Int J Legal Med (2009) 123:395–401
DOI 10.1007/s00414-009-0356-5

ORIGINAL ARTICLE

Forensic botany: species identification of botanical trace evidence using a multigene barcoding approach

Gianmarco Ferri · Milena Alù · Beatrice Corradini · Giovanni Beduschi

Received: 26 February 2009 / Accepted: 12 May 2009 / Published online: 7 June 2009

Abstact Forensic botany can provide significant supporting evidence during criminal investigations. However, it is still an underutilized field of investigation with its most common application limited to identifying specific as well as suspected illegal plants. The ubiquitous presence of plant species can be useful in forensics, but the absence of an accurate identification system remains the major obstacle to the present inability to routinely and correctly identify trace botanical evidence. Many plant materials cannot be identified and differentiated to the species level by traditional morphological characteristics when botanical specimens are degraded and lack physical features. By taking advantage of a universal barcode system, DNA sequencing, and other biomolecular techniques used routinely in forensic investigations, two chloroplast DNA regions were evaluated for their use as "barcoding" markers for plant identification in the field of forensics. We therefore investigated the forensic use of two non-coding plastid regions, psbA-trnH and trnL-trnF, to create a multimarker system for species identification that could be useful throughout the plant kingdom. The sequences from 63 plants belonging to our local flora were submitted and registered on the GenBank database. Sequence comparison to set up the level of identification (species, genus, or family) through Blast algorithms allowed us to assess the suitability of this method. The results confirmed the effectiveness of our botanic universal multimarker assay in forensic investigations.

Keywords Species identification · Forensic botany · DNA barcoding · Plant DNA database

Introduction

The ubiquitous presence of plant species makes botanical trace evidence useful for many aspects of criminal investigations, but remains underused in forensics due to a lack of experience and botanical knowledge among investigators as well as the inability to routinely identify degraded specimens by morphology. Plant identification can help to determine a sample's geographic origin, provide links between crime scene and individuals, test alibis, ascertain the possession or trade in forbidden or endangered species, and more [1–8]. Even if the value of botanical trace evidence in criminal and civil cases has been clearly demonstrated and it is widely accepted as suitable scientific evidence by the courts, it remains ignored by many investigators. Morphology and anatomy rarely confirm the source of the plant, particularly for degraded and fragmented material. Taking advantage of DNA sequencing and other biomolecular techniques, identification of land plants becomes useful in forensics, reducing the expertise necessary in botanical identification and producing successful results in casework application. Analysis of trace botanical evidence is a developing forensic discipline and techniques based on molecular biology should support and complement traditional forensic botany in the identification of evidence. The problem of reproducibility and standardization, especially from degraded specimens, prevents the use of classical methods in routine forensic

Electronic supplementary material The online version of this article (doi:10.1007/s00414-009-0356-5) contains supplementary material, which is available to authorized users.

G. Ferri (✉) · M. Alù · B. Corradini · G. Beduschi
Department of Diagnostic and Laboratory Services and Legal Medicine, Section of Legal Medicine, University of Modena and Reggio Emilia, Modena, Italy
e-mail: gianmarco.ferri@unimore.it

investigations [9–15] but the use of genetic markers could have the potential to overcome shortcomings associated with traditional forensic botany, and it has been recently proposed in forensic applications [16–18].

Recently, the use of short orthologous standard DNA sequences as a tool for species identification with a standardized protocol, known as DNA barcoding, has been proposed and initiated to facilitate biodiversity and taxonomic studies and enhance forensic analyses [19–21]. The main purpose of DNA barcoding is to provide rapid and accurate identification of unidentified organisms whose DNA barcodes have already been registered in a sequence library (DNA database). In order for a region of DNA to be effective as a barcode, it must simultaneously contain enough variability to be informative for identification, be short enough to sequence in a single reaction, and contain conservative regions, which can be used to develop universal primers. A portion of the mitochondrial cytochrome *c* oxidase I gene sequence is currently being used as a universal barcode in certain animal groups and has been proposed also in the forensic field [21–23]. Based on the existing literature, there is currently no available universally usable region, and it is generally agreed that a multilocus approach based on plastid (chloroplast) data is the most effective strategy for species identification and recognition in plants [24–29]. A variety of loci have been recently suggested as potential DNA barcodes in plants, in both the nuclear and plastid genomes [28–33]. Our laboratory has selected two regions of the chloroplast genome, psbA-trnH and trnL-trnF, frequently enclosed in different barcoding selections, to quantify how well these alleged barcoding regions work for a regional flora and for assessing the utility of barcoding markers for forensic investigations.

The intergenic spacer psbA-trnH is one of the most variable non-coding regions of the plastid genome in angiosperms (in terms of having the highest percentages of variable sites), meaning that it could ensure high level of species discrimination [24, 28, 30–33]. The other intergenic spacer trnL-trnF has been found to be capable of detecting inter- and intraspecies variation in a range of botanical species and has been employed in several phylogenetic studies [34–36].

The ability of selected markers to resolve land plant identification was further evaluated by comparing known and unknown sequences with those from the nucleotide sequence database GenBank. To evaluate how well the present setup of the NCBI nucleotide database allows correct plant identification by applying its implemented sequence similarity search tools, identification matches using BLAST and MegaBLAST algorithms were performed using the nucleotide sequences obtained for each plant specimen.

Most potential users of this method are not botanical experts and so would need a quick, easy, and accurate system for identification. Even if identification to the species level is not always possible, identification to higher taxonomic ranks can be helpful in forensic investigations, as in many cases a relatively crude diagnosis could be acceptable. In many of the situations in which this method would be applied, the application of a broad species concept is accepted. Our efforts have therefore focused on developing a DNA-based identification system that, together with traditional morphological methods, provides criteria that progressively identify an unknown plant sample to a given taxonomic rank also by any non-specialist botanist.

Materials and methods

Samples were collected from 63 individual plants from 33 families and 53 genera, which usually grow in our region, and each species was confirmed by morphological identification analysis by a botanical expert. The set of taxa included angiosperms, gymnosperms, and bryophytes. The sequences analyzed for both markers (n. 124) were submitted to GenBank database (the plant species with the corresponding GenBank accession numbers are listed in Table S1). For each plant species, two different samples were collected to determine any intraspecies variation. The selected taxa represent some of the most common species in the landscape of the Emilia-Romagna region, obviously including those more likely to be encountered at random during forensic investigations (grasses, Pinaceae, oaks, olive, etc.) [13, 14, 16]. Different types of tissue were collected from the samples (seeds, roots, and leaves), and the starting material was transferred into a 1.5-mL reaction tube to perform DNA extraction.

Total genomic DNA was isolated from 0.5 to 1.0 g of the different types of tissue (stored in fresh or dried form) with QIAGEN DNeasy plant mini kit (Qiagen, Valencia, CA, USA) following the manufacturer's guidelines, except for the complete exclusion of liquid nitrogen used for tissue disruption and the addition of 1 M DTT. Primers were selected from the literature [24, 34], taking into account the ability of amplification on the widest range of plant taxa under standard PCR conditions and the ability of providing diagnostic information for most plant groups, without the need for species-specific optimisation. The primer sequences are given in Table 1. PCR amplification was performed under the same conditions for all samples.

Approximately 1–5 ng of template DNA was used in a 12.5–25 μL PCR reaction consisting of 1× reaction buffer (Applied Biosystems, Foster City, CA, USA), 2 mM $MgCl_2$, 0.2 μM of each dNTP in an equimolar ratio, 0.75 μM of each primer, and 0.75 units of TaqGold DNA

Table 1 List of universal primers for tested loci

Locus	Primer sequence (5′ to 3′)
psbA-trnH	F GTTATGCATGAACGTAATGCTC
	R CGCGCATGGTGGATTCACAATCC
trnL-trnF	F GGTTCAAGTCCCTCTATCCC
	R ATTTGAACTGGTGACACGAG

polymerase (5 U/μL; Applied Biosystems). Amplification was performed for 30 cycles on a ABI 9700 Thermal Cycler under the following conditions: an initial cycle at 94°C for 10 min, denaturation at 94°C for 30 s, annealing at 55°C for 30 s, extension at 72°C for 60 s, for 32 cycles with a final extension step at 72°C for 10 min. Samples that produced aspecific bands on a gel were repeated at more stringent conditions (56–58°C annealing temperatures).

PCR products were separated on 2% agarose gels containing 0.5 μg/mL of ethidium bromide. Products showing more than one band were subsequently re-amplified. Single PCR products were sequenced using both the forward and reverse PCR primers. All amplification products with correct fragment size were purified for sequencing using EXOSAP (exonuclease I and shrimp-alkaline phosphotase, USB, Cleveland, OH, USA) and subsequently used as the template in a 5 μL sequence reaction with BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems) according to the protocol. Sequencing reaction products were finally purified from residual dye terminators using Edge BioSystems columns (Gaithersburg, MD, USA) and analyzed with POP-4 performance optimized polymer (Applied Biosystems) on an ABI 3130 sequencer. Each locus was quantified for PCR amplification success, which was defined as the percentage recovery and successful sequencing of the selected locus for each species.

To compare the power of discrimination of the two loci psbA-trnH and trnL-trnF, DNA sequences of the 63 selected plants were aligned with the program Clustal W version 1.8 using the Jalview Java alignment editor, which enables fast viewing and editing of multiple, large sequences while analyzing them for a level of sequence divergence [37]. The gap creation penalty and gap extension penalty were 5 and 1 for trnL-trnF and psbA-trnH, respectively. Alignments at family or genera level were unambiguous. Molecular diversity indices (mean number of pair-wise differences, nucleotide diversity, and number of polymorphic sites/number of total sites) were computed with the Arlequin software to find the sequence variation among different taxonomic levels [38]. Sequences selected from the GenBank database were also included (Tables S2 and S3) to improve the statistical significance of the statistical analysis among species belonging to the same genus.

Known and unknown sequences were used as queries in the BLASTn and MEGABLAST search to assess identification matches with the available GenBank sequences [39–41]. We then estimated the accuracy of the two different sequence similarity-homology search algorithms for molecular based species diagnosis, evaluating the default settings. We also determined whether it was possible to detect the correct family or genus from the results of the input sequences, even when the exact species was not actually present in the database, but related species from the same genera, or different genera from the same family were.

Results and discussion

Species identification is a basic and very important emerging and underestimated task in forensic botany. In practical terms, forensic botany can present additional information in many forensic cases involving plant evidence that may be useful to link a suspect, a victim, or a vehicle to the crime scene. During the commission of outdoor crimes, plant material may be transferred from the crime scene to the victim or perpetrator, and this kind of trace may be probative due to restricted and specific geographical distribution. Plant re-colonization (succession) of land that has been disturbed will follow a specific set of pattern that can be useful to estimate the time of death. Identification of vegetable contents in stomach or partially plant digested in feces can support or disprove an alibi. Again, molecular botany can be used to identify primary or secondary crime scene, tracking drug distribution network (cannabis) and in insurance fraud case [2]. In our specific casework application, the identification of root found in the eye socket of the buried skull was useful to evaluate a minimum estimate of the burial time by botanical expert which at first was unable to identify the plant through the root fragment. In our experience, in real and simulated casework, the extraction starting with specimens of about 0.25 cm^2 in fresh or dried form (0.25 to 0.5 g of tissue) provide enough quantity and quality of DNA to obtain successful PCR amplification and sequencing. The high copy number of chloroplasts DNA leads to a successful analysis even from low and highly degraded DNA.

This research represents an attempt to show the functionality and usefulness of forensic genetic botanical species diagnosis with a DNA barcoding multimarker approach to successful discriminate between the plant species investigated. The accuracy of species identification was estimated by comparing the sequences of 63 native plant species analyzed for two markers, psbA-trnH and trnL-trnF, through GenBank leading toward a standardized multilocus solution. Up to now, no single region has been shown to have the ability to resolve species identification to

a higher taxonomic rank, and it is unlikely that other single markers exist that would [25–29].

Firstly, the study confirmed the success of our modified DNA extraction protocol, which excluded the liquid nitrogen step and also of the high degree of versatility in the applied set of primers. The selected regions psbA-trnH and trnL-trnF were successfully amplified for all the plants tested, showing a single band on a gel in all of the species investigated except in one case. The amplification of psbA-trnH region of the *Quercus robur* (Oak) resulted in a pattern of additional bands confirmed by multiple amplifications of different specimens from the same species. DNA dilution and/or stringent PCR condition were not enough to achieve amplification of a single fragment, and the subsequent sequence analysis was not performed. Isolation of the correct band or alternative primer design should resolve this sample. A high rate of amplification success was obtained from the currently proposed regions (for psbA-trnH nearly 100% amplification and for trnL-trnF 100% amplification). The size of the PCR products was found to be highly variable among selected samples particularly for the psbA-trnH marker and for intrafamily groups (Table S4). The PCR products ranging from 268 to 740 bp (average 444 bp) for the psbA-trnH and from 173 to 608 bp (average 399 bp) for the trnL-trnF intergenic spacer (IGS) even if most of the amplicons ranged between 300 and 500 bp and between 370 and 460 bp, respectively. As reported by other authors, the psbA-trnH showed higher sequence size length variation among plant species belonging to the same genera, principally due to the presence of insertions and deletions [24–26, 28]. For the marker trnL-trnF, size variation was observed especially between species belonging to different families but a slight variation also emerged between highly correlated species belonging to the same genera or family.

In some cases, we experienced difficulty in obtaining complete sequences mainly due to the existence of heteroplasmic regions that were observed in 18 samples for the psbA-trnH marker and in seven samples for the trnL-trnF IGS out of the 63 total species studied. This phenomenon has been confirmed by different authors especially for the psbA-trnH marker [24, 30–33]. The complete nucleotide sequences were successfully determined by analyzing both forward and reverse strands. For the *Gingko biloba* species, only a double heteroplasmic region prevented the complete sequencing of both strands for the psbA-trnH marker. In this case, as for mixed or contaminated trace evidences, single sequence can be obtained by cloning. Bever and Cimino [42] via PCR, cloning, and sequencing were able to obtain individual component sequences of a eight plants mixture. The universal primers' intrinsic property to allow amplification across a wider taxonomic range of species complicates the mixture identification and single sample selection, and isolation is possible only when differences between samples are large and significant. Mixture results and length heteroplasmy highlight the possible need for cloning protocol.

All the sequences showed a unique haplotype. We analyzed two samples for each of the 63 plant species recovered, and no intraspecies sequence variation was found for the two markers analyzed. The diversity indices were analyzed first at the family level (for all families with at least two species sampled) and then to improve the statistical significance of the analysis, at the genus level by adding other sequences selected from the GenBank database (Tables S2 and S3). The statistical values and diversity indices computed by Arlequin are shown in Tables S5 and S6. The range of sequence diversity between different species of the same family (number of polymorphic sites/number of total sites) ranged from a minimum of 7.7% within the grasses (Poaceae) to a maximum of 65% within the Asteraceae for the psbA-trnH IGS and from a minimum 3.5% within the Magnoliaceae to a maximum of 67% within the Brassicaceae for the trnL-trnF IGS. The values of sequence divergence were confirmed by different statistical tests (Table S5). The genetic distances between species within the same genus ranged from 2.8% within the *Magnolia* genus to 33% for the *Prunus* genus for psbA-trnH and from 1.7% for the genera *Populus* and *Magnolia* to 67% for the *Solanum* for the trnL-F IGS, reflecting the variance obtained by similar indices available in Table S6.

With the purpose of establishing taxonomic identification methods using sequence data, we evaluated the extent to which the present NCBI database setup allows users to reliably conduct routine plant identification with the sequence similarity and homology search tools BLAST and MEGABLAST [39]. The taxonomic identification levels of the match and the relative number of species per genus previously registered in GenBank were examined (Tables S7 and S8). The BLAST results are presented as an output file providing a list of similar sequences or optionally showing sequences in pair-wise alignment with the query. High-scoring segment pairs are presented, consisting of the query sequence and a database entry in combination with a similarity value reporting the significance of the match. The result also contains a list of database entries, which are sorted by decreasing sequence similarity, with the first entry defined as maximum scoring segment pair (MSP), which displays the best concordance with the query sequence. The output sequences can appear in different order, depending on the setup of the statistics selected, even if the default BLAST output format provides a percentage of base pair identity with the resulting best hit as well as the *E* value per pair-wise alignment. This primary step in the species identification of an unknown sample is

highly influenced by the number and the quality of the nucleotide sequences registered in the database as well by the bioinformatics support used for the management and search of sequence data (see text ESM 1).

With the corresponding species registered in the database, identification was achieved correctly as MSP with a best match having the best sequence similarity value with the corresponding species for 17 (85%) and 28 (61%) species for the psbA-trnH and trnL-trnF markers, respectively (Table S7; see text ESM 1).

As explained above, we emphasize that some technical difficulties in the sequence similarity database search, such as the influence of the registered sequence length and quality, as well as the number of ambiguities, affect the results and the success of species–species identification (see text ESM 1).

With the purpose of standardizing a method for the interpretation of these results and to evaluate when exact species returned as first best match, we organized the reference sequence results in the output file based on the percentage of identity (see if the exact species returned as the first best hit compared with the query). In all the cases of psbA-trnH, the program returned a single best match, and therefore, we obtain the identification of the exact species as the best option (see text ESM 1).

Looking at the best match with the query sequence with the current status of nucleotide database, we strongly recommend firstly considering the max identity value and then to evaluate the other statistical measurements. However, when a complete global reference library of DNA sequences, approved as an effective barcode region, will be available in the public domain, and universal data standards are applied to DNA barcode records, the scientific community will be able to obtain reliable information concerning the barcode sequence for the unknown plant specimens, and the problems concerning the interpretation of sequence search results will be completely resolved (see text ESM 1).

Our results show that if complete sequence data from the correct species are present in the GenBank dataset, the BLAST search shows them as best hits, whereas correct genus or family affiliation was also obtained when no data for the appropriate species have been published. Difficulties can occur when partial sequence coverage between query sequence and reference sequence exists. In these cases, the BLAST algorithm will place longer sequences first, as the most likely similarity in preference to a shorter one, even when the shorter one matches the input sequence exactly. When a correct species was not found in the database, the biological significance of the matching result strongly depended on the availability of closely related species. This result was then coupled with lower sequence similarity values (Table S8).

The effect of number of sequences available for a genus in GenBank on the incidence of unique identifications was statistically significant for both psbA-trnH and trnL-trnF. Success in identifying the correct species was clearly higher if the two loci were both analyzed. Among the 48 plant species having one or both markers in GenBank, it was possible to identify 30 samples (62.5%) at the correct species level, considering those having one or both markers with 100% of query coverage and those having one or both markers with a partial query coverage of the subject sequence. In four cases (8.3%), the query sequence showed a best correspondence both with the exact plant species and others within the same genera; therefore, it was possible to identify the sequence at higher taxonomic level (see text ESM 1).

It is straightforward that a multilocus approach for plants will be necessary to greatly improve identification ability and reliability, as one single marker may easily result in a misidentification due to the particular evolution of plant's genomes, the complexity of the plant kingdom, and the poor coverage of the databases [25–33]. As underlined by other authors [17, 43, 44], forensic botanical comparison by nucleotide sequence data can be hampered by the lack of appropriate DNA databases as while DNA sequence databases for many mitochondrial loci have been established for animal species diagnosis, and the use of genetic markers has been validated in forensic laboratories for species identification [21, 45, 46], scientists have a poor knowledge of the genomes of plants [47].

We considered the barcoding criteria necessary to develop a comparative method based on DNA sequencing to identify unknown botanical evidence. Up to now, the application of DNA barcoding in plants has been constrained by a lack of consensus regarding the most variable and practical DNA region, unlike animal species identification where the mitochondrial cytocrome *c* oxidase I represents an ideal marker. Therefore, we selected two plastid non-coding markers among those proposed by taxonomists, taking into account sequence recovery rates combined with the ability to discriminate between common flora. We developed a species identification method based on the combined amplification and sequencing of two markers useful for botanical species determination in the forensic and law enforcement fields. The two sets of universal primers tested on our local flora seem to be effective for almost all the plant species examined showing sufficient information for most species determination. Furthermore, we provide an estimate of the effectiveness of combined markers using the BLAST procedure to gain the identification of unknown botanical evidence.

These loci have been found to be able to detect the inter- and intraspecies variation in a range of botanical species with high individual resolution rates among different plant

groups. Considering all the species examined, with the combined use of psbA-trnH and trnL-trnF, we were able to resolve correctly monophyletic species in nearly 60% of cases (Tables S7 and S8). For the remaining samples, the identification was achieved at higher taxonomic levels. The availability of two or more markers in DNA-based plant identification will be necessary as noted in the literature and confirmed in this study because combining regions improved the proportion of species resolved within genus-level clades.

Difficulties and potential major drawbacks highlighted here include the large numbers of plant species globally (420,000 vs 50,000 species for animal vertebrates) [26], the poor coverage of species in the sequence databases, and the lack of authenticated reference DNA sequences data are likely to limit the application of this approach in botanical forensics for the time being.

This molecular identification system would reduce plant species identification to a set of simple PCR tests in a routine analysis based on the standardized technique reported here, which outlines every step while minimizing the expertise. The central concept in species recognition is to match the sequence of the evidence sample to a reference sequence through DNA sequence similarity searches. Even if identification to the species level is not always possible, the confident forensic identification of samples to a higher taxonomic rank may be informative for the direction of an investigation. The failure to incorporate botanical evidence in investigations is mainly due to the lack of knowledge about plants and specific techniques by the investigator. However, we show here that fairly successful results can be obtained in correct plant identification.

Given the development of the barcoding initiative, the associated standard protocols and their ability for processing thousands of references sequences in a short time, the future application of selected barcoding markers, together with new or existing loci, should lead to the application of a powerful molecular technique that meets forensic standards. The development of a global reference DNA database with highly authenticated sequence data of specific barcode genes based on validated international guidelines for the plant kingdom, as successfully developed in the animal context, is a key aim of the Barcode project and a fundamental resource for forensic requirements. Therefore, the perspective of an effective, easy, and reproducible analytical procedure will modify in a relevant manner the approach of the investigator team to the interpretation on botanical trace evidence leading to a major awareness of the importance of these kinds of samples.

Acknowledgment The authors would like to thank Francesca Ferrari of the Department of Diagnostic and Laboratory Service and Legal Medicine, Section of Legal Medicine, University of Modena and Reggio Emilia, for her efforts with this research. The authors would also like to thank George Busby of the Department of Zoology, University of Oxford, for his great involvement in manuscript preparation.

References

1. Coyle HM, Ladd C, Palmbach T, Lee HC (2001) The green revolution: botanical contributions to forensic and drug enforcement. Croat Med J 42(3):340–345
2. Coyle HM, Lee CL, Lee HC, Lin WY, Palmbach TM (2005) Forensic botany: using plant evidence to aid in forensic death investigation. Croat Med J 46(4):606–612
3. Quatrehomme G, Lacoste A, Bailet P, Grevin G, Ollier A (1997) Contribution of microscopic plant anatomy to postmortem bone dating. J Forensic Sci 42(1):140–143
4. Mildenhall DC (2006) An unusual appearance of a common pollen type indicates the scene of the crime. Forensic Sci Int 163:236–240
5. Mildenhall DC (2006) *Hypericum* pollen determines the presence of burglars at the scene of a crime: an example of forensic palynology. Forensic Sci Int 163:231–235
6. Bock JH, Norris DO (1997) Forensic botany: an under-utilized resource. J Forensic Sci 42(3):364–367
7. Yoon CK (1993) Botanical witness for the prosecution. Science 260:894–895
8. Dunbar M, Murphy TM (2009) DNA analysis of natural fiber rope. J Forensic Sci 54:1–6
9. Lee CL, Coyle HM, Lee HC (2007) Genetic analysis of individual seeds by amplified fragment length polymorphism method. Croat Med J 48:563–565
10. Congiu L, Chicca M, Cella R, Rossi R, Bernacchia G (2000) The use of random amplified polymorphic DNA (RAPD) marker to identify strawberry varieties: a forensic application. Mol Ecol 9:229–232
11. Bless C, Palmeter H, Wallace M (2006) Identification of *Acer rubrum* using amplified fragment length polymorphism. J Forensic Sci 51(1):31–38
12. Coyle MH (2003) An overview of DNA methods for the identification and individualization of marijuana. Croat Med J 44:315–321
13. Stambuk S, Sutlovic D, Bakaric P, Petricevic S, Andelinovic S (2007) Forensic botany: potential usefulness of microsatellite-based genotyping of Croatian Olive (Olea europaea L.) in forensic casework. Croat Med J 48:556–562
14. Craft KJ, Owens JD, Ashley MV (2007) Application of plant DNA markers in forensic botany: genetic comparison of *Quercus* evidence leaves to crime scene trees using microsatellites. Forensic Sci Int 165:64–70
15. Korpelainen H, Virtanen V (2003) DNA fingerprinting of mosses. J Forensic Sci 48:804–807
16. Ward J, Peakall R, Gilmore SR, Robertson J (2005) A molecular identification system for grasses: a novel technology for forensic botany. Forensic Sci Int 152:121–131
17. Tsai LC, Yu YC, Hsieh HM, Wang JC, Linacre A, Lee JCI (2006) Species identification using sequences of the trnL intron and the trnL-trnF IGS of chloroplast genome among popular plants in Taiwan. Forensic Sci Int 164:193–200
18. Tsai LC, Yu YC, Hsieh HM, Liu KL, Linacre A, Lee JCI (2008) Bidens identification using the non coding regions of chloroplast genome and nuclear ribosomal DNA. Forensic Sci Int Genet 2:35–40
19. Hebert PDN, Cywinska A, Ball SL, de Waard JR (2003) Biological identifications through DNA barcodes. Proc R Soc Lond B 270:313–321

20. Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopaedia of life: an introduction to DNA barcoding. Phil Trans R Soc Lond B 360:1805–1811
21. Dawnay N, Ogden R, McEwing R, Carvalho GR, Thorpe RS (2007) Validation of the barcoding gene COI for use in forensic genetic species identification. Forensic Sci Int 173:1–6
22. Hebert PDN, de Waard JR, Ratnasingham S (2003) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc R Soc Lond B 270:S96–S99
23. Ferri G, Alù M, Corradini B, Licata M, Beduschi G (2009) Species identification through DNA "barcodes". Genet Test Mol Biomarkers 13(3):421–426
24. Kress JW, Wurdack Kenneth J, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci USA 102(23):8369–8374
25. Chase MW, Salamin N, Wilkinson M, Dunwell JM, Prasad Kesanakurthi R, Haidar N, Savolainen V (2005) Land plants and DNA barcodes: short-term and long-term goals. Phil Trans R Soc Lond B 360:1889–1895
26. Cowan RS, Chase MW, Kress JW, Savolainen V (2006) 300 000 species to identify: problems, progress and prospects in DNA barcoding of land plants. Taxon 55:611–616
27. Hollingsworth PM (2008) DNA barcoding plants in biodiversity hot spots progress and outstanding questions. Heredity. doi:10.1038/hdy.2008.16
28. Chase MW, Cowan RS, Hollingsworth PM et al (2007) A proposal for a standardised protocol to barcode all land plants. Taxon 56:295–299
29. Pennisi E (2007) Wanted: a barcode for plants. Science 318:190–191
30. Kress WJ, Erickson DL (2007) A two locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region. PloS ONE 2(6):e508. doi:10.1371/journal.pone.0000508
31. Sass C, Little D, Stevenson DW, Specht CD (2007) DNA barcoding in the Cycadales: testing the potential of proposed barcoding markers for species identification of Cycads. PloS ONE 2(11):e1154. doi:10.1371/journal.pone.00154
32. Fazekas AJ, Burgess KS, Kesanakurti PR et al (2008) Multiple multilocus DNA barcodes from the Plastid Genome discriminate plant species equally well. PloS ONE 3(7):e2802. doi:10.1371/journal.pone.0002802
33. Newmaster SG, Fazekas J, Steeves AD, Janovec J (2008) Testing candidate plant barcode regions in the Myristicaceae. Mol Ecol Notes 8:480–490
34. Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol 17:1105–1109
35. Stech M, Quandt D, Frey W (2003) Molecular circumscription of the hornworts (Anthocerotophyta) based on the chloroplast DNA trnL-trnF region. J Plants Res 116:389–398
36. Taberlet P, Coissac E, Pompanon F et al (2007) Power and limitations of the chloroplast *trn*L (UAA) intron for plant barcoding. Nucleic Acids Res 35:1–8
37. Chenna R, Sugawara H, Koike T, Lopez R, Gibson T, Higgins DJ, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res 31:3497–3500
38. Excoffier L, Laval G, Schneider S (2005) Arlequin ver 3.0: an integrated software package for population genetics data analysis. Evol Bioinformatics Online 1:47–50
39. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (2008) GenBank. Nucleic Acids Res 36:25–30
40. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
41. McGinnis S, Madden T (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res 32:20–25
42. Bever R, Cimino M (2001) Forensic molecular botany: identification of plants from trace evidence. Proceedings of the Promega 12th International Symposium on Human Identification, Oct 9–12, Biloxi, MS, USA
43. Gemeinholzer B, Oberprieler C, Bachmann K (2006) Using GenBank data for plant identification: possibilities and limitations using the ITS 1 of Asteraceae species belonging to the tribes Lactuceae and Anthemideae. Taxon 55:173–187
44. Little DP, Stevenson DWM (2007) A comparison of algorithms for the identification of specimens using DNA barcodes:examples from gymnosperms. Cladistics 23:1–21
45. Parson W, Pegoraro K, Niederstätter H, Föger M (2000) Species identification by means of the cytochrome b gene. Int J Legal Med 114:23–28
46. Imaizumi K, Akutsu T, Myasaka S, Yoshino M (2007) Development of species identification tests targeting the 16S ribosomal RNA coding region in mitochondrial DNA. Int J of Legal Med 121:184–191
47. Crane PR (2004) Documenting plant diversity: unfinished business. Phil Trans R Soc Lond B 359:735–737