CrossMark

REVIEW

Advances in DNA metabarcoding for food and wildlife forensic species identification

Martijn Staats[1] · Alfred J. Arulandhu[1] · Barbara Gravendeel[2] · Arne Holst-Jensen[3] ·
Ingrid Scholtens[1] · Tamara Peelen[4] · Theo W. Prins[1] · Esther Kok[1]

Abstract Species identification using DNA barcodes has
been widely adopted by forensic scientists as an effective mo-
lecular tool for tracking adulterations in food and for analysing
samples from alleged wildlife crime incidents. DNA
barcoding is an approach that involves sequencing of short
DNA sequences from standardized regions and comparison
to a reference database as a molecular diagnostic tool in spe-
cies identification. In recent years, remarkable progress has
been made towards developing DNA metabarcoding strate-
gies, which involves next-generation sequencing of DNA
barcodes for the simultaneous detection of multiple species
in complex samples. Metabarcoding strategies can be used
in processed materials containing highly degraded DNA e.g.
for the identification of endangered and hazardous species in
traditional medicine. This review aims to provide insight into
advances of plant and animal DNA barcoding and highlights
current practices and recent developments for DNA
metabarcoding of food and wildlife forensic samples from a
practical point of view. Special emphasis is placed on new
developments for identifying species listed in the
Convention on International Trade of Endangered Species
(CITES) appendices for which reliable methods for species
identification may signal and/or prevent illegal trade.
Current technological developments and challenges of DNA
metabarcoding for forensic scientists will be assessed in the
light of stakeholders' needs.

Introduction

Genetic identification of species plays a key role in the
investigation of illegal trade of protected or endangered
wildlife [1] and in the detection of species mislabelling
and fraud in the food industry [2]. Currently, DNA
barcoding is an established molecular technique that is
used for differentiating and assigning taxonomy to species
using standardized short DNA sequences (Box 1).
Application of DNA barcoding for food authentication
has gained much attention because of food safety con-
cerns, including incorrect food labelling, food substitu-
tions or food contamination [3–5]. DNA barcoding has
been effective in the traceability of many processed food
products in particular seafood and meat products [2]. For
instance, DNA barcoding has made an impact by demon-
strating widespread mislabelling or substitution of fish
and seafood products in markets and restaurants in New
York (USA) and Canada [4, 5]. Proper identification of
species present in food and food supplements is of vital
importance to protect consumers against potential food
adulteration, ingredient mislabelling or food poisoning.
Given its utility, DNA barcoding is being used by the
US Food and Drug Administration as a replacement for

✉ Martijn Staats
    martijn.staats@wur.nl

1   RIKILT Wageningen UR, P.O. Box 230, 6700
    AE Wageningen, The Netherlands

2   Naturalis Biodiversity Center, Sylviusweg 72, P.O. Box 9517,
    Leiden, The Netherlands

3   Norwegian Veterinary Institute, Ullevaalsveien 68, P.O. Box 750,
    Sentrum 0106, Oslo, Norway

4   Dutch Customs Laboratory, Kingsfordweg 1, 1043
    GN Amsterdam, The Netherlands

the time-consuming technique of protein isoelectric focusing for fish and fish products [6].

Another established application of DNA barcoding to forensic science is in investigations of wildlife crimes such as illegal collection and trade of flora and fauna. More than 35,000 species of flora and fauna are categorized as endangered by the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). Protected species are listed in Appendices I, II and III, according to how severe a certain population is threatened by extinction [7]. Besides the regulated legal trade, a significant portion of the trade in endangered flora and fauna is illegal. In the European Union (EU), enforcement is mainly focused at the borders, where illegally imported wildlife products, plants or animals will be seized by customs and CITES authorities. The mailings on the EU-TWIX (European Union Trade in Wildlife Information eXchange; www.eutwix.org) network of wildlife-protecting enforcement bodies are very clear; seizures of wildlife and products containing wildlife are everyday practice. In some cases, the species identification of seized specimens is not very difficult, because the specific morphological characteristics can be readily observed, though often requiring taxonomic expertise for decisive identification. Identification will be more difficult when only parts of an animal or plant without distinctive morphological characteristics are present, or when plant or animal parts have been pulverized and have become ingredients of food supplements (e.g. traditional medicines, TMs). Currently, CITES lists species encompassing a wide diversity of species of terrestrial plants such as cycads, cacti and orchids, in addition to vertebrates such as fish, amphibians, reptiles, birds and mammals, and invertebrates such as lobsters, crabs and corals [8]. Customs laboratories will obviously benefit from applying standardized, fast and reliable methods when dealing with samples of which no a priori knowledge on the species composition is known. It is these benefits that have made DNA barcoding the method of choice for customs laboratories when trying to establish the presence of biological material from endangered species within processed products [9–12].

A complicating aspect for DNA barcoding in the analysis of food supplements such as TMs and other mixed products is that they are composed of more than one ingredient. Such samples often contain multiple species that can only be efficiently analysed if multiple DNA barcode templates can be sequenced in parallel, something that next-generation sequencing (NGS) technologies do effectively [13]. Current NGS platforms yield millions of DNA reads in a relatively short period of time, and the sequencers' performance improves every year [14]. NGS combined with DNA barcoding is referred to as metabarcoding [15]. Metabarcoding thus uses universal PCR primers to mass-amplify one or more taxonomically informative targets. The prefix 'meta' refers to the collection of barcode sequences from different species. The general strategy consists of (1) extracting DNA from food or (wildlife) forensic samples, (2) amplifying a specific DNA barcode or other target region of taxonomic value, (3) sequencing the corresponding DNA amplicons using NGS technology, (4) analysing the sequences using appropriate bioinformatics pipelines, (5) identifying the species in the sample from which DNA has been extracted and (6) screening for CITES species among these [16, 17]. Metabarcoding has been applied in many diverse environmental samples, such as faeces [15], soil [18], marine water [19] and bulk samples of tropical arthropods [20]. However, there are only a few published applications of metabarcoding to food and (wildlife) forensic samples. Coghlan et al. [21] demonstrated the power of metabarcoding in detecting species in complex traditional Chinese medicine (TCM) samples presented in the form of powders, crystals, capsules, tablets and herbal tea. Their screening revealed that some of the TCM samples contained CITES listed species, including the Asiatic black bear (*Ursus thibetanus*) and the Saiga antelope (*Saiga tatarica*), as well as unlisted ingredients, and potentially toxic and allergenic plants. Cheng et al. [22] performed metabarcoding analyses on well-defined TCM preparations based on a six-herb formula named Liuwei Dihuang Wan, which is widely used in China. They concluded that there are significant differences in quality and safety among commercial TCM preparations, as the unlisted species *Senna obtusifolia* was identified in some preparations that may potentially pose safety risks to consumers. Tillmar et al. [23] developed a metabarcoding method for the identification of species of mammals in human forensic tissues, with which the presence of low quantities of DNA from the genus *Canis* could be identified.

Although metabarcoding may seem easy to apply, researchers often face limitations in obtaining a representative assessment of species composition. First, different preprocessing conditions and production procedures from samples with different composition and matrices (e.g. TMs and other processed and complex products) may result in highly variable DNA quality and concentration. DNA integrity has a significant influence on the effectiveness of the metabarcoding and other molecular methodologies for species identification [22, 24, 25]. Secondly, while there are many bioinformatics methods available for the analysis of metabarcoding data, the discriminating power of these methods is directly related to prior choices on barcode marker and reference database composition [26, 27]. PCR bias caused by variable primer–template mismatches across species may limit the quantitative potential of DNA metabarcoding and may cause species to be missed [28, 29]. Furthermore, DNA metabarcoding wholly relies on the presence of high-quality barcode sequence reference databases that are based on good taxonomy and barcode coverage. The goal of this paper is to review the advancements and current practices of plant and animal metabarcoding, with an emphasis on complex food and forensic wildlife samples

for identifying, in particular, species listed by CITES. This effort is complementary to recent work focussing on metabarcoding for biodiversity assessments in environmental samples [17, 30–32] and an extension of the work on DNA barcoding of food and forensic samples [1–3, 33–36]. This overview will address the opportunities and challenges that must be faced to allow the customs laboratories and other routine laboratories to perform efficient and reliable metabarcoding analysis that can broadly identify any species present in a sample under investigation.

Box 1. DNA barcoding and the International Barcode of Life project (iBOL)

DNA barcoding is a rapid method of differentiating and assigning taxonomy to species using standardized short DNA sequences. For animals, the most commonly used sequence is a 658-bp (base pair) region of the mitochondrial cytochrome c oxidase subunit I gene (COI, COX1, CO1). DNA barcoding allows for fast, reliable, automatable and cost-effective species identification by users with little or no taxonomic experience [37]. Identifications are usually made by comparing unknown sequences against known species DNA barcodes via alignment searching (BLAST) [38] or distance-based tree construction [39].

A suitable barcode for identification at the species level should be sufficiently variable between species (typically at least 3 % difference between closely related species but this may vary amongst taxonomic groups) and display either low or no intraspecific variations. Also, barcodes should be widely studied for a large number of species to enable comparison of the nucleotide sequence from an unknown sample with reference sequences in a database. Accurate species identification wholly relies on the taxonomic coverage of barcodes in a reference database. If the query sequence lacks a conspecific (belonging to the same species) target sequence in the database, species-level barcoding-based identification of the query will fail. Instead, the closest matches in the database may be identified and the sample barcode scored as a "new" taxon (operational taxonomic unit, OTU). From a practical point of view, therefore, DNA barcoding requires a comprehensive reference database. Such reference data sets are being assembled by the barcoding campaigns initiated by the International Barcode Of Life project (iBOL; www.ibol.org), resulting in considerably improved species coverage for target taxa of such DNA barcoding campaigns [40]. Official barcode sequences generated by the iBOL initiatives are deposited and organized in the Barcode Of Life Data (BOLD) Systems (http://boldsystems.org; [41]). BOLD is a large-scale and rigorously curated DNA barcode storage database, and most of the sequence information contained within BOLD has been derived from voucher specimens with authoritative taxonomic identifications. Barcoding campaigns focussing on fish, birds, mammals, insects and fungi have been initiated e.g. the Fish Barcode of Life Initiative (FISH-BOL, www.fishbol.org), the Marine Barcode Of Life Initiative (MarBOL, www.marinebarcoding.org), the Shark Barcode Of Life project (SharkBOL; www.sharkbol.org) and the Barcode of Wildlife Project (BWP; www.barcodeofwildlife.org). For plants there are initiatives to barcode e.g. the world's tree species in TreeBOL and grasses and grass-like plants in GrassBOL.

Barcodes and a variety of alternative taxonomically informative genes that have been generated from general scientific research are deposited in the International Nucleotide Sequence Database Collaboration (INSDC) and can be used for taxonomic assignment in barcoding studies. The iBOL initiative aims to create a database of 5 million standardized DNA sequences, which can be used to identify 500,000 species, by 2015.

Scientific literature on the utility of DNA barcoding in the recognition, discrimination and discovery of plant and animal species has been reviewed extensively by Savolainen et al. [42], Kress and Erickson [43], Bucklin et al. [44], Hollingsworth et al. [26], Fazekas et al. [45], Ortea et al. [33], Nicolè et al. [34], Bhargava and Sharma [46], Kvist [47] and Sandionigi et al. [48].

DNA extraction and DNA integrity

The initial sample preparation and extraction step in the analysis of DNA from food products is probably the most crucial step in the process of species identification in complex forensic samples. This step can be very difficult to standardize and optimize because of the complexity and diversity of the matrices encountered, each presenting different problems. For instance, it can be difficult to ensure that a representative sample is obtained from heterogeneous samples that are composed of many ingredients (e.g. TMs), and in such cases sufficient homogenization is particularly critical prior to DNA extraction. Forensic samples, such as food samples and TM may contain only very low amounts of DNA or contain ingredients that have been subjected to various treatments during the production process (e.g. cooking, high pressure, pH modification, grinding or drying), which may cause the DNA to be highly degraded [22, 49–51]. Furthermore, failure to eliminate potential inhibitory components and interfering substances from the material under investigation (e.g. protein, lipids, polyphenols, polysaccharides) may severely influence PCR analysis. Needless to say, any factor that may contribute to downstream bias needs to be minimized.

Different DNA extraction methods which can be used for analysis of forensic samples are now available; extraction is either based on in-house developed protocols or commercially available kits. Commercial kits offer a means for standardizing DNA extraction from forensic samples, as the protocol can be easily implemented in any laboratory. However, in many laboratories user-specific protocols have been developed to improve DNA extraction efficiency on a case-by-case basis. DNA extraction using cetyltrimethylammonium bromide (CTAB) extraction buffer combined with additional silica or a resin-based purification step are efficient for a wide range of plants and plant-derived products, in particular for separation of polysaccharides from DNA [25, 51, 52]. Ivanova et al. [53] developed a cost-efficient and automation-friendly DNA extraction protocol for animal tissues that consists of a tissue lysis step (SDS and proteinase K) followed by silica-based purification of DNA using inexpensive glass fibre filtration plates. The latter method has been used to process thousands of animal species at the Canadian Centre for DNA Barcoding (CCDB) as part of the iBOL initiative. Despite these efforts in

standardizing the DNA extraction method, the most suitable method is generally strongly dependent on the matrix, and there is no "universal" method that could be used for all food and (wildlife) forensic samples [49].

As suspect samples may often contain degraded DNA, it is a requirement that metabarcoding methods are able to identify species on the basis of short DNA sequences that may still be present in highly processed materials [51]. In such forensic samples, DNA degradation often prevents the amplification of PCR fragments longer than approximately 300 bp [24, 25, 54, 55]. The use of shorter barcode regions, so-called mini-barcodes, may overcome this problem. Owing to their reduced size, mini-barcodes are often amplified with higher efficiency in degraded samples than standard, full-length barcodes, which are typically 650–900 bp in length [25, 56]. On the other hand, the rate of taxonomic discrimination is generally positively correlated with the length of the mini-barcode. The use of universal mini-barcodes that will only allow identification of taxa above the species level, as a result of saturation of the taxonomic discrimination, should generally be avoided unless identification at the genus or family level is warranted.

Animal DNA barcodes and mini-barcodes

For animals, the standard barcode is a 658-bp region in the gene encoding mitochondrial cytochrome *c* oxidase I (COI or COX1, CO1) [57]. COI has long been used in animal molecular systematics to study relationships of closely related species because of its high level of interspecific variation [58]. Its popularity within the barcoding community is clearly reflected in the large public databases such as National Centre for Biotechnology Information (NCBI) GenBank (www.ncbi.nlm.nih.gov/genbank) and BOLD (Box 1). Universal primer sets for amplifying the COI barcode across major taxonomic groups have been developed by Ivanova et al. [59], and primer cocktails have been reported that are effective in fish, mammals, amphibians and reptiles (Table 1). A good discriminatory power in the identification of birds (98–100 % identification success rate [39]), fish (93–98 % identification success rate [64]), spiders (100 % identification success rate [78]), butterflies (97.9 % identification success rate [63]) and reptiles (72.7–100 % identification success rate [61]) has been shown for the COI barcode.

Despite its proven effectiveness, COI is not always suitable and effective for identifying all animal species. For endangered organisms such as sea snails (the mollusc class Gastropoda) and corals the COI barcoding region and other mitochondrial markers were found to offer insufficient resolution to allow for reliable discrimination between closely related species [79–82]. Using a DNA metabarcoding approach, Elbrecht et al. [29] demonstrated that species may go undetected in complex artificial mixtures of freshwater

invertebrate taxa because of universal COI primer–template mismatches. The use of group-specific primers or alternative degenerate primers may prevent species from being missed using COI [28, 29].

The traceability of mammalian meat including meat of ranched and hunted game species heavily relies on the use of the mitochondrial cytochrome *b* (cyt*b*) region [2, 83]. The choice of cyt*b* instead of COI is due mainly to practical reasons. The early availability of universal primers for cyt*b* [63, 78], long before the use of COI became popular, led to the deposition of several thousand cyt*b* sequences of a large range of edible mammalian species in public databases. Thus, its use became well established. Nonetheless, DNA barcoding based on COI has also proven effective in the identification of edible meat, including bush meat species [2, 83–86]. The FishTrace consortium (www.fishtrace.org) has promoted the use of cyt*b* through the development of universal cyt*b* primers for teleost fish species and the release of validated sequence data of many hundreds of European marine fish species [69].

Additional activities have taken place in finding suitable short DNA regions and related PCR primers for barcoding of species in widely diverse food and forensic samples, but so far no true mini-barcode standard has been adopted. Efforts in designing short broad coverage COI barcodes (i.e. mini-barcodes) to accommodate identification of a diversity of animal species in samples with degraded DNA has proven to be difficult. The use of the 130-bp COI mini-barcode primers designed by Meusnier et al. [24] has been limited [87], because the priming sites in the COI gene used in the mini-barcode design are not sufficiently conserved to cover a broad range of taxa [27, 66]. Leray et al. [83] have adopted a thorough approach and used the COI barcodes provided by the Moorea BIOCODE project, an "All Taxa Biotic Inventory" (www.mooreabiocode.org), consisting of more than 64,000 sequences across all phyla to design conserved universal COI mini-barcoding primers to target a 313-bp region. The newly designed primers were reported to perform well across metazoan diversity, with a higher success rate than the versatile primer sets traditionally used for DNA barcoding, i.e. the "Folmer primers" HCO2198 and LCO1490 [60] (Table 1).

Mitochondrial cyt*b*, 12S and 16S rRNA genes are the most commonly used genetic markers for species discrimination in degraded samples [88]. Universal primers for the amplification of short regions of cyt*b* have been developed for various animal taxa [70, 71]. Their use has been demonstrated in different problematic forensic samples that may contain degraded DNA including hair shafts, bones, feathers and meat products [70, 83].

Mini-barcodes based on the 12S and 16S rDNA mitochondrial genes have recently been demonstrated by several studies to be suited for identifying a wide range of animal species in environmental samples [89] and processed food and wildlife forensic products including TMs [21, 23, 73, 76]. The 12S and

Table 1 Non-exhaustive list of primers for amplifying animal DNA barcodes and mini barcodes

DNA marker	Target taxonomic group	Primer name	Primer sequences (5′–3′)	Amplicon length (bp)	Remark	Reference
COI	Various phyla	LCO1490 HCO2198	GGTCAACAAATCATAAAGATATTGG TAAACTTCAGGGTGACCAAAAAATCA	648		Folmer et al. [60]
COI	Reptiles	RepCOI-F RepCOI-R	TNTTMTCAACNAACCACAAAGA ACTTCTGGRTGKCCAAARAATCA	664		Nagy et al. [61]
COI-1	Birds	BirdF1 BirdR1	TTCTCCAACCACAAAGACATTGGCAC ACGTGGGAGATAATTCCAAATCCTG	648	Forward Reverse	Hebert et al. [39]
COI	Insects and amphibians	LepF1 LepR1	ATTCAACCAATCATAAAGATATTGG TAAACTTCTGGATGTCCAAAAAATCA	648		Hebert et al. [62]
		MLepF1 MLepR1	GCTTTCCCACGAATAAATAATA (use with LepR1) CCTGTTTCCAGCTCCATTTTC (use with LepF1)			Hajibabaei et al. [63]
COI-2	Mammals, fish, reptiles and amphibians	LepF1_t1 VF1_t1	TGTAAAACGACGGCCAGTATTCAACCAATCATAAAGATATTGG TGTAAAACGACGGCCAGTTCTCAACCAACCACAAAGACATTGG	648	M13-tailed cocktail; mix ratio 1:1:1:3:1:1:1:3	Ivanova et al. [59]
		VF1d_t1	TGTAAAACGACGGCCAGTTCTCAACCAACCACAARGAYATYGG			
		VF1i_t1	TGTAAAACGACGGCCAGTTCTCAACCAACCAIAAIGAIATIGG			
		LepR1_t1	CAGGAAACAGCTATGACTAAACTTCTGGATGTCCAAAAAATCA			
		VR1d_t1	CAGGAAACAGCTATGACTAGACTTCTGGGTGGCCRAARAAYCA			
		VR1_t1	CAGGAAACAGCTATGACTAGACTTCTGGGTGGCCAAAGAATCA			
		VR1i_t1	CAGGAAACAGCTATGACTAGACTTCTGGGTGICCIAAIAAICA			
COI-3	Fish and mammals	VF2_t1 FishF2_t1	TGTAAAACGACGGCCAGTCAACCAACCACAAAGACATTGGCAC TGTAAAACGACGGCCAGTGACTAATCATAAAGATATCGGCAC		M13-tailed cocktail; mix ratio 1:1:1:1	Ivanova et al. [59]
		FishR2_t1	CAGGAAACAGCTATGACACTTCAGGGTGACCGAAGAATCAGAA			
		FR1d_t1	CAGGAAACAGCTATGACACCTCAGGGTGTCCGAARAAYCARAA			
COI	Sharks	FishR2 Shark-int	ACTTCAGGGTGACCGAAGAATCAGAA ATCTTTGGTGCATGAGCAGGAATAGT	550		Ward et al. [64]
COI	Echinodermata phylum	COIceF COIceR	ACTGCCCACGCCCTAGTAATGATATTTTTTATGGTNATGCC TCGTGTGTCTACGTCCATTCCTACTGTRAACATRTG	>550		Hoareau and Boissin [65]
COI	Universal animal mini-barcode	mlCOIintF jgHCO2198	GGWACWGGWTGAACWGTWTAYCCYCC TAIACYTCIGGRTGICCRAARAAYCA	313		Leray et al. [66] Geller et al. [67]
COI	Universal animal mini-barcode	Uni-MinibarR1 Uni-MinibarF1	GAAAATCATAATGAAGGCATGAGC TCCACTAATCACAARGATATTGGTAC	130		Meusnier et al. [24]
cytb	Universal mammal	L14724 H15915	CGAAGCTTGATATGAAAAACCATCGTTG AACTGCAGTCATCTCCGGTTTACAAGAC	1140	Full-length cytb	Irwin et al. [68]
cytb	Universal fish	FishcytB-F CytB1-5R	ACCACCGTTGTTATTCAACTACAAGAAC GGTCTTTGTAGGAGAAGTATGGGTGGAA	750	cytb-5′ fragment	Sevilla et al. [69]

Table 1 (continued)

DNA marker	Target taxonomic group	Primer name	Primer sequences (5'–3')	Amplicon length (bp)	Remark	Reference
cytb	Universal vertebrate animal mini-barcode	L14816 / H15173	CCATCCAACATCTCAGCATGATGAAA / CCCCTCGAATGATATTTGTCCTCA	357		Parson et al. [70]
cytb	Universal animal mini-barcode	L14841 / H15149	AAAAAGCTTCCATCCAACATCTCAGCATGATGAAA / AAACTGCAGCCCCTCAGAATGATATTTGTCCTCA	307		Kocher et al. [71]
16S	Universal animal	16sar-L / 16sbr-H	CGCCTGTTTATCAAAAACAT / CCGGTCTGAACTCAGATCACGT	500–650	Forward Reverse	Palumbi [72]
16S	Universal animal mini-barcode	16S-forward / 16S-reverse	AYAAGACGAGAAGACCC / GATTGCGCTGTTATTCC	250		Sarri et al. [73]
16S	Fish, cephalopods and crustaceans	16S1F / 16S2R	GACGAKAAGACCCTA / CGCTGTTATCCCTADRGTAACT	250		Deagle et al. [74]
16S	Universal mammal mini-barcode	16S-forward / 16S-reverse	GACGAGAAGACCCTATGGAGC / TCCGAGGTCACCCAACCTCCG	100		Tillmar et al. [23]
16S	Universal vertebrate mini-barcode	L2513 / H2714	GCCTGTTTACCAAAAACATCAC / CTCCATAGGGTCTTCTCGTCTT	244		Kitano et al. [75]
16S	Universal animal mini-barcode	16S-HF / 16S-HR1 / 16S-HR2	ATAACACGAGAAGACCCT / CCCACGGTCGCCCCAAC / CCCGCGGTCGCCCCAAC	80–125		Horreo et al. [76]
12S	Universal vertebrate mini-barcode	L1085 / H1259	CCCAAACTGGGATTAGATACCC / GTTTGCTGAAGATGGCGGTA	215		Kitano et al. [75]
12S	Universal vertebrate mini-barcode	12SV5-F / 12SV5B2	TTAGATACCCCACTATGC / TAGAACAGGCTCCTCTAG	98		Riaz et al. [77]

16S rDNA contain internal regions that are strongly conserved across taxa, suitable for designing universal primers, alternated with short hypervariable regions that are species-specific. Sarri et al. [73] developed an approximately 250-bp barcode marker (Table 1) which allowed for the successful amplification of the 16S region across different sample types (e.g. cheese, processed meats, frozen fish fillets) and the correct identification of a wide range of animals in food products, including fishes, birds, reptiles, crustaceans and European mammals. Kitano et al. [75] developed 12S and 16S mini-barcodes for the identification of a large number of vertebrates (mammals, birds, reptiles, amphibians and fish). Similarly, Karlsson and Holmlund [88] used short 12S and 16S regions to identify a total of 28 different mammals including domestic and game species.

Plant DNA barcodes and mini-barcodes

In plants, the COI gene and other mitochondrial regions are a poor choice for species identification because the mitochondrial genome in plants has evolved too slowly to allow it to be used for DNA barcoding [90]. Research into a COI analogue in plants has focused on the plastid genome, but the selection of a standard plant barcode marker has been complicated by the trade-off that arises between the high requirements of universality and high variability among plants [43]. So far, no single barcode marker has been found that is expected to discriminate all of the more than 200,000 species of plants. The Consortium for the Barcode Of Life (CBOL) plant working group has opted for the use of a core set of two (rbcL and matK) coding sequences from plastids as the "core" DNA barcode (Table 2) [26]. The rbcL barcode consists of a 599-bp region at the 5' end of the gene. It is easy to amplify, sequence and align in most land plants, but it has only modest discriminatory power. Newmaster et al. [102] analysed over 10,000 rbcL sequences from GenBank and found that rbcL could discriminate samples in approximately 85 % of pairwise comparisons of congeneric species. The matK barcode region consists of a ca. 841-bp region at the centre of the gene, which is one of the most rapidly evolving regions of the plastid genome. The matK is perhaps the closest plant analogue to the COI animal barcode [103]. Ogden et al. [104] developed a single-nucleotide polymorphism (SNP) genotyping approach based on matK DNA barcodes to distinguish between traded timber products of Ramin (Gonystylus) species, which are all CITES protected. Unfortunately, matK can be difficult to amplify, particularly in non-angiosperms, because of the lack of sufficiently universal primers [92, 94].

The two most widely used supplementary loci are the nuclear ribosomal ITS (nrITS) [95] and plastid intergenic spacer psbA-trnH region [105]. The nrITS region had previously been discounted as a standard DNA barcode because of concerns over paralogy and the presence of putative pseudogenes which led to sequencing difficulties in many plant groups [106]. However, the increased resolution of nrITS over plastid DNA barcodes in many studies suggested that it should continue to be explored as part of the plant DNA barcode [95, 107]. Some authors have noted that just using a subset of the ribosomal cassette (nrITS2) can lead to greater amplification and sequencing success compared to the entire nrITS region [95]. By testing the discriminating ability of nrITS2 in more than 6600 medicinal plants and closely related samples, Chen et al. [95] found that the rate of successful identification was 92.7 % at the species level, and they proposed that the nrITS2 region should be the standard barcode for investigating forensic samples containing medicinal plants. Newmaster et al. [102] used rbcL and nrITS2 DNA barcodes to highlight species substitution and contamination in herbal products.

The psbA-trnH region is straightforward to amplify across land plants and is one of the more variable intergenic spacers in plants [108]. It has been used successfully in a range of barcoding studies [109, 110]. One of the main concerns associated with the use of psbA-trnH as a standard barcode is the premature termination of sequence reads by mononucleotide repeats leading to unidirectional reads in up to 30 % of sequences [111].

In plants, the design of suitable universal mini-barcode markers has proven difficult. The length constraints to allow working with highly degraded DNA severely limit the taxonomic resolution of mini-barcodes compared to that of the 500- to 800-bp-long standardized barcodes (rbcL, matK). Primers for the amplification of an approximately 180-bp region of chloroplast rbcL have been used, but in most cases this system only allows the identification of families, not genera or species [112]. Little [56] evaluated a variety of rbcL primers in silico and found the discriminatory power of the best rbcL mini-barcode to be less than 38.2 %. Taberlet et al. [100] used the chloroplast tRNA^{Leu} (UAA) intron sequences [trnL (UAA), 254–767 bp] and a shorter fragment of this intron (the P6-loop, 10–143 bp) for identifying plant species in processed food and ancient permafrost samples. The number of trnL (UAA) intron sequences available in databases is high, by far the most numerous among non-coding chloroplast DNA sequences. The trnL (UAA) region had overall low resolution. However, Taberlet et al. [100] concluded that only closely related species are not resolved and that the region can effectively be used to identify commonly eaten plants (e.g. potato, tomato, maize, but not almond). The trnL (UAA) has been extensively used in the food industry [113], forensic sciences [12] and diet studies based on faeces [114].

For some applications, a plant mini-barcode with relatively modest discriminatory power at the genus or higher taxonomic level can be useful. For example, it is often an entire genus or family that is listed by CITES, rather than individual plant species. For many plant families listed by CITES (e.g. Cycadaceae, Orchidaceae, Cactaceae, Euphorbia) identification to a larger group is therefore all that is required. This does

Table 2 Non-exhaustive list of primers for amplifying plant DNA barcodes and mini-barcodes

DNA marker	Target taxonomic group	Primer name	Primer sequences (5'–3')	Amplicon length (bp)	Remark	Reference
*rbc*L	Universal plant	rbcL a-F	ATGTCACCACAAACAGAGACTAAAGC	654		Levin et al. [91]
		rbcL a-R	GTAAAATCAAGTCCACCRCG			Kress and Erickson [92]
*mat*K	Angiosperms & Gymnosperms	matK-KIM1R	ACCCAGTCCATCTGGAAATCTTGGTTC	656–889	Forward	Fazekas et al. [45]
		matK-KIM3F	CGTACAGTACTTTTGTGTTTACGAG		Reverse	
*mat*K	Angiosperms & Gymnosperms	matK-390f	CGATCTATTCATTCAATATTTC	656–889	Forward	Cuénoud et al. [93]
		matK-1326r	TCTAGCACACGAAAGTCGAAGT		Reverse	
*mat*K	Gymnosperms	NY552F	CTGGATYCAAGATGCTCCTT	656–889	Forward	Fazekas et al. [45]
		NY1150R	GGTCTTTGAGAAGAACGGAGA		Reverse	
*mat*K	Gymnosperms	matKpkF4	CCCTATTCTATTCAYCCNGA	656–889	Forward	Fazekas et al. [94]
		matKpkR1	CGTATCGTGCTTTTRTGYTT		Reverse	
nrITS2	Universal plant	S2F	ATGCGATACTTGGTGTGAAT		Forward	Chen et al. [95]
		ITS4	TCCTCCGCTTATTGATATGC		Reverse	White et al. [96]
nrITS2	Universal plant	S2F	ATGCGATACTTGGTGTGAAT	160–320	Forward	Chen et al. [95]
		S3R	GACGCTTCTCCAGACTACAAT		Reverse	
nrITS	Universal angiosperm	17SE	ACGAATTCATGGTCCGGTGAAGTGTTCG	800	Forward	Sun et al. [97]
		26SE	TAGAATTCCCCGGTTCGCTCGCCGTTAC		Reverse	
*trn*H-*psb*A	Universal plant	psbAF	GTTATGCATGAACGTAATGCTC	264–792		Sang et al. [98]
		trnH2	CGCGCATGGTGGATTCACAATCC			Tate and Simpson [99]
*trn*L (UAA)	Universal plant mini-barcode	g	GGGCAATCCTGAGCCAA	10–143	p-loop region of *trn*L	Taberlet et al. [100]
		h	CCATTGAGTCTCTGCACCTATC			
*trn*L (UAA)	Universal plant	c	CGAAATCGGTAGACGCTACG	767		Taberlet et al. [101]
		d	GGGGATAGAGGGACTTGAAC			
*trn*L (UAA)	Universal plant mini-barcode	c	CGAAATCGGTAGACGCTACG	250		Taberlet et al. [101]
		h	CCATTGAGTCTCTGCACCTATC			Taberlet et al. [100]

not apply to all illegally traded plant genera though, such as tree ferns of the genus *Cibotium*, of which only *C. regale* is legally protected. In such cases, an alternative approach could be to design species-specific mini-barcodes to distinguish between closely related species from the CITES listed species, as was done for instance for *Rauvolfia serpentina* [12].

Box 2. Approximate number of sequences of DNA barcodes and other taxonomically informative genes available in GenBank (December 2014). GenBank sequences were retrieved with a query of the sequence annotations using the nucleotide database e.g. 'COI' OR 'cytochrome *c* oxidase' AND eukaryote'. The query headers were then downloaded and additionally filtered using the GNU/Linux command line tools (e.g. awk and grep). The number of unique genera and species were estimated from the sequence annotations and should be considered only as an approximation.

Sequencing of DNA barcodes using NGS technology

There are many excellent reviews on NGS platforms, and also their fundamentals and broad characteristics are described elsewhere [13, 14, 115, 116]. We will focus on the important steps in the NGS workflow and only provide a brief overview of NGS technologies relevant for DNA metabarcoding.

Early DNA metabarcoding studies have employed the 454 pyrosequencing technology of Roche because it was the first commercially available NGS system and because of its longer sequence read-outs allowing for a more informative fraction of DNA barcodes to be sequenced. Pyrosequencing has been used for DNA metabarcoding of raw materials of the diet of several animals [66, 114], environmental monitoring [15, 117, 118] and for analysing ancient DNA

Number of barcoding sequences deposited in GenBank

	COI	16S	cyt*b*	*mat*K	*rbc*L	*trn*L	*psb*A-*trn*H	nrITS
Approx. number of accessions	940,687	264,931	324,769	94,246	134,784	172,493	44,581	378,711
Approx. number of species	102,919	60,928	34,230	43,039	47,675	63,172	20,891	84,670
Approx. number of genera	30,923	21,691	10,822	8759	10,978	10,895	3836	14,338

Box 3. Benchtop next-generation sequencing system and their characteristics. Benchtop instruments are scaled-down, economical NGS platforms driven by the need for cheaper and faster sequencing, and which are suited for metabarcoding of typical food and forensic samples. The system specificities are listed with expected maximum performance by the beginning of 2016.

Instrument	Company	Machine run time (h)	Reads/run	Read length (base)	Output
454 GS Junior Plus[a]	Roche	18	70,000	ca. 700	70 Mb
MiniSeq[b]	Illumina®	24	44–50 million	2 × 150	6.6–7.5 Gb
MiSeq[c]	Illumina®	56	44–50 million	2 × 300	13.2–15 Gb
NextSeq 500[d]	Illumina®	29	Up to 800 million	2 × 150	100–120 Gb
Ion PGM™ System[e]	Ion Torrent™	7.3	4–5.5 million	400	1.2–2.0 Gb
Ion Proton™ System[f]	Ion Torrent™	4	60–80 million	200	Up to 10 Gb

[a] Adopted from http://454.com/products/gs-junior-plus-system/index.asp. Roche announced that 454 sequencers will be phased out in mid-2016

[b] adopted from http://www.illumina.com/systems/miniseq/specifications.html

[c] Adopted from http://www.illumina.com/systems/miseq/performance_specifications.html

[d] Adopted from http://www.illumina.com/systems/nextseq-sequencer/performance-specifications.html

[e] Adopted from https://tools.lifetechnologies.com/content/sfs/brochures/PGM-Specification-Sheet.pdf

[f] Adopted from https://tools.lifetechnologies.com/content/sfs/brochures/CO06326_Proton_Spec_Sheet_FHR.pdf

extracted from museum specimens [87]. The 454 technology is, however, no longer mainstream and Roche announced that 454 sequencers will be phased out in mid-2016.

Recently, benchtop sequencers have emerged that, owing to their compact format, lower set-up and running costs, and faster data turnaround times, have made NGS accessible for routine testing laboratories. The 454 GS Junior System (Roche), the MiSeq and MiniSeq (Illumina®), the NextSeq 500 (Illumina®), the Ion Proton™ System (Ion Torrrent™) and Ion PGM™ System (Ion Torrrent™) have sequencing capacities large enough for most metabarcoding projects (Box 3). Tillmar et al. [23] used to Roche 454 GS Junior system for the detection of animal species using the 16S rRNA gene. The same benchtop sequencer and the 454 GS-Titanium sequencer were used to identify plant and animal species in TMs [21, 22, 119]. Bertolini et al. [120] used the Ion Torrent PGM™ System for the identification of DNA from meat species using 12S and 16S rRNA genes.

The choice of NGS technology for DNA metabarcoding may depend on several parameters such as the barcode length, the number of barcodes used and the number of samples that need to be analysed.

An advantage of Illumina® sequencing is that sequencing data with very low error rates (greater than 0.1 %) are produced, compared to 454 and Ion Torrent™ sequencing [121]. The most common error types on the 454 and Ion Torrent™ platforms are insertions and deletions (indels), in particular when reading homopolymer regions. This results in an overall error rate of approximately 1.5 % [121, 122]. Sequencing errors can lead to spurious identification of species. Bertolini et al. [120] reported that when Ion Torrent data are quality filtered during downstream bioinformatics processing, the error rates do not introduce any bias that could prevent the correct assignment of meat species.

The high output combined with relatively short length have limited the use of Illumina sequencing technology mainly to profiling of bacterial communities using short 16S rDNA hypervariable regions [123, 124]. However, recent developments allowed the MiSeq platform to double the amount of output per flow cell by producing read lengths of 300 bp (Box 3). Because the Illumina platform can generate amplicon sequences in a paired-end format, paired reads can be directly matched and assembled into amplicons of up to ca. 550 bp.

This development has allowed the MiSeq sequencer to compete with 454 sequencing technology as it allows for generating sequence data from barcode regions with sufficient taxonomic resolution for animal and plant species identification.

An important step in the NGS workflow is to generate a library of the amplicons of interest. Fundamental for library construction is the modification of the DNA amplicons into a form that is compatible with the NGS platform to be used. The library is constructed by enzymatically ligating adapter sequences to the DNA amplicons or by adding them by PCR. The adaptors include specific sequences that are required for clonal amplification of the library on a solid surface (bead or glass slide). The choice of these adapter sequences is dictated by the NGS platform (Box 3). The adapter sequences may additionally contain a 6- to 10-nucleotide-long multiplex identifier (MID) that is used to pool amplicons from several independent samples in one run. MIDs are typically added to make more efficient use of the sequencing capacity of the NGS sequencers i.e. the number of reads generated by each NGS technology is usually higher than required per sample. Adapter sequences with different MIDs need to be used for each sample when multiple samples are sequenced in a single NGS experiment. The number of samples that can be pooled depends on (1) the number of available MIDs, (2) the

sequencing capacity of the NGS platform, (3) the number of amplicons per sample and (4) the required sequencing depth [125, 126]. After NGS, the resulting combined sequence data from different samples are subsequently sorted in silico by MID using bioinformatics tools.

Bioinformatics tools

Bioinformatics has played a crucial role in the advancement of metabarcoding. In recent years, many bioinformatics tools have been developed and are constantly being improved to efficiently and effectively perform various steps involved in the metabarcoding process. After obtaining NGS data, quality filtering is the first essential step, because it removes erroneous data that may otherwise potentially lead to misidentification of species. Sequencing errors introduced during NGS can be recognized because raw reads have predicted error probabilities for each base indicated by Phred quality scores. Sequence errors can be removed during quality filtering and trimming e.g. by truncating reads at the position where their quality begins to drop. A Phred score of 20, which corresponds to a 1 % error rate in base calling, is often used as a minimum threshold in quality filtering. Bokulich et al. [127] published guidelines for quality-filtering strategies to enable efficient extraction of high-quality data from Illumina amplicon sequencing data. In their studies on TMs, Coghlan et al. [21, 119] used the commercially available software Geneious [128]. Other software tools for quality filtering of reads include PRINSEQ [129] and Trimmomatic [130].

Following quality control, the sequences can either be directly matched to a reference library of DNA barcodes or processed further using clustering analysis. Clustering analysis is often performed to improve throughput by removing redundancy in the data such that the input can be used for the more computationally intensive analysis of assigning taxonomy. Clustering methods group reads into operational taxonomic units (OTUs) based on their similarity to other sequences in the samples, and from which representative or consensus sequences are selected. Commonly used clustering algorithms are CD-HIT [131], BlastClust [132] and UPARSE [133]. An OTU is commonly defined as a cluster of reads with 97 % similarity, which would be considered as belonging to a unique species according to the DNA barcoding standard [57]. However, the traditionally used 97 % similarity threshold is only an approximation. Sometimes two closely related species may have identical barcode sequences [134] or conversely single species may have two or more copies of a DNA barcode marker that differ by more than 3 % [135].

Next, tree-based methods and similarity-based methods are most commonly used for assigning query sequences to taxonomy. Tree-based methods assign query sequences to species on the basis of their membership of clusters (or clades) in a barcode tree. This approach is usually based on neighbour joining (NJ)

developed by Saitou and Nei [136], and is implemented in BOLD by Ratnasingham and Hebert [41]. The underlying assumption in NJ barcode matching is that distinct species form discrete clusters in an NJ tree [57]. For identification, query sequences are induced in the NJ tree to see which cluster they appear in. Similarity-based BLAST (Basic Local Alignment Search Tool [132]) is probably the most widely used method for classifying DNA sequences in practice. BLAST aligns the query sequence against those present in a selected target database using nearly exact matches of short nucleotide strings (e.g. 10 nucleotides). A similarity score is computed from the portion of the query aligned to the reference sequence. The reference sequence(s) with the highest similarity score is (are) presented along with an indication of the Expect value (E-value), which is the number of hits one can "expect" to see by chance when searching a database of a particular size.

A number of dedicated software pipelines exist that allow processing of metabarcoding data sets followed by taxonomic annotation, including jMOTU and Taxonerator [137], CLOTU [138], QIIME [139], Mothur [140] and UPARSE [133]. These software tools have been developed for studying microbial communities using the 16S rRNA gene fragment, but they can also be used for metabarcoding samples containing plants and animals [21, 141]. The HTS barcode checker pipeline is an application for automated processing of NGS data to determine whether these contain DNA barcodes obtained from species listed on the CITES appendices [16]. DNA metabarcodes are automatically converted into taxonomic identifications by matching with names on the CITES appendices. By inclusion of a blacklist and additional names database, the HTS barcode checker pipeline prevents false positives and resolves taxonomic heterogeneity.

In DNA metabarcoding, the availability of curated reference databases is of major importance to the assignment of sequences to species. A prerequisite is that reference database should contain accurate sequences that are correctly assigned to taxa with adequate sampling and taxon coverage to fully evaluate both the intraspecific and interspecific variations. Unbalanced representation of certain species, which is expected when dealing with CITES species, may greatly affect the analysis. Currently, there are many barcoding campaigns initiated by iBOL to generate DNA barcode data from well-identified and vouchered samples (Box 1). Worldwide sequencing efforts have already resulted in more than 2 million COI records from nearly 170,000 species in BOLD. The Barcode Index Number System (BINs) introduced by BOLD is an online framework that automatically clusters animal COI barcode sequences, generating a wiki Web page for each cluster [142]. Since clusters show high concordance with species, the framework can be used to verify species identifications as well as to document potential new animal species without taxonomic information. BOLD has already reached a good level of standardization and accuracy in terms of the identification of animals but the situation for plants is quite

different. The debate about the correct marker(s) to be used as universal barcode has led to a delay in the introduction of plant sequences in the BOLD database [26]. There is also valuable sequence data archived by the International Nucleotide Sequence Database Collaboration (INSDC; www.insdc.org), which, besides the COI region, is particularly extensive for nrITS and cyt*b* (Box 2). However, many of the existing INSDC sequences lack validation in the form of voucher information, making it difficult to detect and remove misidentified specimens or contaminated sequences. Currently the use of local curated reference data sets is often

preferred when DNA barcoding is used in plants []. Luo et al. [86] developed a custom DNA barcoding database for medicinal plant materials, and it accepts plastid DNA markers and nuclear nrITS regions as input (www.cuhk.edu.hk/icm/mmdbd.htm). Furthermore, an online identification module for herbal plant materials has been developed (www.tcmbarcode.cn), which is based around a selection of nrITS2 and *psb*A-*trn*H barcodes from selected medicinal species and their adulterants, substitutes and closely related species.

Non-exhaustive list of software available for DNA metabarcoding

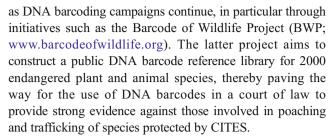|	Description	Reference
Software for quality filtering of reads		
PRINSEQ	Application for filtering, reformatting and quality trimming of metagenomic datasets. The software is publicly available through a user-friendly Web interface and as stand-alone version	Schmieder and Edwards [129] http://edwards.sdsu.edu/cgi-bin/prinseq/prinseq.cgi
Geneious	Commercially available suite of molecular tools	Kearse et al. [128] www.geneious.com
Trimmomatic	A flexible read trimming tool for Illumina NGS data	Bolger et al. [130] http://www.usadellab.org/cms/?page=trimmomatic
Software for cluster analysis of reads		
CD-HIT	A fast program for clustering of next-generation sequencing data. The software is publically available through a user-friendly interface and as stand-alone version	Fu et al. [131] http://weizhongli-lab.org/cdhit_suite/cgi-bin/index.cgi
BLASTclust	A program to make non-redundant sequence sets	Altschul et al. [132] http://toolkit.tuebingen.mpg.de/blastclust
Software for assigning reads to taxonomy		
BOLD identification	Species identification system of the Barcode of Life Data Systems (BOLD)	Ratnasingham and Hebert [41] http://www.boldsystems.org/
BLAST	The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program is publically available through a user-friendly Web interface and as stand-alone version	Altschul et al. [132] http://blast.ncbi.nlm.nih.gov/Blast.cgi
HTS-barcode-checker	A tool for automated detection of illegally traded species from high-throughput sequencing data	Lammers et al. [16] https://github.com/naturalis/HTS-barcode-checker
Software pipelines for DNA metabarcoding		
jMOTU and Taxonerator	Software for turning DNA barcode sequences into annotated OTUs	Jones et al. [137]
QIIME	Quantitative Insights Into Microbial Ecology: bioinformatics pipeline for microbiome analysis from raw DNA sequence data	Caporaso et al. [139] http://qiime.org/
CLOTU	Software for processing amplicon reads followed by taxonomic annotation	Kumar et al. [138]
UPARSE	Pipeline for clustering NGS amplicon reads into OTUs	Edgar et al. [133] http://drive5.com/uparse/
Mothur	Open-source, platform-independent, community-supported software for describing and comparing microbial communities	Schloss et al. [140] http://www.mothur.org/

Outlook

Next-generation sequencing of DNA barcodes, commonly referred to as DNA metabarcoding, is more and more becoming a standard approach for the simultaneous identification and detection of multiple species in complex samples. The approach is similar for both species identification to prevent food fraud and for tracing possible cases of illegal trade of CITES species. A large variety of informative barcodes and mini-barcodes in both the animal and plant area is available, potentially allowing for a clear-cut identification of species present in a sample of interest. However, comprehensive identification of (endangered and/or protected) species in complex forensic samples is not yet fully feasible at this moment. This is due to a number of reasons. In the first place, no truly universal DNA isolation method is available for all the different matrixes seized by the customs and CITES authorities. In-house-developed protocols or commercially available kits or a combination of both are typically assessed in an attempt to obtain amplifiable DNA from forensic samples, which will increase time and cost. In many cases, the poor success of extraction and PCR amplification of DNA from forensic samples hinders effective identification of species. Accordingly, systematic studies are needed to optimise DNA isolation methods and efficiency to satisfy the stakeholders' needs, which are to obtain a robust and rapid DNA isolation method that can be applied across a wide range of (wildlife) forensic samples and one that would maximize DNA purity and yield, whilst reducing any further DNA damage.

Secondly, forensic samples are often heavily processed and may contain severely fragmented DNA, thus hampering the ability to PCR amplify full-length barcodes. In such cases, mini-barcodes are often the only alternative, but these do not always provide species-level resolution, and truly universal primers for mini-barcode amplification have been found difficult to design. Universal primers should be used that minimize PCR bias caused by variable primer–template mismatches across species to ensure that all species can be detected [28, 29]. Several mini-barcodes have been proposed, but especially for plants no universal mini-barcode standard to provide species-level resolution has so far been adopted. The power of DNA metabarcoding is that a panel of different barcodes and mini-barcodes can efficiently be analysed in parallel. Such a strategy will provide improved resolution at the species level when some barcodes fail to resolve, while verifying species with multiple barcodes contributes to enhanced quality assurance.

Thirdly, the current underrepresentation of DNA barcodes from species protected under CITES and closely related species critically hampers their identification. This will improve

as DNA barcoding campaigns continue, in particular through initiatives such as the Barcode of Wildlife Project (BWP; www.barcodeofwildlife.org). The latter project aims to construct a public DNA barcode reference library for 2000 endangered plant and animal species, thereby paving the way for the use of DNA barcodes in a court of law to provide strong evidence against those involved in poaching and trafficking of species protected by CITES.

Finally, it will be necessary to develop and validate bioinformatics pipelines for the detection and identification of endangered species using DNA metabarcoding strategies. Several dedicated software tools have been developed, but there is a need to validate pipelines for clustering of reads into OTUs, using benchmarked algorithms for quality control, denoising, chimera removal and OTU picking.

Conclusion

The DNA metabarcoding approach holds great promise for detecting and identifying endangered plant and animal species in complex forensic samples. However, validation of the approach should be performed before DNA metabarcoding can be applied in a routine setup. By making use of DNA-barcoded reference species in well-characterized complex products or as internal controls in real-life samples, one can assess whether the DNA metabarcoding procedure is able to accurately and concurrently identify various target plant and animal species. Only when DNA metabarcoding has been demonstrated to be robust and transferable across laboratories can the method truly be implemented in routine testing. In that sense, we are just at the beginning of exploring the broad applications of DNA metabarcoding to reveal the composition of complex products in the light of, for instance, food fraud and the illegal trading of endangered plant and animal species.

Compliance with ethical standards

Conflicts of interest The authors have declared that no competing interests exist.

References

1. Iyengar A. Forensic DNA, analysis for animal protection and biodiversity conservation: a review. J Nat Conserv. 2014;22(3):195–205.

2. Galimberti A, De Mattia F, Losa A, Bruni I, Federici S, Casiraghi M, et al. DNA barcoding as a new tool for food traceability. Food Res Int. 2013;50(1):55–63.

3. Fajardo V, González I, Rojas M, García T, Martín R. A review of current PCR-based methodologies for the authentication of meats from game animal species. Trends Food Sci Technol. 2010;21(8):408–21.

4. Wong EH-K, Hanner RH. DNA barcoding detects market substitution in North American seafood. Food Res Int. 2008;41(8):828–37.

5. Hanner R, Becker S, Ivanova NV, Steinke D. FISH-BOL and seafood identification: geographically dispersed case studies reveal systemic market substitution across Canada. Mitochondrial DNA. 2011;22(sup1):106–22.

6. Yancy HF, Zemlak TS, Mason JA, Washington JD, Tenge BJ, Nguyen N-LT, et al. Potential use of DNA barcodes in regulatory science: applications of the Regulatory Fish Encyclopedia. J Food Protect. 2008;71(1):210–7.

7. CITES. CITES. 2015. Accessed 6 Apr 2016.

8. Speciesplus. Speciesplus. 2015. Accessed 15 Oct 2015.

9. Chen F, Chan H, Wong K-L, Wang J, Yu M-T, But P, et al. Authentication of Saussurea lappa, an endangered medicinal material, by ITS DNA and 5S rRNA sequencing. Planta Med. 2008;74(8):889–92.

10. Lahaye R, Van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, et al. DNA barcoding the floras of biodiversity hotspots. Proc Natl Acad Sci U S A. 2008;105(8):2923–8.

11. Gathier G, Niet T, Peelen T, Vugt RR, Eurlings M, Gravendeel B. Forensic identification of CITES protected slimming Cactus (Hoodia) using DNA barcoding. J Forensic Sci. 2013;58(6):1467–71.

12. Eurlings M, Lens F, Pakusza C, Peelen T, Wieringa JJ, Gravendeel B. Forensic identification of Indian snakeroot (*Rauvolfia serpentina* Benth. ex Kurz) using DNA barcoding. J Forensic Sci. 2013;58(3):822–30.

13. Shokralla S, Spall JL, Gibson JF, Hajibabaei M. Next-generation sequencing technologies for environmental DNA research. Mol Ecol. 2012;21(8):1794–805.

14. Glenn TC. Field guide to next-generation DNA sequencers. Mol Ecol Resour. 2011;11(5):759–69.

15. Ji Y, Ashton L, Pedley SM, Edwards DP, Tang Y, Nakamura A, et al. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. Ecol Lett. 2013;16(10):1245–57.

16. Lammers Y, Peelen T, Vos RA, Gravendeel B. The HTS barcode checker pipeline, a tool for automated detection of illegally traded species from high-throughput sequencing data. BMC Bioinformatics. 2014;15:44.

17. Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. Towards next-generation biodiversity assessment using DNA metabarcoding. Mol Ecol. 2012;21(8):2045–50.

18. Nielsen UN, Wall DH. The future of soil invertebrate communities in polar regions: different climate change responses in the Arctic and Antarctic? Ecol Lett. 2013;16(3):409–19.

19. Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, et al. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc Natl Acad Sci U S A. 2006;103(32):12115–20.

20. Gibson J, Shokralla S, Porter TM, King I, van Konynenburg S, Janzen DH, et al. Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. Proc Natl Acad Sci U S A. 2014;111(22):8007–12.

21. Coghlan ML, Haile J, Houston J, Murray DC, White NE, Moolhuijzen P, et al. Deep sequencing of plant and animal DNA contained within traditional Chinese medicines reveals legality issues and health safety concerns. PLoS Genet. 2012;8(4), e1002657.

22. Cheng X, Su X, Chen X, Zhao H, Bo C, Xu J, et al. Biological ingredient analysis of traditional Chinese medicine preparation based on high-throughput sequencing: the story for Liuwei Dihuang Wan. Sci Rep. 2014;4:5147.

23. Tillmar AO, Dell'Amico B, Welander J, Holmlund G. A universal method for species identification of mammals utilizing next generation sequencing for the analysis of DNA mixtures. PLoS One. 2013;8(12), e83761.

24. Meusnier I, Singer GA, Landry JF, Hickey DA, Hebert PD, Hajibabaei M. A universal DNA mini-barcode for biodiversity analysis. BMC Genomics. 2008;9:214.

25. Särkinen T, Staats M, Richardson JE, Cowan RS, Bakker FT. How to open the treasure chest? Optimising DNA extraction from herbarium specimens. PLoS One. 2012;7(8), e43808.

26. Hollingsworth PM, Graham SW, Little DP. Choosing and using a plant DNA barcode. PLoS One. 2011;6(5), e19254.

27. Ficetola GF, Coissac E, Zundel S, Riaz T, Shehzad W, Bessiere J, et al. An in silico approach for the evaluation of DNA barcodes. BMC Genomics. 2010;11:434.

28. Piñol J, Mir G, Gomez-Polo P, Agustí N. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. Mol Ecol Resour. 2015;15(4):819–30.

29. Elbrecht V, Leese F. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass–sequence relationships with an innovative metabarcoding protocol. PLoS One. 2015;10(7), e0130324.

30. Coissac E, Riaz T, Puillandre N. Bioinformatic challenges for DNA metabarcoding of plants and animals. Mol Ecol. 2012;21(8):1834–47.

31. Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH. Environmental DNA. Mol Ecol. 2012;21(8):1789–93.

32. Valentini A, Pompanon F, Taberlet P. DNA barcoding for ecologists. Trends Ecol Evol. 2009;24(2):110–7.

33. Ortea I, Pascoal A, Canas B, Gallardo JM, Barros-Velazquez J, Calo-Mata P. Food authentication of commercially-relevant shrimp and prawn species: from classical methods to foodomics. Electrophoresis. 2012;33(15):2201–11.

34. Nicolè S, Negrisolo E, Eccher G, Mantovani R, Patarnello T, Erickson DL, et al. DNA barcoding as a reliable method for the authentication of commercial seafood products. Food Technol Biotech. 2012;50:387–98.

35. Alacs EA, Georges A, FitzSimmons NN, Robertson J. DNA detective: a review of molecular approaches to wildlife forensics. Forensic Sci Med Pathol. 2009;6(3):180–94.

36. Veldman S, Otieno J, Gravendeel B, Andel Tv, Boer Hd. Conservation of endangered wild harvested medicinal plants: use of DNA barcoding. Novel plant bioresources: applications in food, medicine and cosmetics. 2014:81–8.

37. Hebert PD, Gregory TR. The promise of DNA barcoding for taxonomy. System Biol. 2005;54(5):852–9.

38. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.

39. Hebert PD, Stoeckle MY, Zemlak TS, Francis CM. Identification of birds through DNA barcodes. PLoS Biol. 2004;2:1657–63.

40. Kwong S, Srivathsan A, Meier R. An update on DNA barcoding: low species coverage and numerous unidentified sequences. Cladistics. 2012;28(6):639–644.

41. Ratnasingham S, Hebert PD. BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). Mol Ecol Notes. 2007;7(3):355–64.

42. Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R. Towards writing the encyclopedia of life: an introduction to DNA barcoding. Philos Trans R Soc Lond B Biol Sci. 2005;360(1462):1805–11.

43. Kress WJ, Erickson DL. DNA barcodes: genes, genomics, and bioinformatics. Proc Natl Acad Sci U S A. 2008;105(8):2761–2.

44. Bucklin A, Steinke D, Blanco-Bercial L. DNA barcoding of marine metazoa. Annu Rev Mar Sci. 2011;3:471–508.

45. Fazekas AJ, Kuzmina ML, Newmaster SG, Hollingsworth PM. DNA barcoding methods for land plants. Methods Mol Biol. 2012;858:223–52.

46. Bhargava M, Sharma A. DNA barcoding in plants: evolution and applications of in silico approaches and resources. Molec Phylogenet Evol. 2013;67(3):631–41.

47. Kvist S. Barcoding in the dark: a critical view of the sufficiency of zoological DNA barcoding databases and a plea for broader integration of taxonomic knowledge. Molec Phylogenet Evol. 2013;69(1):39–45.

48. Sandionigi A, Galimberti A, Labra M, Ferri E, Panunzi E, De Mattia F, et al. Analytical approaches for DNA barcoding data – how to find a way for plants? Plant Biosyst. 2012;146(4):805–13.

49. Bernardo GD, Gaudio SD, Galderisi U, Cascino A, Cipollaro M. Comparative evaluation of different DNA extraction procedures from food samples. Biotechnol Progr. 2007;23(2):297–301.

50. Fernandes TJ, Oliveira MBP, Mafra I. Tracing transgenic maize as affected by breadmaking process and raw material for the production of a traditional maize bread, broa. Food Chem. 2013;138(1):687–92.

51. Gryson N. Effect of food processing on plant DNA degradation and PCR-based GMO analysis: a review. Anal Bioanal Chem. 2010;396(6):2003–22.

52. Olexová L, Dovičovičová Ľ, Kuchta T. Comparison of three types of methods for the isolation of DNA from flours, biscuits and instant paps. Eur Food Res Technol. 2004;218(4):390–3.

53. Ivanova NV, Dewaard JR, Hebert PD. An inexpensive, automation-friendly protocol for recovering high-quality DNA. Mol Ecol Notes. 2006;6(4):998–1002.

54. Bergerová E, Hrnčírová Z, Stankovská M, Lopašovská M, Siekel P. Effect of thermal treatment on the amplification and quantification of transgenic and non-transgenic soybean and maize DNA. Food Anal Methods. 2010;3(3):211–8.

55. Rasmussen RS, Morrissey MT, Hebert PD. DNA barcoding of commercially important salmon and trout species (Oncorhynchus and Salmo) from North America. J Agric Food Chem. 2009;57(18):8379–85.

56. Little DP. A DNA mini-barcode for land plants. Mol Ecol Resour. 2014;14(3):437–46.

57. Hebert PD, Cywinska A, Ball SL. deWaard JR. Biological identifications through DNA barcodes. Proc R S Lond B Biol Sci. 2003;270(1512):313–21.

58. Trontelj P, Machino Y, Sket B. Phylogenetic and phylogeographic relationships in the crayfish genus Austropotamobius inferred from mitochondrial COI gene sequences. Molec Phylogenet Evol. 2005;34(1):212–26.

59. Ivanova NV, Zemlak TS, Hanner RH, Hebert PD. Universal primer cocktails for fish DNA barcoding. Mol Ecol Notes. 2007;7(4):544–8.

60. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol. 1994;3(5):294–9.

61. Nagy ZT, Sonet G, Glaw F, Vences M. First large-scale DNA barcoding assessment of reptiles in the biodiversity hotspot of Madagascar, based on newly designed COI primers. PLoS One. 2012;7(3), e34506.

62. Hebert PD, Penton EH, Burns JM, Janzen DH, Hallwachs W. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. Proc Natl Acad Sci U S A. 2004;101(41):14812–7.

63. Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PD. DNA barcodes distinguish species of tropical Lepidoptera. Proc Natl Acad Sci U S A. 2006;103(4):968–71.

64. Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PD. DNA barcoding Australia's fish species. Philos Trans R S B Biol Sci. 2005;360(1462):1847–57.

65. Hoareau TB, Boissin E. Design of phylum-specific hybrid primers for DNA barcoding: addressing the need for efficient COI amplification in the Echinodermata. Mol Ecol Resour. 2010;10(6):960–7.

66. Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, et al. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Front Zool. 2013;10(1):34.

67. Geller J, Meyer C, Parker M, Hawk H. Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. Mol Ecol Resour. 2013;13(5):851–61.

68. Irwin DM, Kocher TD, Wilson AC. Evolution of the cytochrome b gene of mammals. J Mol Evol. 1991;32(2):128–44.

69. Sevilla RG, Diez A, Norén M, Mouchel O, Jérôme M, Verrez-bagnis V, et al. Primers and polymerase chain reaction conditions for DNA barcoding teleost fish based on the mitochondrial cytochrome b and nuclear rhodopsin genes. Mol Ecol Notes. 2007;7(5):730–4.

70. Parson W, Pegoraro K, Niederstätter H, Föger M, Steinlechner M. Species identification by means of the cytochrome b gene. Int J Legal Med. 2000;114(1–2):23–8.

71. Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villablanca FX, et al. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc Natl Acad Sci U S A. 1989;86(16):6196–200.

72. Palumbi S. Simple fool's guide to PCR. Honolulu: University of Hawaii; 1991.

73. Sarri C, Stamatis C, Sarafidou T, Galara I, Godosopoulos V, Kolovos M, et al. A new set of 16S rRNA universal primers for identification of animal species. Food Control. 2014;43:35–41.

74. Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. Biol Lett. 2014;10(9).

75. Kitano T, Umetsu K, Tian W, Osawa M. Two universal primer sets for species identification among vertebrates. Int J Legal Med. 2007;121(5):423–7.

76. Horreo JL, Ardura A, Pola IG, Martinez JL, Garcia-Vazquez E. Universal primers for species authentication of animal foodstuff in a single polymerase chain reaction. J Sci Food Agric. 2013;93(2):354–61.

77. Riaz T, Shehzad W, Viari A, Pompanon F, Taberlet P, Coissac E. ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. Nucleic Acids Res. 2011;39(21):e145.

78. Barrett RD, Hebert PD. Identifying spiders through DNA barcodes. Can J Zool. 2005;83(3):481–91.

79. Shearer T, Van Oppen M, Romano S, Wörheide G. Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). Mol Ecol. 2002;11(12):2475–87.

80. Meyer CP, Paulay G. DNA barcoding: error rates based on comprehensive sampling. PLoS Biol. 2005;3(12), e422.

81. Wörheide G. Low variation in partial cytochrome oxidase subunit I (COI) mitochondrial sequences in the coralline demosponge Astrosclera willeyana across the Indo-Pacific. Mar Biol. 2006;148(5):907–12.

82. Huang D, Meier R, Todd PA, Chou LM. Slow mitochondrial COI sequence evolution at the base of the metazoan tree and its implications for DNA barcoding. J Mol Evol. 2008;66(2):167–74.

83. D'Amato ME, Alechine E, Cloete KW, Davison S, Corach D. Where is the game? Wild meat products authentication in South Africa: a case study. Investig Genet. 2013;4(1):6.

84. Cai Y, Zhang L, Shen F, Zhang W, Hou R, Yue B, et al. DNA barcoding of 18 species of Bovidae. Chin Sci Bull. 2011;56(2):164–8.

85. Bitanyi S, Bjornstad G, Ernest EM, Nesje M, Kusiluka LJ, Keyyu JD, et al. Species identification of Tanzanian antelopes using DNA barcoding. Mol Ecol Resour. 2011;11(3):442–9.

86. Luo A, Zhang A, Ho SY, Xu W, Zhang Y, Shi W, et al. Potential efficacy of mitochondrial genes for animal DNA barcoding: a case study using eutherian mammals. BMC Genomics. 2011;12:84.

87. Shokralla S, Zhou X, Janzen DH, Hallwachs W, Landry JF, Jacobus LM, et al. Pyrosequencing for mini-barcoding of fresh and old museum specimens. PLoS One. 2011;6(7), e21252.

88. Karlsson AO, Holmlund G. Identification of mammal species using species-specific DNA pyrosequencing. Forensic Sci Int. 2007;173(1):16–20.

89. Elbrecht V, Taberlet P, Dejean T, Valentini A, Usseglio-polatera P, Beisel J-N, et al. Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. Peer J Preprints. 2016;4: e1855v1.

90. Palmer JD, Herbon LA. Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. J Mol Evol. 1988;28(1–2):87–97.

91. Levin RA, Wagner WL, Hoch PC, Nepokroeff M, Pires JC, Zimmer EA, et al. Family-level relationships of Onagraceae based on chloroplast rbcL and ndhF data. Am J Bot. 2003;90(1):107–15.

92. Kress WJ, Erickson DL. A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region. PLoS One. 2007;2(6), e508.

93. Cuénoud P, Savolainen V, Chatrou LW, Powell M, Grayer RJ, Chase MW. Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid rbcL, atpB, and matK DNA sequences. Am J Bot. 2002;89(1):132–44.

94. Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, et al. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS One. 2008;3(7), e2802.

95. Chen S, Yao H, Han J, Liu C, Song J, Shi L, et al. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS One. 2010;5(1), e8613.

96. White TJ, Bruns T, Lee S, Taylor J. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. PCR Protocols: Guide Methods Appl. 1990;18:315–22.

97. Sun Y, Skinner D, Liang G, Hulbert S. Phylogenetic analysis of Sorghum and related taxa using internal transcribed spacers of nuclear ribosomal DNA. Theor Appl Genet. 1994;89(1):26–32.

98. Sang T, Crawford D, Stuessy T. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of Paeonia (Paeoniaceae). Am J Bot. 1997;84(9):1120–36.

99. Tate JA, Simpson BB. Paraphyly of Tarasa (Malvaceae) and diverse origins of the polyploid species. System Bot. 2003;28(4):723–37.

100. Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, et al. Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Res. 2007;35(3), e14.

101. Taberlet P, Gielly L, Pautou G, Bouvet J. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol. 1991;17(5):1105–9.

102. Newmaster SG, Grguric M, Shanmughanandhan D, Ramalingam S, Ragupathy S. DNA barcoding detects contamination and substitution in North American herbal products. BMC Med. 2013;11(1):222.

103. Hilu K, Liang H. The matK gene: sequence variation and application in plant systematics. Am J Bot. 1997;84(6):830–9.

104. Ogden R, McGough HN, Cowan RS, Chua L, Groves M, McEwing R. SNP-based method for the genetic identification of ramin Gonystylus spp. timber and products: applied research meeting CITES enforcement needs. Endanger Species Res. 2009;9(3):255–61.

105. Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci U S A. 2005;102(23):8369–74.

106. Group CPW, Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, et al. A DNA barcode for land plants. Proc Natl Acad Sci U S A. 2009;106(31):12794–7.

107. Li D-Z, Gao L-M, Li H-T, Wang H, Ge X-J, Liu J-Q, et al. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. Proc Natl Acad Sci U S A. 2011;108(49):19641–6.

108. Shaw J, Lickey EB, Schilling EE, Small RL. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am J Bot. 2007;94(3):275–88.

109. Gonzalez MA, Baraloto C, Engel J, Mori SA, Pétronelli P, Riéra B, et al. Identification of Amazonian trees with DNA barcodes. PLoS One. 2009;4(10), e7483.

110. Kress WJ, Erickson DL, Jones FA, Swenson NG, Perez R, Sanjur O, et al. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. Proc Natl Acad Sci U S A. 2009;106(44):18621–6.

111. Devey DS, Chase MW, Clarkson JJ. A stuttering start to plant DNA barcoding: microsatellites present a previously overlooked problem in non-coding plastid regions. Taxon. 2009;58(1):7–15.

112. Poinar HN, Hofreiter M, Spaulding WG, Martin PS, Stankiewicz BA, Bland H, et al. Molecular coproscopy: dung and diet of the extinct ground sloth Nothrotheriops shastensis. Science. 1998;281(5375):402–6.

113. Faria M, Magalhães A, Nunes M, Oliveira M. High resolution melting of trnL amplicons in fruit juices authentication. Food Control. 2013;33(1):136–41.

114. De Barba M, Miquel C, Boyer F, Mercier C, Rioux D, Coissac E, et al. DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. Mol Ecol Resour. 2014;14(2):306–23.

115. Quail M, Smith ME, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers. BMC Genomics. 2012;13(1):341.

116. Metzker ML. Sequencing technologies-the next generation. Nat Rev Genet. 2010;11(1):31–46.

117. Hajibabaei M, Shokralla S, Zhou X, Singer GA, Baird DJ. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PLoS One. 2011;6(4), e17497.

118. Carew ME, Pettigrove VJ, Metzeling L, Hoffmann AA. Environmental monitoring using next generation sequencing: rapid identification of macroinvertebrate bioindicator species. Front Zool. 2013;10(1):45.

119. Coghlan ML, Maker G, Crighton E, Haile J, Murray DC, White NE, et al. Combined DNA, toxicological and heavy metal analyses provides an auditing toolkit to improve pharmacovigilance of traditional Chinese medicine (TCM). Sci Rep. 2015;5.

120. Bertolini F, Ghionda MC, D'Alessandro E, Geraci C, Chiofalo V, Fontanesi L. A next generation semiconductor based sequencing approach for the identification of meat species in DNA mixtures. PLoS One. 2015;10(4), e0121701.

121. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, et al. Performance comparison of benchtop high-

throughput sequencing platforms. Nat Biotechnol. 2012;30(5): 434–9.

122. Salipante SJ, Kawashima T, Rosenthal C, Hoogestraat DR, Cummings LA, Sengupta DJ, et al. Performance comparison of Illumina and ion torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. Appl Environ Microbiol. 2014;80(24):7583–91.

123. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 2012;6(8):1621–4.

124. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol. 2013;79(17):5112–20.

125. Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat Methods. 2008;5(3):235–7.

126. Smith AM, Heisler LE, Onge RPS, Farias-Hesson E, Wallace IM, Bodeau J, et al. Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. Nucleic Acids Res. 2010;38(13), e142.

127. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, et al. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Methods. 2013;10(1):57–9.

128. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647–9.

129. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011;27(6):863–4.

130. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014:btu170.

131. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28(23):3150–2.

132. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25(17):3389–402.

133. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods. 2013;10(10):996–8.

134. Stoeckle MY, Gamble CC, Kirpekar R, Young G, Ahmed S, Little DP. Commercial teas highlight plant DNA barcode identification successes and obstacles. Sci Rep. 2011;1(42).

135. Song J, Shi L, Li D, Sun Y, Niu Y, Chen Z, et al. Extensive pyrosequencing reveals frequent intra-genomic variations of internal transcribed spacer regions of nuclear ribosomal DNA. PLoS One. 2012;7(8), e43971.

136. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4(4):406–25.

137. Jones M, Ghoorah A, Blaxter M. jMOTU and taxonerator: turning DNA barcode sequences into annotated operational taxonomic units. PLoS One. 2011;6(4), e19259.

138. Kumar S, Carlsen T, Mevik B-H, Enger P, Blaalid R, Shalchian-Tabrizi K, et al. CLOTU: an online pipeline for processing and clustering of 454 amplicon reads into OTUs followed by taxonomic annotation. BMC Bioinformatics. 2011;12(1):182.

139. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7(5):335–6.

140. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75(23):7537–41.

141. Yu DW, Ji Y, Emerson BC, Wang X, Ye C, Yang C, et al. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods Ecol Evol. 2012;3(4):613–23.

142. Ratnasingham S, Hebert PD. A DNA-based registry for all animal species: The Barcode Index Number (BIN) System. PLoS One. 2013;8(7), e66213.