

DNA Barcoding for Identification and Detection of Species

Rashmi Mathur ¹, Isha Gunwal ², Nitin Chauhan ³, Yamini Agrawal ^{4,*}

1 Sri Aurobindo College, University of Delhi; isha.gunwal83@gmail.com (I.G.);

2 Sri Aurobindo College, University of Delhi; rashmimathur16@gmail.com (R.M.);

3 Swami Shraddhanand College, University of Delhi; nitinchauhankg@gmail.com (N.C.);

4 Shaheed Rajguru College of Applied Sciences for Women, University of Delhi; yaminiagrawal04@gmail.com (Y.A.);

* Correspondence: yaminiagrawal04@gmail.com (Y.A.);

Scopus Author ID 57197080880

Received: 3.06.2021; Revised: 10.07.2021; Accepted: 14.07.2021; Published: 5.09.2021

Abstract: Identification is a very important part of the taxonomy. Since a species represents the basic unit of biological classification, identifying species is important to understand the systematics and the precise phylogenetic position of particular species. In recent years, species identification and delimitation have seen major improvements because of the incorporation of DNA sequence data. This review provides a comprehensive list of commonly employed nuclear and chloroplast regions used for the barcoding of plants.

Keywords: cytochrome C oxidase 1 (Cox1); nrITS; *trnH-psbA*; Rubisco Large subunit (*rbcL*); *matK*.

© 2021 by the authors. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

DNA barcoding is a technology for species-level identification and detection. It relies on DNA sequence variations in selected and small regions of nuclear and/or cytoplasmic genomes to provide unique molecular recognition tags to species. Thus, DNA barcodes are short sequences of DNA from standardized and globally agreed-upon locus/loci of either nuclear or cytoplasmic genome or both. These can be from coding or non-coding regions. The concept of DNA barcoding was introduced by Paul Hebert of the University of Guelph in 2003, based on his pioneering study on 200 closely allied Lepidopteran species and subsequent investigations on birds, fishes, and insects [1–4]. DNA sequence that was found to be effective in his pioneering and subsequent studies on insects, fishes, and birds, "Folmer's region" at 5' end of cytochrome C oxidase 1 (Cox1) having 658 base pairs was proposed as the universal barcode for all eukaryotes [3,5–7]. Short standardized gene regions (DNA barcodes) such as the 5.8S ribosomal RNA gene and flanking internal transcribed spacers 1 & 2 (ITS) region have been employed for the rapid and accurate identification of many species [8,9].

DNA barcoding had shown tremendous progress in global research programs since its beginning in 2003. Initially, the major drawback in applying DNA barcoding was its dependence on reference databases which was limited primarily. Now millions of barcode sequences have been produced, and a good amount of reference databases is available for researchers. The advantage of DNA barcoding over the current taxonomic identification methods is that a species can be identified even if a small amount of its tissue/DNA is available.

Within two years of promulgating the concept, it was realized that because of the low substitution rate of nucleotides present in plant mitochondria, cox 1 and other regions in the mitochondrial genome cannot be used for DNA barcoding of plants, except for some

macroalgae [10–12]. Thus, the search for a barcode for plants began earnestly in 2005 [11]. Therefore, many such regions from the chloroplast as well as from the nuclear genomes were tested as possible barcodes for plants. However, before commenting on the tested loci, it would be worthwhile to list characteristics, which a locus should possess to become a suitable barcode. The desirable attributes of a barcode locus are that it should be (i) short (~800bp), so that it can be easily sequenced in one reaction, (ii) the one that evolves fast to provide sufficient sequence variations at the species level, (iii) having sufficiently conserved flanking regions so that universal primers could be used for their amplification, (iv) variable enough to allow species distinction but with a little intraspecific variation, to provide distinct barcode gap (v) technically simple to sequence (vi) easy to align for development of effective bioinformatics tools (vi) recoverable from herbarium specimens and other degraded DNA samples [11, 13].

2. Nuclear and Chloroplast regions used for barcoding of plants

2.1. Nuclear Ribosomal Internal Transcribed Spacer (*nrITS*).

ITS part Internal Transcribed Spacer 2 (ITS 2), are the most commonly sequenced loci [11, 14], with the size of the former ranging from 400 to > 1000 bp [15].

2.2. *tRNA for histidine and photosystem II protein D1 protein*.

It is an intergenic spacer between the genes coding for tRNA for histidine and photosystem II protein D1 protein (*trnH-psbA*). This is the most variable region in angiosperms. Its length varies from 100-800bp [11, 16].

2.3. *Rubisco Large subunit (rbcL)*.

The enzyme Rubisco (Ribulose biphosphate carboxylase) is the most widespread and commonly known to participate in catalysis reactions in carbon assimilation. It comprises both small and large subunits, where a large subunit (*rbcL*) carries the site for carbon fixation exhibiting greater than 20% amino acid conserveness among plant species [17]. The length of *rbcL* is approximately 650 bp. It is easy to amplify, sequence, and align in most land plants and is a widely used phylogenetic marker [18]. Although this locus is not suitable for species-level identification, it evolves too slowly to broadly use intra-species delimitation [19].

2.4. *Maturase K (matK)*.

A portion (~950 bp) of Maturase K (mat K), a group II intron having a total length of 1536 base pairs, present in the *trnK* gene, was recommended as the single locus barcode for plants [19]. CBOL [20] recommended it as one of the loci in the two-locus barcode.

2.5. Several other loci (*atpB-rbcL*, *atpB*, *ndhF*, *psbM-trnB*, *rpl36-rps8*, *rpoB*, *rpoC1*, *rps16*, *trnC-ycf6*, *trnK-rps16*, *trnL*, *trnL-F*, *trnV-atpE*, *ycf6-psbM intron*).

Have also been tested from the chloroplast DNA. However, none of these were found to be as suitable as the one described earlier across a large group of plants.

3. Applications of DNA barcoding

DNA barcoding has several applications, such as:

- In assigning unidentified individuals to a species [5].
- In hastening the biodiversity incentivisation and analysis even if the organisms or their cells/tissue are not available using metabarcoding [21].
- As genetic resource tags [13], it could enhance the discovery of new species [22].
- In identifying cryptic and polymorphic species [3].
- In connecting distinct stages of the life cycle, which otherwise would be difficult to identify based on their lack of any common feature [23].
- In establishing botanical identities of herbals by detecting the various constituents involved in herbal formulations and foodstuffs, and their other substitutes [24–26].
- Paleo Barcoding, a method to study the effect of climate change on life forms [27].
- In forensic investigations [28].
- In screening of propagules of invasive species just during its confinement [29].
- In identification of endangered species of both plants and animals to tackle illegal trade, even if fragments of these are traded [30].
- In recognizing complex food webs by DNA analysis of animal gut [31].
- In Bio-surveillance of habitats for Biosecurity [32].
- Provides additional evidence for taxonomic circumscription [33].

4. Laboratory Protocol

The steps involved for DNA barcoding are as follows:

4.1. Collection, documentation, identification, preservation, and deposition of the samples.

This is the most critical step. Each specimen needs to be numbered, and all details about the site of collection (GPS coordinates), habit, and habitat are to be recorded precisely with a number assigned to each. The digital images of the specimens are taken as soon as possible. The specimens are identified correctly by the expert if needed. The plant is preserved, so that tissue for DNA isolation remains available in the distant future and deposited in a reputed repository. The accession number of each deposited specimen is obtained [14].

4.2. Isolation of genomic DNA.

Total genomic DNA is isolated using any methods, such as original C-TAB protocol [34], modified C-TAB methods, or Genomic DNA extraction kits (e.g., Qiagen, Fermentas, etc.). The quality and quantity of extracted DNA are estimated either spectrophotometrically by taking the absorbance at 260/280, or electrophoretically by resolving DNA fragments using 0.8% agarose gel.

4.3. Selection/designing of primers of the selected barcode loci.

Primers for the selected loci can be made from the genes' flanking regions on an allied taxon's available chloroplast genome sequence. Complete chloroplast genome sequences of many plant species are available on the NCBI database.

4.4. PCR amplification, cleaning of PCR products, and sequencing of the amplicons.

The selected loci are amplified using any DNA polymerase that also possesses proof-reading ability. The thermal cycle varies according to the primer used, the basic steps of which involve denaturation, annealing, and extension of the target region [14]. The PCR products are cleaned using the enzymatic method or the gel extraction method. The enzymatic purification method involves the use of Exonuclease-I and shrimp Alkaline Phosphatase (Exo/SAP). Exonuclease I degrades single-stranded DNA molecules (remaining primers), while Shrimp Alkaline Phosphatase removes phosphate groups from the remaining dNTPs. After purification, amplicons are sequenced bi-directionally using Sanger's sequencing method.

4.5. Analysis and calculating divergence values.

For assembling and having accurate base calling (PHRED score > 20, base calling is more than 99% correct), codon-code aligner or any other software is used.

6. Limitations of the DNA barcoding

DNA barcoding, like other technologies, is not 100% perfect, though less than 100% species resolution in some instances could be due to the similar less than perfect nature of taxonomic delimitation methods. Thus, the conflicts arising out of the DNA barcode data may require reexamination by taxonomists. Furthermore, the utility of an individual locus as the barcode depends not only on its species delimitation ability but also on its success in easy amplification and sequencing reactions [11–20]. Many researchers have signified ITS as a potential universal DNA marker due to its ability to evolve rapidly and discriminate closely related species with ease [15, 35]. However, it imposes a few inherent limitations, such as incidences of intra-genomic variability due to the presence of divergent paralogous copies within the individuals [36]; pseudogenes [37], which could hinder getting good quality sequences. Based on these limitations, CBOL Plant Working Group [20] has not included it in plants' core barcode (matK+ rbcL).

Major problems for authenticating the herbal material using DNA barcoding are the absence of authentic reference sequences in the GenBank associated with vouchered specimens submitted in herbaria. The isolation of good-quality DNA is important for the successful application of molecular methods. However, sometimes it becomes quite challenging to extract high molecular weight DNA being in a highly degraded state or due to the presence of higher amounts of polysaccharides, polyphenols, secondary metabolites in the processed medicinal plant material [38].

Primarily, most of the DNA barcode-based studies were based on Sanger's sequencing. These days, Next-Generation Sequencing (NGS) is also a preferred technique particularly for analyzing samples: showing varying levels of DNA degradation, and/or acquired from multiple species, and/or containing fillers or contaminants. In addition, it offers numerous advantages over Sanger sequencing, including multiple parallel sequencing reactions at a time, clonal templates separation, superior sensitivity, and faster turnaround time [39].

7. Conclusions

Nowadays, DNA barcoding is the central molecular technique for species-level identification. DNA sequence variations in selected and small regions of nuclear and/or

cytoplasmic genomes are employed to provide unique molecular recognition tags to species. A list of DNA barcodes (about 1.3 M public records) is available in the BOLD system specific to animal, fungal, and plant species with a data retrieval interface. Such a system would ease the application of DNA barcoding to identify species apart from routine taxonomic methods.

Funding

This research received no external funding.

Acknowledgments

The authors gratefully acknowledge the University of Delhi for support.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Hleap, J.S.; Littlefair, J.E.; Steinke, D.; Hebert, P.D.; Cristescu, M.E. Assessment of current taxonomic assignment strategies for metabarcoding eukaryotes. *Mol Ecol Resour.* **2021**, <https://doi.org/10.1111/1755-0998.13407>.
2. Raven, P.H.; Miller, S.E. Here today, gone tomorrow **2020**, 370, 149, <https://doi.org/10.1126/science.abf1185>.
3. Hebert, P.D.; Penton, E.H.; Burns, J.M.; Janzen, D.H.; Hallwachs, W. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc Natl Acad Sci* **2004**, *101*, 14812-14817, <https://doi.org/10.1073/pnas.0406166101>.
4. Hebert, P.D.; Stoeckle, M.Y.; Zemplak, T.S.; Francis C.M. Identification of birds through DNA barcodes, *PLoS Biol* **2004**, *2*, 1657-1663.
5. Hebert, P.D.; Cywinska, A.; Ball, S.L.; Dewaard, J.R. Biological identifications through DNA barcodes. *Proc Biol Sci.* **2003**, *270*, 313-321, <https://doi.org/10.1098/rspb.2002.2218>.
6. Hebert, P.D.; Ratnasingham, S.; De Waard, J.R. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc Biol Sci.* **2003**, *270*, <https://doi.org/10.1098/rsbl.2003.0025>.
7. Hebert, P.D.; Gregory, T.R. The Promise of DNA Barcoding for Taxonomy. *Syst Biol* **2005**, *54*, 852–859, <https://doi.org/10.1080/10635150500354886>.
8. Schoch, C.L.; Seifert, K A.; Huhndorf, S.; Robert, V.; Spouge, J.L.; Levesque, C.A. Fungal Barcoding Consortium. Fungal Barcoding Consortium. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc Natl Acad Sci* **2012**, *109*, 6241-6246, <https://doi.org/10.1073/pnas.1117018109>.
9. Shenoy, B.D.; Jeewon, R.; Hyde, K.D. Impact of DNA sequence-data on the taxonomy of anamorphic fungi. *Fungal Divers* **2007**, *26*, 1-54.
10. Saunders, G.V.; Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *The Royal Society* **2005**, <https://doi.org/10.1098/rstb.2005.1719>.
11. Kress, W.J.; Wurdack, K.J.; Zimmer, E.A.; Weigt L.A.; Janzen D.H. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci USA* **2005**, *102*, 8369-8374.
12. Rubinoff, D.; Cameron, S.; Will, K. A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. *J hered* **2006**, *97*, 581-594.
13. Parveen, I.; Singh, H.K.; Raghuvanshi, S.; Pradhan, U.C.; Babbar, S.B. DNA barcoding of endangered Indian *Paphiopedilum* species. *Mol Ecol Resour* **2012**, *12*, 82-90.
14. Malik, S.; Priya, A.; Babbar, S.B. Employing barcoding markers to authenticate selected endangered medicinal plants traded in Indian markets. *Physiol Mol Biol Plants.* **2019**, *25*, 327-337, <https://doi.org/10.1007/s12298-018-0610-8>.
15. Li, F.W.; Kuo, L.Y.; Rothfels, C.J.; Ebihara, A.; Chiou, W.L.; Windham, M.D.; Pryer, K.M. *rbcL* and *matK* earn two thumbs up as the core DNA barcode for ferns. *PLoS One* **2011**, *6*, 26597.

16. Pang, X.; Liu, C.; Shi, L.; Liu, R.; Liang, D.; Li, H.; Chen, S. Utility of the *trnH-psbA* Intergenic Spacer Region and Its Combinations as Plant DNA Barcodes: A Meta-Analysis. *PLoS One* **2012**, *7*, 48833, <https://doi.org/10.1371/journal.pone.0048833>.
17. Miziorko, H.M.; Lorimer, G.H. Ribulose-1,5-bisphosphate carboxylase-oxygenase. *Annu Rev Biochem.* **1983**, *52*, 507-35, <https://doi.org/10.1146/annurev.bi.52.070183.002451>.
18. Chase, M.W.; Salamin, N.; Wilkinson, M.; Dunwell, J.M.; Kesanakurthi, R.P.; Haidar, N.; Savolainen, V. Land plants and DNA barcodes: short-term and long-term goals. *Philos Trans R Soc Lond B Biol Sci.* **2005**, *360*, 1889-1895.
19. Lahaye, R.; Van der Bank, M.; Bogarin, D.; Warner, J.; Pupulin, F.; Gigot, G.; Savolainen, V. DNA barcoding the floras of biodiversity hotspots. *Proc Natl Acad Sci USA* **2008**, *105*, 2923-2928.
20. Hollingsworth, P.M.; Graham, S.W.; Little, D.P. Choosing and using a plant DNA barcode. *PLoS One* **2011**, *6*, 19254.
21. Stern, R.F.; Andersen, R.A.; Jameson, I.; Küpper, F.C.; Coffroth, M.A.; Vaulot, D.; Keeling, P.J. Evaluating the Ribosomal Internal Transcribed Spacer (ITS) as a Candidate Dinoflagellate Barcode Marker. *PLoS One* **2012**, *7*, e42780, <https://doi.org/10.1371/journal.pone.0042780>.
22. Bell, K.L.; De Vere, N.; Keller, A.; Richardson, R.T.; Gous, A.; Burgess, K.S.; Brosi, B.J. Pollen DNA barcoding: current applications and future prospects, *Genome* **2016**, *59*, 629-640, <https://doi.org/10.1139/gen-2015-0200>.
23. Valdez-Moreno, M.; Vásquez-Yeomans, L.; Elías-Gutiérrez, M.; Ivanova, N.V.; Hebert, P.D. Using DNA barcodes to connect adults and early life stages of marine fishes from the Yucatan Peninsula, Mexico: potential in fisheries management. *Mar Freshw Res* **2010**, *61*, 665–671.
24. Chen, S.; Yao, H.; Han, J.; Liu, C.; Song, J.; Shi, L.; Leon, C. Validation of ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS One* **2010**, *5*, 8613.
25. Valentini, A.; Miquel, C.; Taberlet P. DNA barcoding for honey biodiversity. *Divers* **2010**, *2*, 610-617.
26. Srirama, R.; Senthilkumar, U.; Sreejayan, N.; Ravikanth, G.; Gurumurthy, B.R.; Shivanna, M.B.; Shaanker, R. U. Assessing species admixtures in raw drug trade of *Phyllanthus*, a hepato-protective plant using molecular tools. *J Ethnopharmacol* **2010**, *130*, 208-215.
27. Jørgensen, T.; Kjaer, K.H.; Haile, J.; Rasmussen, M.; Boessenkool, S.; Andersen, K.; Willerslev, E. Islands in the ice: detecting past vegetation on Greenlandic nunataks using historical records and sedimentary ancient DNA Meta-barcoding: islands in the ice. *Mol. Ecol.* **2012**, *21*, 1980-1988, <https://doi.org/10.1111/j.1365-294X.2011.05278.x>.
28. Amendt, J.; Richards, C.S.; Campobasso, C.P.; Zehner, R.; Hall, M.J. Forensic entomology: applications and limitations. *Forensic Sci Med Pathol* **2011**, *7*, 379-92, <https://doi.org/10.1007/s12024-010-9209-2>.
29. Madden, M.J.; Young, R.G.; Brown, J.W.; Miller, S.E.; Frewin, A.J.; Hanner, R.H. Using DNA barcoding to improve invasive pest identification at U.S. ports-of-entry. *PLoS One* **2019**, *14*, e0222291, <https://doi.org/10.1371/journal.pone.0222291>.
30. Eaton, J.A.; Shepherd, C.R.; Rheindt, F.E.; Harris, J.B.C.; Van Balen, S.; Wilcove, D.S.; Collar, N. J. Collar: Trade-driven extinctions and near-extinctions of avian taxa in Sundaic Indonesia, *Forktail* **2015**, *31*, 1-12.
31. Roslin, T.; Majaneva, S. The use of DNA barcodes in food web construction- terrestrial and aquatic ecologists unite. *Genome* **2016**, *59*, 603-628, <https://doi.org/10.1139/gen-2015-0229>.
32. Armstrong, K.F.; Ball, S.L. DNA barcodes for biosecurity: invasive species, *Philos Trans R Soc Lond B Biol Sci.* **2005**, *360*, 1813-1823, <https://doi.org/10.1098/rstb.2005.1713>.
33. Hubert, N.; Hanner, R. DNA Barcoding, species delineation and taxonomy: a historical perspective. *DNA Barcodes* **2015**, *3*, 44-58, <https://doi.org/10.1515/dna-2015-0006>.
34. Doyle, J. J.; Doyle, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* **1987**, *19*, 11-15.
35. Tripathi, A. M.; Tyagi, A.; Kumar, A.; Singh, A.; Singh, S.; Chaudhary, L. B.; Roy, S. The Internal Transcribed Spacer (ITS) Region and *trnH-psbA* Are Suitable Candidate Loci for DNA Barcoding of Tropical Tree Species of India. *PloS One* **2014**, *9*, e105914, <https://doi.org/10.1371/journal.pone.0105914>.
36. Álvarez, I. J. F. W.; Wendel, J. F. Ribosomal ITS sequences and plant phylogenetic inference. *Mol Phylogenet Evol.* **2003**, *29*, 417-434.
37. Bailey, C.D.; Carr, T.G.; Harris, S.A.; Hughes, C.E. Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. *Mol Phylogenet Evol.* **2003**, *29*, 435-55, <https://doi.org/10.1016/j.ympev.2003.08.021>.

38. Tehen, N.; Parveen, I.; Pan, Z.; Khan, I.A. DNA barcoding of medicinal plant material for identification. *Curr. Opin. Biotechnol.* **2014**, *25*, 103-110, <https://doi.org/10.1016/j.copbio.2013.09.010>.
39. Ivanova, N.V.; Kuzmina, M.L.; Braukmann, T.W.; Borisenko, A.V.; Zakharov, E.V. Authentication of Herbal Supplements Using Next-Generation Sequencing. *Plos One* **2016**, *11*, e0156426, <https://doi.org/10.1371/journal.pone.0156426>.