Forensic Science International: Genetics xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Forensic botany II, DNA barcode for land plants: Which markers after the international agreement?

G. Ferri [1,*], B. Corradini [1], F. Ferrari, A.L. Santunione, F. Palazzoli, M. Alu'

Dipartimento di Medicina Diagnostica, Clinica e di Sanità Pubblica, Struttura Complessa di Medicina Legale, Università di Modena e Reggio Emilia, Italy

ARTICLE INFO

Article history:

Keywords:
Forensic botany
Species identification
DNA barcode
BOLD_IDS
GenBank

ABSTRACT

The ambitious idea of using a short piece of DNA for large-scale species identification (DNA barcoding) is already a powerful tool for scientists and the application of this standard technique seems promising in a range of fields including forensic genetics. While DNA barcoding enjoyed a remarkable success for animal identification through cytochrome c oxidase I (COI) analysis, the attempts to identify a single barcode for plants remained a vain hope for a longtime. From the beginning, the Consortium for the Barcode of Life (CBOL) showed a lack of agreement on a core plant barcode, reflecting the diversity of viewpoints. Different research groups advocated various markers with divergent set of criteria until the recent publication by the CBOL–Plant Working Group. After a four-year effort, in 2009 the International Team concluded to agree on standard markers promoting a multilocus solution (rbcL and matK), with 70–75% of discrimination to the species level. In 2009 our group firstly proposed the broad application of DNA barcoding principles as a tool for identification of trace botanical evidence through the analysis of two chloroplast loci (trnH-psbA and trnL-trnF) in plant species belonging to local flora. Difficulties and drawbacks that were encountered included a poor coverage of species in specific databases and the lack of authenticated reference sequences for the selected markers. Successful preliminary results were obtained providing an approach to progressively identify unknown plant specimens to a given taxonomic rank, usable by any non-specialist botanist or in case of a shortage of taxonomic expertise. Now we considered mandatory to update and to compare our previous findings with the new selected plastid markers (matK + rbcL), taking into account forensic requirements.

Features of all the four loci (the two previously analyzed trnH-psbA + trnL-trnF and matK + rbcL) were compared singly and in multilocus solutions to assess the most suitable combination for forensic botany.

Based on obtained results, we recommend the adoption of a two-locus combination with rbcL + trnH-psbA plastid markers, which currently best satisfies forensic needs for botanical species identification.

1. Introduction

The ambitious idea of using a relatively short piece of DNA for large-scale species identification (named DNA barcoding) is already a powerful tool for scientists and the application of this standard technique seems promising in a widespread range of fields including forensic genetics [1–6]. While DNA barcoding enjoyed a remarkable success for animal identification through cytochrome c oxidase I (COI) analysis, the attempts to identify a single barcode for plants remained a vain hope for a long time. From the beginning, the *Consortium for the Barcode of Life* (*CBOL*) showed lack of agreement on a core plant barcode, reflecting the diversity of viewpoints [7–12]. A plethora of different recommendations led to spreading of resources and different research groups advocated different markers with divergent set of criteria, until the recent publication by the CBOL–Plant Working Group (CBOL-PWG).

After a four-year effort, in 2009 the International Team of 52 scientists concluded to agree on standard markers and the tortuous search for a universal barcode for plants has been a close call with a definitive loci selection [14].

The CBOL-PWG promoted a multilocus solution comprising portions of the plastid genes *rbcL* and *matK* as the core system for land plants identification, attaining the 70–75% of discrimination to the species level.

This percentage of success, lower than in animal identification, is a good enough starting point for standardization and is useful for specific applications where resolution to the species level is not

* Corresponding author at: Via del Pozzo, 71, Policlinico, 41124 Modena, Italy. Tel.: +39 059 4224876/4223030; fax: +34 059 4224450.
E-mail address: gianmarco.ferri@unimore.it (G. Ferri).
[1] Joint authorship: these authors contributed equally to this paper.

http://dx.doi.org/10.1016/j.fsigen.2014.10.005
1872-4973/

always required, as in forensic casework investigation [7]. DNA barcoding in plants presents challenges that are not encountered in animals [15] and given the logistic difficulties of undertaking identification of something like 380.000–400.000 land plant species, this solution still offers the opportunity to harness a powerful universal DNA-based identification method for plants. It should be also taken in consideration that plant species resolution is constrained by wide genetic variability and it seems unlikely that adding more markers would significantly improve the discriminatory power [7–9,14,16]. The gap between intra- and interspecific genetic distance is not always marked in plants and polyploidy, hybridization and apomixes preclude the application of a single species concept.

The *rbcL* locus offers a high level of recoverability and a good but not outstanding discrimination power whereas *matK* offers higher species resolution but requires further development particularly as for the extent of universal primers. This combination may represent a pragmatic solution to the complex trade-off between cross-species application, sequence quality, power of discrimination and cost.

Nevertheless, the Working Group also argued that, besides the proposed core barcode, the employment of supplementary loci meeting the criteria of universality, sequence quality and discrimination power required by the Barcoding Project should be considered [14].

The non-coding plastid region *tnrH-psbA* is another strongly supported barcode candidate proposed by several laboratories as a third barcode marker, but despite its strong potential, it suffered from technical problems that may require manual sequencing editing [14].

Within the forensic context, it should be taken in consideration that not all techniques are transferable between laboratories or meet forensic standards but we should also consider that the Barcoding Project can guarantee the level of reproducibility and standardization necessary to presents evidences in a Court. Species identification is strictly dependent on comparison of the unknown sequence of the sampled evidence against a reference database of sequence data. Reviewing literature, it became apparent that relatively few DNA datasets exist for plants outside of barcoding research and CBOL will ease the building of a large-scale reference library with pooled data across laboratories and organizations. Sequence data archived into a database without the necessary care, standardization or quality control further complicate the identification process and are not suitable for forensic comparison [17–19]. In contrast, the build-up of a barcode library relies on the publication of sequences from well-identified voucher specimens adhering to the BARCODE standard in GenBank (http://www.ncbi.nlm.nih.gov/genbank/barcode.html) or to the Barcode of Life Datasystems (BOLD) submission guidelines and thus ensures greater reliability.

After the seminal papers on forensic botany from the Miller–Coyle group in the early 2000s [20,21] and the interesting work of Ward et al. [22,23], our group proposed for the very first time the broad application of DNA barcoding principles as a tool for identification of botanical trace evidence through the analysis of two chloroplast loci (*trnH-psbA* and *trnL-trnF*) in 63 plant species belonging to local flora [24,25]. This multimarker approach was found to be able to correctly resolve species in nearly 65% of cases with the remaining samples identified at higher taxonomic level and provided a proof of concept for such applications in forensics, as recently pointed out by Hollingsworth et al. [7]. Difficulties and drawbacks that were encountered included the poor coverage of species in the DNA sequence database and the lack of authenticated reference sequences for the selected markers. On the other hand, the successful results provided an approach to progressively identify unknown plant specimens to a given taxonomic rank, usable by any non-specialist botanist or in case of a shortage of available taxonomic expertise. Following the international agreement on a barcode core system [14], we considered mandatory to update and to compare our previous findings with the newly selected plastid markers (*matK* + *rbcL*), taking into consideration forensic requirements. Features of all the four loci (*trnH-psbA*, *trnL-trnF*, *matK*, *rbcL*) were compared singly and in pairwise combinations to assess the most suitable solution for forensic botany.

2. Materials and methods

Details on sample collection and DNA extraction of the 63 analyzed species belonging to 53 genera in 33 families are provided in an earlier work [24].

Following suggestions of CBOL-PWG, matK primers designed by Kim et al. (3F/1R) [14] were used for angiosperms. Two alternative set of primers were further employed [26] to increase taxonomic coverage in particular for non-angiosperms, including a pair specifically developed for gymnosperms [27].

Primer sequences and sources are shown in Table S1.

PCR protocols, purification and sequencing conditions are illustrated in text ESM1.

GenBank accession number of sequences are listed in Table S2.

Sequences of each locus were aligned using ClustalW [28] with default parameters.

Molecular diversity indices (number of polymorphic sites/number of total sites N_p/N, mean number of pairwise differences, nucleotide diversity) were computed with the Arlequin software [29] to describe sequence variation among different taxonomic levels.

Comparisons between species belonging to the same genera required the inclusion of additional DNA sequences retrieved from GenBank. The accessions numbers are listed in Tables S3 and S4. Diversity indices at family level were calculated only for families with at least two sampled species (see Table S2).

The species discrimination power of selected barcodes was firstly evaluated by sequence similarity search through a BLAST search in GenBank.

Then, the probability of correct species identification with rbcL and matK sequences was further assessed consulting the official barcode reference library BOLD (latest version 3.0), through the Barcode of Life Data Systems Identification Engine (BOLD-IDS), the free bioinformatics platform specifically developed to manage the large volume of DNA barcoding data.

This identification engine has begun only recently to allow the identification of plant samples for the DNA barcode regions *rbcL* and *matK*, together with COI and ITS sequences for animal and fungal identification respectively [30].

3. Results and discussion

We initially focused on of the specific features of the two agreed barcodes markers (*matk* and *rbcL)* using the previously developed plant dataset [24] with the addition of data on *Cannabis sativa* and *Chimonobambusa quadrangularis*.

In the second part of the study, the four analyzed plastid markers were considered singularly by investigating only the 63 plant species shared by both studies (this study and [24]). Finally, all possible two-locus combinations were tested to select the most appropriate barcode system for forensic plant species testing.

A total of 118 sequences were successfully recovered from the 65 starting plant species, considering both core barcoding markers (*rbcL* and *matK*).

Please cite this article in press as: G. Ferri, et al., Forensic botany II, DNA barcode for land plants: Which markers after the international agreement? Forensic Sci. Int. Genet. (2014), http://dx.doi.org/10.1016/j.fsigen.2014.10.005