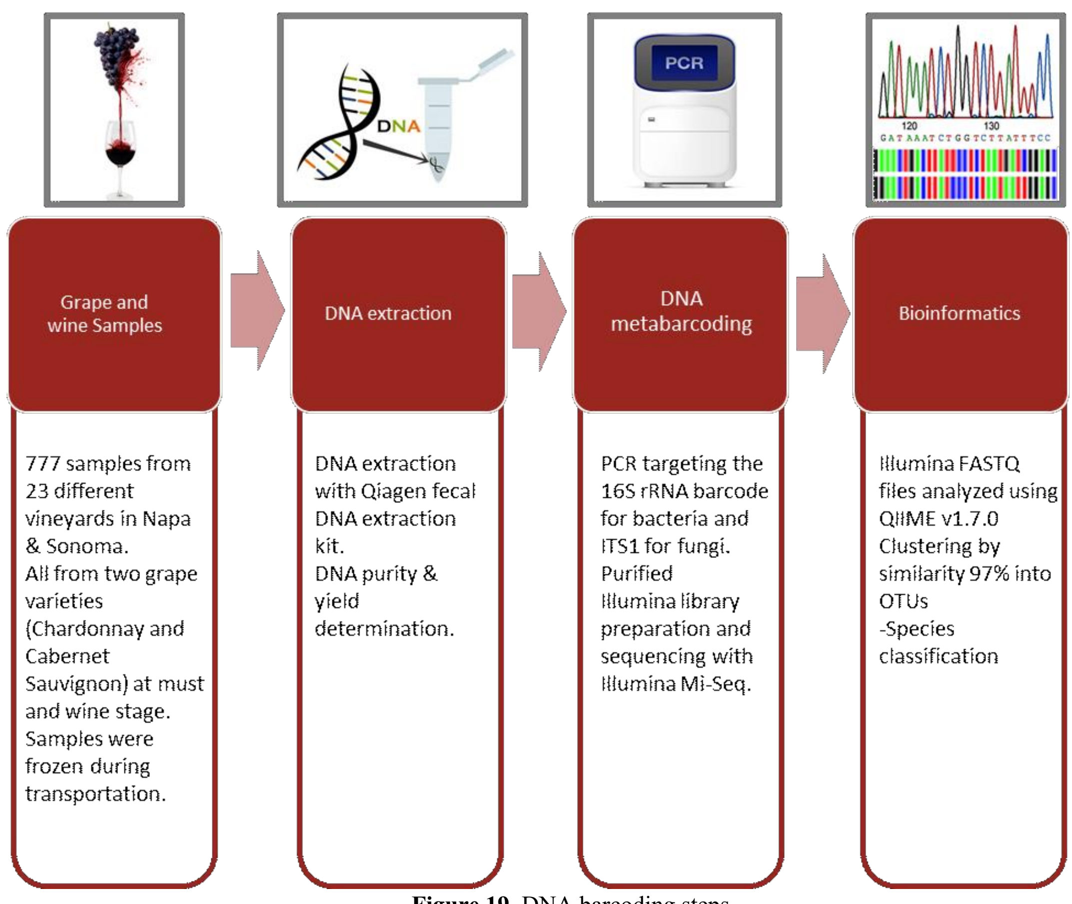


Figure 19. DNA barcoding steps.

Results

Individual AVAs and vineyards were distinguished based on the microbial consortia present in the grape must/juice.

A. Principal Component Analysis (PCoA)

After sequencing, samples represented by almost 500 bacterial and 100 fungal sequences following all quality-filtering steps were obtained. Beta-diversity (similarity between samples) was calculated within QI-IME using the weighted UniFrac distance between samples to assess similarity among bacterial communities and Bray-Curtis dissimilarity for fungal communities. Principal coordinates were computed from the resulting distance matrices to compress dimensionality intro three-dimensional principal-coordinate analysis (PCoA) plots, enabling visualization of sample relationships. These results can be seen in the next **Figure 20.**

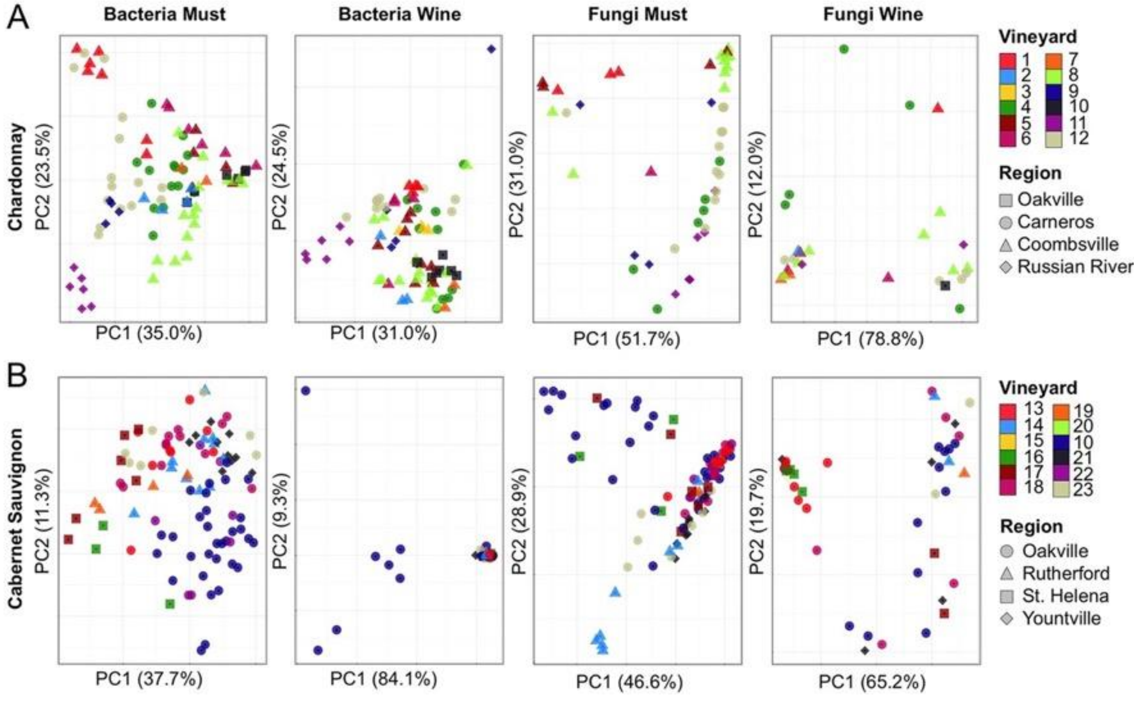


Figure 20. PCoA comparisons of bacterial weighted UniFrac distance (left two columns) and fungal Bray-Curtis dissimilarity (right two columns) in musts and wines (see column labels), categorized by vineyard (color) and AVA source (shape). (A) Chardonnay, (B) Cabernet Sauvignon. Each point represents an individual sample, and sample proximity on the plot is a function of similarity in bacterial and fungal community composition.

Figure 20 shows that microbiota exhibits regional variation in musts and wines. Both AVA and vineyard-specific microbial signatures diminished a little bit during fermentation (**Figure 20**) as growth of fermentative organisms reshaped the community structure, richness, and diversity of the wines. This effect was largely dependent on grape variety and winery; Chardonnay vineyards and AVAs retained significantly different bacterial profiles at end of fermentation, and Cabernet fungi differentiated vineyard origin of at least one vineyard, but Cabernet bacterial profiles became less distinct due to growth of *Leuconostocaceae* (O. *oeni)* during malolactic fermentation (MLF).

B. Random Forest Classification:

Random forest classification models confirm that all vineyards are distinguishable at accuracies between 79% (Chardonnay juice) and 82% (Chardonnay wine), [note the 81% (Cabernet Wine) and almost 83% (Cabernet Wine)] based on microbial profiles in the finished wine, indicating that vineyard-specific signatures (eventhough diminished as it has been seen on **Figure 20**) are do still retained through fermentation (**Table 11**)

	Error Rate[a]	Random Rate[b]	ER/Random[c]
Cabernet Must	17.05	75.00	4.40
Cabernet Wine	19.35	85.71	4.43
Chardonnay Juice	20.45	85.71	4.19
Chardonnay Wine	17.65	66.00	3.74

Table 11. Random Forests Models predict vineyards origin of grape musts; [a], Error rate: percentage of misclassification of out-of-bag must samples to the wrong vineyard; [b], Random rate: percentage of misclassification expected due to random error; c, ER/Random: Error rate / Random rate.

Vineyard origin was proven through its microbiome at an accuracy level of 80%. Regional strain diversity in the many other bacterial and fungal species involved in wine production may similarly contribute to microbial *terroir* and deserves further investigation. Thus, local conditions appear to modulate microbial communities in addition to regional effects.

It is also to be noticed that the discrimination was also dependent upon the grape variety, indeed Chardonnay demonstrated stronger AVA differentiation for both bacterial and fungal profiles than Cabernet Sauvignon.

5.5. Case Study 5- DNA barcoding for highly processed products: the Açaí case.

This case of study refers to Lugon et al. (2021) research.

Problem Statement

Açaí is an Amazon superfruit with antioxidant and anti-inflammatory properties; it has high contents of bioactive compounds such as anthocyanins, flavonoids, and phenolic acids. Açaí products marketed as pulps, juices, sorbets, popsicles, and powder are so processed that it is hard to verify, visually, if they are truly açaí (*Euterpe oleracea* or *E. precatoria*), or another species, known as juçara fruit (*E. edulis*) that present morphological similarities and may be mistaken with açai.

The objective of the study was to define a unique and universal system in order to identify the *Euterpe* species using DNA barcodes; nine regions were studied as barcode candidates: *rbcL, matK, trnH-psbA, ycf1b, trnL* intron, *trnL-F, psbK-I, ETS*, and *ITS2*.

Methods Workflow

Fifty commercial products where bought directly from the market, including pure açai products and also mixed products (containing ingredients such as açaí berries, water, banana, guarana, strawberry, and more), aiming for different formats (**Figure 21**). **Figure 22** shows the steps to obtain sequences of the different barcodes.

Figure 21. Açai Market Products; a) açaí capsules; b) açaí powder; c) pure frozen açaí pulp; d) açaí popsicle; e) açaí sorbet.

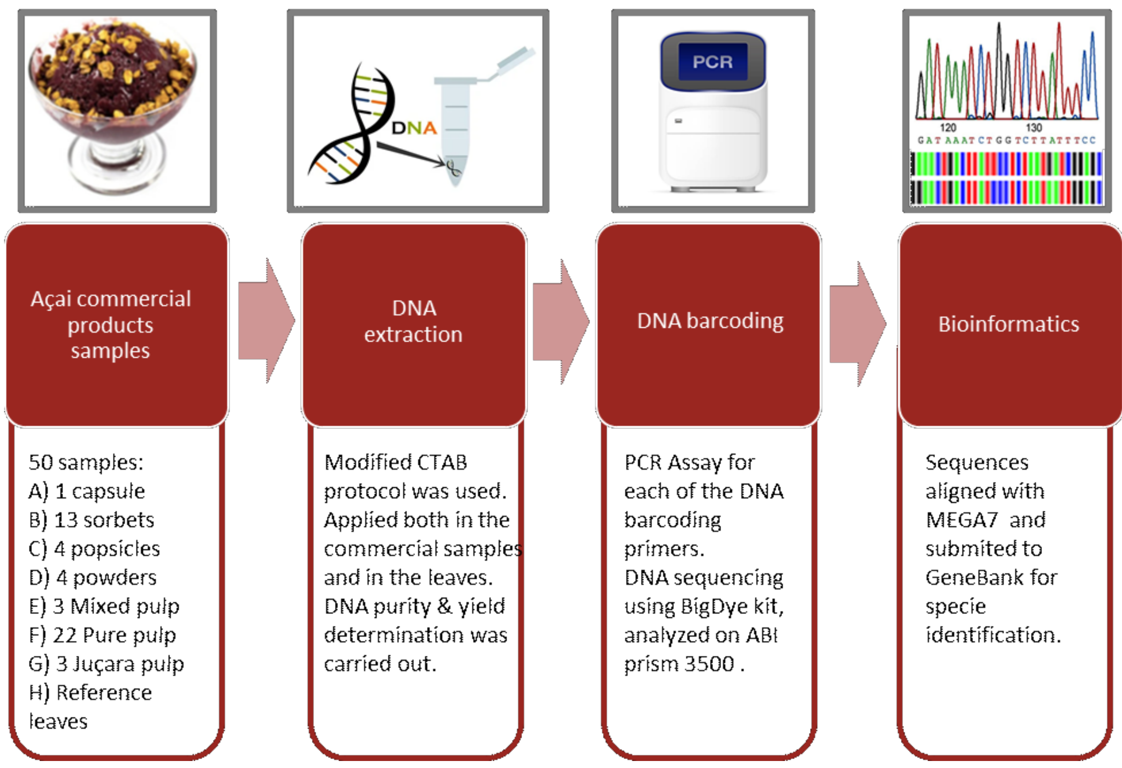


Figure 22. DNA barcoding steps.

Results

A. Ideal Barcode selection

Each of the nine studied barcodes candidates were analyzed in terms of success rate of PCR (**Figure 23a**) and success rate of sequencing (**Figure 23b**) for each of the three studied species using only the references leaves.

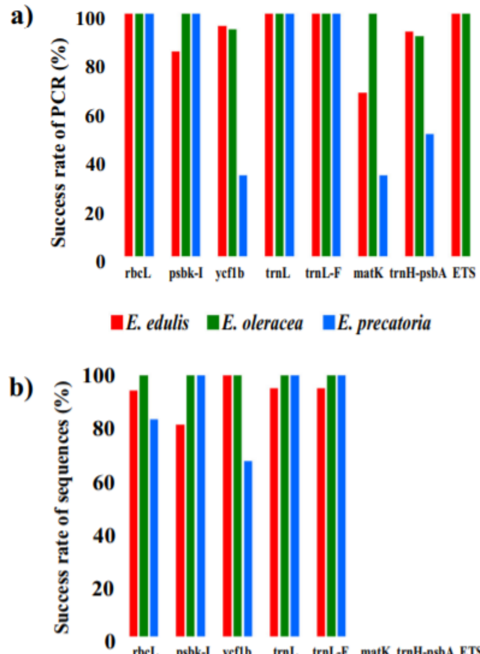


Figure 23. PCR amplification success (a) and sequencing success rates (b) of the eight DNA barcoding regions in reference samples.

The CBOL Plant Working Group proposes the use of *rbcL* and *matK* as the core barcodes and to use *trnH-psbA* and *ITS2* as complementary regions, however, for the species analyzed in this work, these markers were not considered suitable as barcodes, while *psbK-I* was considered the best region to differentiate the three species clearly through sequence alignment, therefore it was the selected barcode to be applied to all 50 commercial samples.

B. Commecial Samples Traceability

There was no *psbK-I* amplification in 12% of the samples (the capsule açaí sample, three popsicle samples, and two powder samples). The remaining samples were sequenced, but 18% of those presented mixed chromatograms, presenting some fluorescent signals overlapping. As a result, a total of 35 high-quality sequences were obtained (70% of the samples) (**Figure24a**).

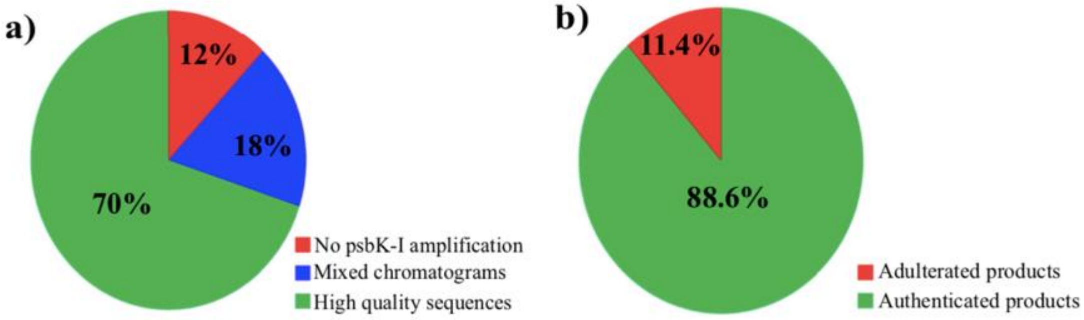


Figure 24. Authentication test using *psbK-I* as DNA barcoding in commercial products. a) a total of 50 products were tested; b) a total of 35 high-quality sequences were identified depending on the concordance between the species expected and sequence obtained.

From those 35 high-quality sequences, 11.4% were found without concordance between the species expected and sequence obtained (**Figure 24b**). These adulterated products corresponded to one açai sorbet and three pure frozen açaí, that presented characteristic sequence of *E. edulis*.

These results reveal concern about the correct identification of species in food. There are a large number of articles in the literature that correlate a species with a particular action (e.g., anti-cancer) and most of the time there was no genetic authentication of these samples.

Chapter 6. CONCLUSIONS

To meet the expectations of customers who demand access to a safe and genuine food product, it is critical to increase quality assurance and transparency along the whole food supply chain by adopting traceability systems, from the sourcing of raw ingredients to the finished food product.

Tecnology providers, researchers, end-users and policy decision makers are more and more concerned on satisfing the needs of food safety and quality, as can be seen in **Figure 25**, that shows the trend in food traceability publications from 1999 to 2018.

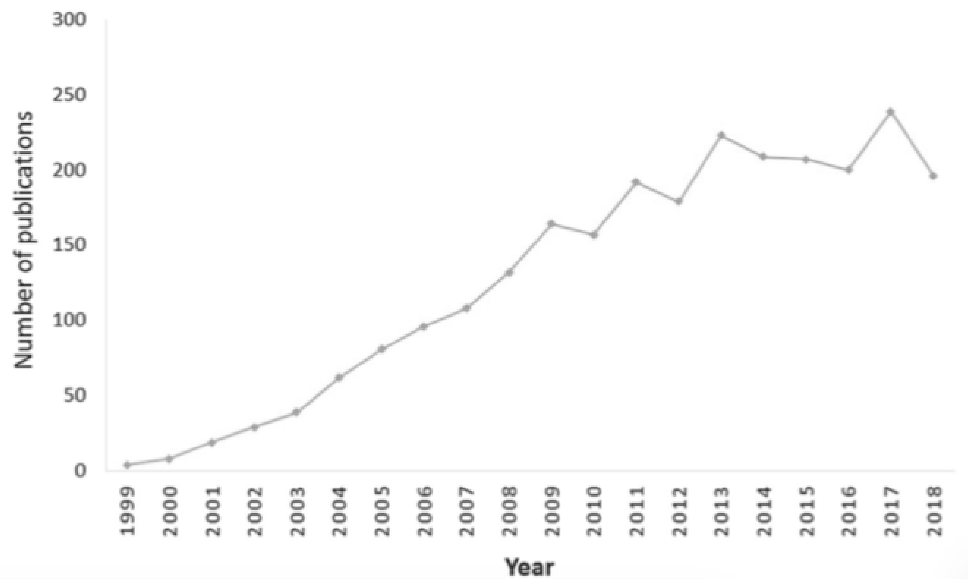


Figure 25. Trend in food traceability publications. Source: Violino, 2019.

Among the different techniques for agri-food traceability and authentication, the molecular approaches are gaining increasing interest due to their significant advantages compared to the physico-chemical approaches. Just the theorical fact that DNA is present in nearly all the cells of a given organism and its sequence remains unchanged during all production phases, makes it the fraud-free tool that is so needed. True, some industrial transformations, such as chemical treatments, can affect the yield, integrity, and quality of DNA, however, the use of DNA Barcodes from mitocondrial and chloroplast genomes, guarantees much less extent of damage, especially due to its abundance and size. Moreover, the technological advances of NGS techniques, along with a cost reduction and more user-friendly options for analysis, are making DNA Barcoding based approaches increasingly widespread in food authenticity.

It is known that some of the major criticisms of DNA Barcoding include the unavoidable complexity of biological matter or the choose of the appropiate bioinformatics tool for the analytical method, as well as the right set of primers. However, developing expertise in evaluating the scenario in a case-by-case manner can solve this concern. DNA barcoding, including all its complementary methodologies (e.g., target Real-Time PCR or metabarcoding), is often the best option not only in terms of accuracy and reliability, but also for its multiplexing capability, low cost and low required specialization of involved operators.

An “ideal” traceability system would follow the “history” of a product from its origin to its final use, taking into account all transformation and commercialization steps. The presented case studies had proven the efficacy and versatility of DNA Barcoding in the many different types of scenarios that can be found in plant-based food produts, from a) **mixed products**, such as the polyfloral honey, where adapted DNA Barcoding technology based on the use of NGS and mini-barcode fragments were applied; b) **highly processed products**, such as coffee grains, for which it was proved that is possible to identify plant materials, even in small traces; c) **products with different food matrices**, such as the grape skin containing complex microbiome; d) **fresh or processed food**; e) to its use in traceability of **validating the food product purity** with absence of adulterants in crops such as has been seen in the lentil case, where combined with HRM DNA barcoding detected adulterants as low as 1%.
Its time to move food production industry safety and quality level not to that of fraud mitigation nor fraud detection but to one of fraud prevention. DNA barcoding promises to attain the goal of ensuring consumer freedom of choice and improving the transparency of food production systems in a worldwide level.

To give a panorama of future applicability scenarios of DNA Barcoding in terms of traceability, it can be recalled how it was exposed on case Study 4 that by analizing the microbioma of a living organism (grapes in that case) hidden information about its characterization, as well as the climate surrounding it, its geographical location, its agricultural practices, and its interactions with other organisms can be discovered. Therefore, if this same tracing is applied to a human and its internal microbiota, the possibilities of characterizing a human-food behavoir and its interactions with the world are practically unlimited, allowing to help a designed personalized food plan for obtaining proper health benefits and preventing pathologies.

Chapter 7. BIBLIOGRAPHY

1. AIJN. 2014. Market Report Liquid Fruit. European Fruit Juice Association. 15-16.
2. Barcaccia, G., Lucchin, M., Cassandro, M. 2015. DNA Barcoding as a Molecular Tool to Track Down Mislabeling and Food Piracy. *Diversity* 8: 2.
3. Beltramo, C., Cerutti, F., Brusa, F., Mogliotti, P., Garrone, A., Squadrone, S., Acutis, P.L., Peletto, S. 2021. Exploring the botanical composition of polyfloral and monofloral honeys through DNA metabarcoding. *Food Control 128*: 108175.
4. 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.
5. Bokulich, N.A. et Al. 2016. Associations among Wine Grape Microbiome, Metabolome, and Fermentation Behavior Suggest Microbial Contribution to Regional Wine Characteristics. ASM Journals mBio 7: 3.
6. Bokulich, N.A., Thorngate, J.H., Richardson, P.M., Mills, D.A. 2014. Microbial biogeography of wine grapes is conditioned by cultivar, vintage, and climate. *Proceedings of the National Academy of Sciences of the United States of America* 111: E139–E148.
7. Bosmali, I., Ganopoulos, I., Madesis, P., Tsaftaris, A. 2012. Microsatellite and DNA-barcode regions typing combined with High Resolution Melting (HRM) analysis for food forensic uses: A case study on lentils (Lens culinari*s). Food Research Int*ernational 46: 141–147.
8. Bosmali, S.A., Ordoudi, M.Z., Tsimidou, P., Madesis. 2017. Greek PDO saffron authentication studies using species specific molecular markers. *Food Research International* 100: 899-907.
9. Brown, B., Emberson, R. M. & Paterson, A. M. 1999 Mitochondrial COI and II provide useful markers for *Weiseana* (Lepidoptera, Hepialidae) species identification. *Bulletin of Entomological Research* 89: 287–294.
10. Bruni, I., De Mattia, F., Galimberti, A., Galasso, G., Banfi, E., Casiraghi, M., et al. 2010. Identification of poisonous plants by DNA barcoding approach. *International Journal of Legal Medicine* 124: 595–603.
11. Bruni, I., Galimberti, A., Caridi, L., Scaccabarozzi, D., de Mattia, F., Casiraghi, M., Labra, M. 2014. A DNA barcoding approach to identify plant species in multiflower honey. *Food Chemistry 170*: 308–315.
12. Bucklin, A., Guarnieri, M., Hill, R. S., Bentley, A. M. & Kaartvedt, S. 1999. Taxonomic and systematic assessment of planktonic copepods using mitochondrial COI sequence variation and competitive, species-specific PCR. *Hydrobiology* 401: 239–254.
13. Campanaro, A., Tommasi, N., Guzzetti, L., Galimberti, A., Bruni, I., Labra, M. 2019. DNA barcoding to promote social awareness and identity of neglected, underutilized plant species having valuable nutritional properties. *Food Research International* 115:1-9.
14. 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. *Brief Bioinform* 11(4): 440-53.
15. Consortium for the Barcode of Life (CBOL) Plant Working Group. 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America* 106: 12794–12797
16. Corrado, G. 2016. Advances in DNA typing in the agro-food supply chain. *Trends in Food Science and Technology* 52: 80–89.

17. Cox, A. J. and Hebert, P. D. N. 2001. Colonization, extinction and phylogeographic patterning in a freshwater crustacean. *Molecular Ecology Notes* 10: 371–386.
18. Danezis, G.P., Tsagkaris, A.S., Camin, F., Brusic, V., Georgiou, C.A. 2016. Food authentication: Techniques, trends & emerging approaches. *Trends Analytical Chemistry* 85: 123–132.
19. De Mattia, F., Bruni, I., Galimberti, A., Cattaneo, F., Casiraghi, M., Labra, M. A. 2011. Comparative study of different DNA barcoding markers for the identification of some members of Lamiaeea. *Food Research International 44*: 693–702.
20. DeSalle, R. 2006. Species discovery versus species identification in DNA barcoding efforts: response to Rubinoff. *Conservation Biology* 20:1545–7.
21. Ding, Y., Jiang, G., Huang, L., Chen, C., Sun, J., Zhu, C. 2020. DNA barcoding coupled with high-resolution melting analysis for nut species and walnut milk beverage authentication. *Journal of Science and Food Agriculture* 100: 2372–2379.
22. Doyle, J. J., Gaut, B. S. 2000. Evolution of genes and taxa: a primer. *Plant Molecular Biology* 42: 1–6.
23. Erickson, D.L., Spouge, J.L., Resch, A., et al. 2008. DNA barcoding in land plants: developing standards to quantify and maximize success. Taxon 13:1304–1316
24. European Commission. 1997. Green Paper on the General Principles of Food Law in the European Union; *Commission of the European Communities:* 1–63.
25. European Commission. 2000 White Paper on Food Safety. *Commission of the European Communities: 1–52.*
26. European Commission. Commission Regulation (EC) No. 1825/2000 of 25 August 2000 laying down detailed rules for the application of Regulation (EC) No 1760/2000 of the European Parliament and of the Council as regards the labelling of beef and beef products. *Official Journal of the European Community* L216: 8–12.
27. European Commission. Council Regulation (EC) No 178/2002 of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety. *Official Journal of the European Community* L31: 1–24.
28. Everstine, K., Spink, J., Kennedy, S. 2013. Economically motivated adulteration (EMA) of food: Common characteristics of EMA incidents. *Journal of Food Protection* 76: 723–735
29. Fanelli, V., Mascio, I., Miazzi, M.M., Savoia, M.A., De Giovanni, C., Montemurro, C. 2021. Molecular Approaches to Agri-Food Traceability and Authentication: An Updated Review. *Foods 10*: 1644.
30. Faria, M.A., Magalhães, A., Nunes, M.E., Oliveira, M.B.P.P. 2013. High resolution melting of trnL amplicons in fruit juices authentication. *Food Control* 33: 136-141.
31. Fazekas, A. J., Burgess, K. S., Kesanakurti, P. R., Graham, S. W., Newmaster, S. G., Husband, B. C., et al. 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *Public Library of Science One* 3: e2802.
32. Fazekas, A. J., Kesanakurti, P. R., Burgess, K. S., Percy, D. M., Graham, S. W., Barrett, S. C. H., et al. 2009. Are plant species inherently harder to discriminate than animal species using DNA barcoding markers?. *Molecular Ecology Resources* 9: 130–139.
33. Ferguson, J.W.H. 2002. On the use of genetic divergence for identifying species. *Biological Journal of the Linnean Society* 75:509–16.
34. 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: 294–299.

35. Frigerio, J., Gorini, T., Galimberti, A., 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.
36. Galimberti, A., Bruno, A., Mezzasalma, V., Mattia, F., Bruni, I., Labra, M. 2015. Emerging DNA-based technologies to characterize food ecosystems. *Food Research International* 69: 424-433.
37. Galimberti, A., Casiraghi M., Bruni I.,et. al. 2019. From DNA barcoding to personalized nutrition: the evolution of food traceability. *Food Science* 28: 41–48
38. Ganopoulos, I., Bazakos, C., Madesis, P., Kalaitzis, P., Tsaftaris, A. 2013. Barcode DNA high-resolution melting (Bar-HMR) analysis as novel close-tubed and accurate tool for olive oil forensic use. *Journal of the Science of Food and Agriculture 93*: 2281–2286.
39. Goldstein, P.Z., DeSalle, R. 2003. Calibrating phylogenetic species formation in a threatened insect using DNA from historical specimens. *Molecular Ecol*ology 12: 1993–8.
40. Hajibabaei, M. and C. McKenna. 2012. DNA Barcodes: Methods and Protocols, Chapter 15: DNA Mini-barcodes: 339-354.
41. Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A. C., and Baird, D. J. 2011. Environmental barcoding: A next-generation sequencing approach for biomonitoring applications using river benthos. *Public Library of Science One* 6: e17497.
42. Hajibabaei, M., Smith, M.A., Janzen, D.H. et. al. 2006. A minimalist barcode can identify can identify a specimen whose DNA is degraded. *Molecular Ecology Notes* 6: 959–964
43. Hammond, P. 1992. Species inventory. *Global biodiversity: status of the earth's living resources* (ed. B. Groombridge):17–39.
44. Hawksworth, D. L. & Kalin-Arroyo, M. T. 1995. Magnitude and distribution of biodiversity. *Global biodiversity assessment* (ed. V. H. Heywood): 107–191.
45. Hebert, P. D. N., Stoeckle, M. Y., Zemlak, T. S., and Francis, C. M. 2004. Identification of birds through DNA barcodes. *Public Library of Science Biology* 2: 1657–1663.
46. Hebert, P.D.N., Cywinska, A., Ball, S.L. and deWaard, J.R. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270: 313-321.
47. Hellberg, R. S., and Morrisey, M. T. 2011. Advances in DNA-based techniques for the detection of seafood species substitution on the commercial market. *Journal of Laboratory Automation* 16: 308–321.
48. Herbert, P., Ward, R.D., Zemlak, T.S., et al. 2005. DNA barcoding Australia's Fish species. *Philosophical Transactions of the Royal Society B* 360:1847–57.
49. Hollingsworth, M. L., Clark, A. A., Forrest, L. L., Richardson, J., Pennington, R. T., Long, D. G., et al. 2009. Selecting barcoding loci for plants: Evaluation of seven candidate loci with species level sampling in three divergent groups of land plants. *Molecular Ecology Resources* 9: 439–457.
50. Hollingsworth, P. M., Graham, S. W., & Little, D. P. 2011. Choosing and using a plant DNA barcode. *Public Library of Science One* 6: e19254.
51. Hurst, G.D.D., Jiggins, F.M. 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society of London Series B Biology* 272:1525–34.

52. Jaakola, L., Suokas, M., and Häggman, H. 2010. Novel approaches based on DNA barcoding and high-resolution melting of amplicons for authenticity analyses of berry species. *Food Chemistry* 123: 494–500.
53. Kane, N., Sveinsson, S., Dempewolf, H., Yang, J.Y., Zhang, D., Engels, J.M.M., Cronk, Q. 2012. Ultra-barcodinging cacao (*Theobroma*spp.; *Malvaceae*) using whole chloroplast genomes and nuclear ribosomal DNA. *American Journal of Botany 99*: 320–329
54. Khansaritoreh, E., Salmaki, Y., Ramezani, E., Akbari Azirani, T., Keller, A., Neumann, K., Alizadeh, K., Zarre, S., Beckh, G., Behling, H. 2020. Employing DNA metabarcoding to determine the geographical origin of honey. *Heliyon 6*: e05596.
55. Knowlton, N. and Weigt, L. A. 1998. New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society of London Series B Biology* 265: 2257–2263.
56. Kress, W.J., Erickson, D.L. 2012 DNA Barcodes: Methods and Protocols (Methods and Protocols), vol 858. *Humana Press*, Totowa, NJ.
57. Kress, W.J., Erickson, D.L., Swenson, N.G. et al. 2010. Advances in the use of DNA barcodes to build a community phylogeny for tropical trees in a Puerto Rican forest dynamics plot. *Public Library of Science One 5*: e15409
58. Kumar, P., Gupta, V. K., Misra, A. K., Modi, D. R., Pandey, B. K. 2009. Potential of molecular markers in plant biotechnology. *Plant omics* 2(4): 141-162.
59. Kumar, S., Gadagkar, S. R. 2000. Efficiency of the neighbour-joining method in reconstructing deep and shallow evolutionary relationships in large phylogenies. *Journal of Molecular Evolution* 51: 544–553.
60. Kumar, S., Tamura, K., Jacobsen, I.B. & Nei, M. 2001. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17(12): 1244-5.
61. Kurtzman, C. P. 1994. Molecular taxonomy of the yeasts. *Yeast* 10: 1727–1740.
62. Lagiotis, G., Stavridou, E., Bosmali, I., Osathanunkul, M., Haider, N., Madesis, P. 2020. Detection and quantification of cashew in commercial tea products using High Resolution Melting (HRM) analysis. *Journal of Food and Science* 85: 1629–1634.
63. Li, D. Z., Gao, L. M., Li, H. T., Wang, H., Gea, X. -J., Liu, J. Q., et al. 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences of the United States of America* 108: 19641–19646.
64. Little, D.P., Stevenson, D.W. 2007. A comparison of algorithms for the identification of specimens using DNA barcodes: examples from gymnosperms. *Cladistics* 23:1–21.
65. Lo, Y.T., Shaw, P.C. 2018. DNA-based techniques for authentication of processed food and food supplements. *Food Chemistry* 240: 767–774.
66. Lockley, A. K., and R. G. Bardsley. 2000. DNA-based methods for food authentication. *Trends in Food Science and Technology* 11.2: 67-77.
67. Lugon, M.D., dos Santos, P.H.D., Oliveira, P.V. et al. 2021. Is Your Açaí Really from Amazon? Using DNA Barcoding to Authenticate Commercial Products. *Food Analytical Methods* 14: 1559-1566.
68. Madesis, P., Ganopoulos, I., Bosmali, I., Tsaftaris, A. 2012. Barcode High Resolution Melting analysis for forensic uses in nuts: A case study on allergenic hazelnuts (*Corylus avellana*). *Food Research International 50*: 351–360.
69. Martins-Lopes, P., Gomes, S., Pereira, L., Guedes-Pinto, H. 2013. Molecular markers for food traceability. *Food Technology Biotechnology* 51: 198–207.

70. Maskova, E., Paulickova, I., Rysova, J., and Gabrovska, D. 2012. Evidence for wheat, rye, and barley presence in gluten free foods by PCR method—Comparison with Elisa method. *Czech Journal of Food Sciences* 29: 45–50.
71. Meusnier, I., Singer, G. A., Landry, J. F., Hickey, D. A., Hebert, P. D., and Hajibabaei, M. 2008. A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics* 9: 214.
72. Meyer, C.P., Paulay, G. 2005. DNA Barcoding: error rates based on comprehensive sampling. *PLOS Biology* 3: e422.
73. Mower, J. P., Touzet, P., Gummow, J., Delph, L. F., and Palmer, J. D. 2007. Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. *BMC Evolutionary Biology* 7: 135.
74. Nanney, D. L. 1982. Genes and phenes. *Tetrahymena Bioscience* 32: 783–788.
75. National Council of Educational Research and Training (NCERT). 2015. Biology Textbook For Class XI. *Regional Books*: India.
76. Nei, M. & Kumar, S. 2000. Molecular evolution and phylogenetics. *Oxford University Press.*
77. Nicolè, S., Barcaccia, G., Erickson, D.L., Kress, W.J., Lucchin, M. 2013. The coding region of the UFGT gene is a source of diagnostic SNP markers that allow single-locus DNA genotyping for the assessment of cultivar identity and ancestry in grapevine (Vitis vinifera L.). *BMC Research Notes* 6: 502.
78. Nicolè, S., Erickson, D.L., Ambrosi, D., Bellucci, E., Lucchin, M., Papa, R., Kress, W.J., Barcaccia, G. 2011. Biodiversity studies in Phaseolus species by DNA barcoding. *Genome* 54: 529–545.
79. Nicolè, S., Negrisolo, E., Eccher, G., Mantovani, R., Patarnello, T., Erickson, D.L., Kress, W.J., Barcaccia, G. 2012. DNA barcoding as a reliable method for the authentication of commercial seafood products. *Food Technology Biotechnology* 50: 387–389.
80. Pace, N. R. 1997. A molecular view of microbial diversity and the biosphere. *Science* 276: 734–740.
81. Pafundo, S., Agrimonti, C., Maestri, E., & Marmiroli, N. 2007. Applicability of SCAR markers to food genomics: Olive oil traceability. *Journal of Agricultural and Food Chemistry* 55: 6052–6059.
82. Prendini, L. 2005. Comment on ‘‘Identifying spiders through DNA barcodes’’. *CanJZool* 83: 498–504.
83. Rasmussen, R. S., Morrissey, M. T., and Hebert, P. D. N. 2009. DNA barcoding of commercially important salmon and trout species Oncorhynchus and Salmon from North America. *Journal of Agricultural and Food Chemistry* 57: 8379–8385.
84. Ratnasingham, S. and Hebert, P.D.N. 2007. BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes 7*: 355–364.
85. Reed, G.H., Wittwer, C.T. 2004. Sensitivity and specificity of single-nucleotide polymorphism scanning by high-resolution melting analysis. *Clinical Chemistry* 50 (10): 1748-1754
86. Saccone, C., DeCarla, G., Gissi, C., Pesole, G. & Reynes, A. 1999. Evolutionary genomics in the Metazoa: the mitochondrial DNA as a model system. *Gene* 238: 195–210.
87. Scarano, D., Rao, R. 2014. DNA markers for food products authentication. *Diversity* 6: 579–596.
88. Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.A., et al. 2012 Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA

barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the United States of America 109: 6241–6246*
89. Seethapathy, G. S., Ganesh, D., Kumar, J. U. S., Senthilkumar, U., Newmaster, S. G., Ragupathy, S., Ravikanth, G. 2015. Assessing product adulteration in natural health products for laxative yielding plants, Cassia, Senna, and Chamaecrista, in Southern India using DNA barcoding. *International Journal of Legal Medicine*,129(4): 693–700.
90. Shaw, J.; Lickey, E.B.; Schilling, E.E.; Small, R.L. *2007.* Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. *American Journal of Botany 94: 275–288*
91. Sickel, W., Ankenbrand, M. J., Grimmer, G., Holzschuh, A., Hartel, S., Lanzen, J., et al. 2015. Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. *BMC Ecology* 15: 20.
92. Simmons, R. B. & Weller, S. J. 2001 Utility and evolution of cytochrome b in insects. *Molecular Phylogenetics Evolution* 20: 196– 210.
93. Sonnenberg, R., Nolte, A. 2007. An evaluation of LSU rDNA D1-D2 sequences for their use in species identification. *Frontiers in Zoology 4:6*
94. Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A. 2007. Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Research* 35(3): e14-e14.
95. Trewick, S. A. 2000. Mitochondrial DNA sequences support allozyme evidence for cryptic radiation of New Zealand *Peripatoides* (Onychophora). *Molecular Ecology* 9: 269–282.
96. Uthicke, S., Byrne, M., and Conand, C. 2010. Genetic barcoding of commercial *Bêchedemer* species (*Echinodermata: Holothuroidea*). *Molecular Ecology Resources* 10: 634–646.
97. Valentini, A., Pompanon, F., and Taberlet, P. 2009. DNA barcoding for ecologists. *Trends In Ecology and Evolution* 24: 110–117.
98. Vincent, S.,Vian, J. M. & Carlotti, M. P. 2000. Partial sequencing of the cytochrome oxidase-b subunit gene I: A tool for the identification of European species of blow flies for post mortem interval estimation. *Journal of Forensic Sciences* 45: 820–823.
99. Violino, S., Antonucci, F., Pallottino, F., Cecchini, C., Figorilli, S. 2019. Food traceability: a term map analysis basic review. *European Food Research and Technology* 245:2089–2099
100. Vogler, A.P., Monaghan, M.T. 2006. Recent advances in DNA taxonomy. *Journal of Zoology and Systems of Evolution Research* 45:1–10.
101. Wares, J. P. and Cunningham, C. W. 2001. Phylogeography and historical ecology of the North Atlantic intertidal. *Evolution* 12: 2455–2469.
102. Wiemers, M., Fiedler, K. 2007. Does the DNA barcoding gap exist? —A case study in blue butterflies (*Lepidoptera: Lycaenidae*). *Frontiers in Zoology 4*: 8.
103. Wilson, K. H. 1995. Molecular biology as a tool for taxonomy. *Clinical Infectious Diseases* 20(Suppl.): 192–208.
104. Wong, L. L., Peatman, E., Lu, J., Kucuktas, H., He, S., Zhou, C., et al. 2011. DNA barcoding of catfish: Species authentication and phylogenetic assessment. *Public Library of Science* One 6: e17812.
105. Zhang, D.-X. & Hewitt, G. M. 1997. Assessment of the universality and utility of a set of conserved mitochondrial primers in insects. *Insect Molecular Biology* 6: 143–150.