# **Chloroplast DNA Barcodes**

#### Subjects: Genetics & Heredity

Contributor: Stavroula Letsiou, Panagiotis Madesis, Efstathios Vasdekis, Cinzia Montemurro, Maria E. Grigoriou, George Skavdis, Vassilios Moussis, Antonios E. Koutelidakis, Andreas G. Tzakos

DNA barcodes are standardized sequences, ideally unique, coding or non-coding, either from the genome of the organism or from its organelles, that are used to identify/classify an organismal group; in short, the method includes amplification of the DNA barcode, sequencing and comparison with a reference database containing the relevant sequences from different species. In plants, the use a universal DNA barcode, such as *COI*, which is used in animals, has not been achieved so far.

Keywords: DNA barcoding ; chloroplast ; matK ; rbcL ; trnH-psbA ; rpoB ; rpoC1 ; psbK ; psbI

#### 1. DNA Barcodes

Standardization, minimalism, and scalability are the three oilers of DNA barcoding. This technique has been successfully used for species identification in animals; a 648-base pair (bp) fragment near the 5'-end of the mitochondrial gene cytochrome c oxidase subunit I (*COI*) has been selected as the standard barcode <sup>[1]</sup>: (a) there is a large copy number per cell resulting in easier amplification from smaller or degraded samples, (b) it is maternally inherited, (c) there is no possibility of recombination with paternal copies, and (d) it rapidly accumulates mutations <sup>[2][3]</sup>. While *COI* is a suitable target for animals, it does not discriminate most plants because of a much slower mutation rate. This has led to the search of alternative barcoding regions <sup>[4][5]</sup>.

The fundamental concept underpinning DNA barcoding is rooted in the notion that throughout the evolution of species, certain DNA lengths within both coding and non-coding regions remain highly conserved, undergoing minor changes. Sequences found in cytoplasmic mitochondrial DNA, chloroplast DNA, and selected segments from nuclear DNA embody these characteristics, making them suitable candidates for DNA barcoding. The utilization of these sequences enables the differentiation of species, providing a molecular signature that facilitates accurate and efficient species identification in the field of DNA barcoding <sup>[G]</sup>. The suitability of such loci or the combination thereof is under discussion for plant species where there is not one easily applied solution. The design of universal primers could enable efficient PCR amplification that following sequencing and bioinformatic analysis and would ideally identify all the known species. Unfortunately, so far, the ideal DNA barcode does not exist in plants <sup>[Z][8][9]</sup>. Several barcodes, single or multiple, have been used and are presented below.

Chloroplast DNA is a circular molecule with a size between 120 and 220 kb and consists of a large and a small singlecopy region (LSC and SSC) intervened by two copies of a large, inverted repeat (Ira and Irb). There are about 100 functional genes that can be used for species identification and, according to some researchers, besides single-locus markers, the whole plastid genome could be used for DNA barcoding besides single-locus markers. DNA barcodes from chloroplast genes are extensively used in plant phylogenetic studies; the design of primers is easy, gene order in the genome of the organelle is conserved, and amplification is much easier due to the high copy number per cell. Nevertheless, compared with the nuclear genome genes of the chloroplast genome, they are characterized by a low evolutionary rate [6][10][11]. Among the chloroplast markers, the following have been successfully used:

#### 2. matK

*matK* (maturase K) is one of the most rapidly evolving chloroplast genes, which has been used for identification at the family, the genus, and even the species level. *matK* exhibits interspecific divergence and a low transition/transversion rate. It is approximately 1550 bp long and encodes maturase K, an enzyme involved in the splicing of type-II introns <sup>[12][13]</sup> [<sup>14][15]</sup>. However, its use as a universal DNA barcode is hampered by technical problems, mainly the design of the universal primer sets, due to the high substitution rate <sup>[16][17][18]</sup>. However, *matK* constitutes a suitable marker for angiosperm, flowering plant, bryophyte, lycophyte, gymnosperm, and monilophyte identification <sup>[15][18]</sup>.

#### 3. rbcL

*rbcL* (ribulose bisphosphate carboxylase/oxygenase large subunit) is a candidate locus for comparing at the levels of family and genus; however, it is not suitable for species identification as it has modest discriminatory power. This marker has been one of the most studied among the plastid genome, with wide representation from all major groups and many available sequences in GenBank <sup>[5][7][8][13]</sup>. It was the first gene sequenced from the plant chloroplast genome and encodes the large subunit of rubilose-1,5-bisphosphate carboxylase/oxygenase (RUBISCO), a critical photosynthetic enzyme <sup>[19]</sup>. *rbcL* is easy amplify and sequence but has a slow evolutionary rate <sup>[13][20]</sup>. Its length is approximately 1430 bp and thus at least two sets of primers are needed to sequence the entire coding sequence <sup>[8][12]</sup>. *rbcL* meets most of the desired criteria and can be used in conjunction with other markers <sup>[5][16]</sup>. It is also widely used for alga, peptidophyte, and angiosperm identification <sup>[9][21]</sup>.

### 4. trnH-psbA

*trnH-psbA* is one of the most variable non-coding plastid loci with an intergenic spacer suitable to offer a high level of species discrimination  $\frac{12}{16}$ . It is easily amplified with universal primers but as it has high rates of insertion/deletion, alignment can be difficult. Moreover, its length varies among different families, with this region containing copies of *rps19* in some cases, as well as a pseudogene that is located between *trnH* and *psbA*; this causes a problem, as despite obtaining high-quality bidirectional sequences, alignment is difficult due to the high length variation. Most researchers have proposed that *trnH-psbA* should be used in combination with one or more loci to provide adequate resolution  $\frac{517}{16}$ . Nevertheless, is has been shown that it is a suitable marker for flowering plants and peptidophytes  $\frac{8121}{2}$ .

## 5. rpoB and rpoC1

*rpoB* (RNA polymerase subunit B) and *rpoC1* (RNA polymerase subunit C1) are plastid genes, encoding subunits of the plastid-encoded plastid RNA polymerase that have been used for the identification at the family level but, due to their slow evolution rate, they cannot be used for species discrimination in many plant families <sup>[15][16]</sup>. Both can be efficiently amplified with a limited range of PCR conditions and primer sets <sup>[16]</sup>. *rpoB*, *rpoC1*, *rpoC2* encode three out of four subunits of the chloroplast RNA polymerase <sup>[22]</sup> and are suitable markers for bryophyte identification <sup>[23]</sup>.

## 6. trnL-trnF (Genic, Intron, and Intergenic Spacer)

The *trnL-trnF* intergenic spacer has been proposed as a universal plastid amplicon and has been widely used in plant systematics and plylogeography since the 1990s  $^{[12][18]}$ . This region is located in the large single copy region of the chloroplast genome  $^{[24]}$ . Despite its slow rate of molecular evolution, the plastid *trnL* intron is suggested as a possible marker because of its conserved sites; hence, it could be a useful tool for evolutionary studies at higher taxonomic levels  $^{[8][16]}$ . Taberlet et al.  $^{[25]}$  established primers that work for 19 species tested including algae, bryophytes, pteridophytes, gymnosperms, and angiosperms.

## 7. psbK-psbl (Intergenic Spacer)

The *psbK* and *psbI* loci encode two low molecular weight polypeptides, K and I, of the photosystem II <sup>[26]</sup>. The non-coding *psbK-psbI* intergenic spacer is conserved and can be easily amplified with PCR, sequenced, and aligned <sup>[27][28]</sup>. It also demonstrates high discriminatory power but low sequence quality and universality <sup>[12]</sup>. Despite its discriminatory power, the CBOL Plant Working Group propose its use as a supplementary locus due to the inconsistency in obtaining bidirectional unambiguous sequences <sup>[5]</sup>. Nevertheless, it constitutes a suitable marker for bryophyte, lycophyte, and monilophyte identification <sup>[18]</sup>.

#### 8. atpF-atpH (Intergenic Spacer)

The non-coding, plastid region *atpF-atpH* could be used as a universal DNA barcoding marker for species-level identification but its discriminatory power is medium. The genes *atpF* and *atpH* encode ATP synthase subunits CFO I and III. The length of *atpF-atpH* sequences vary from 598 to 613 bp and the alignment of these sequences is difficult despite easy PCR amplification. For this reason, it could be useful only as supplementary marker in plant DNA barcoding, providing better resolution on specific projects and taxonomic groups  $\frac{[18][29][30][31]}{29}$ . According to Wang W. et al.  $\frac{[30]}{29}$ , it is a suitable marker for duck seed identification.

#### References

- Coissac, E.; Hollingsworth, P.M.; Lavergne, S.; Taberlet, P. From barcodes to genomes: Extending the concept of DNA barcoding. Mol. Ecol. 2016, 25, 1423–1428.
- 2. Luo, A.; Zhang, A.; Ho, S.Y.; Xu, W.; Zhang, Y.; Shi, W.; Cameron, S.L.; Zhu, C. Potential efficacy of mitochondrial genes for animal DNA barcoding: A case study using eutherian mammals. BMC Genom. 2011, 12, 84.
- 3. Waugh, J. DNA barcoding in animal species: Progress, potential and pitfalls. Bioessays 2007, 29, 188–197.
- 4. 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.
- 5. CBOL Plant Working Group. A DNA barcode for land plants. Proc. Natl. Acad. Sci. USA 2009, 106, 12794–12797.
- Ali, M.A.; Gyulai, G.; Hidvegi, N.; Kerti, B.; Al Hemaid, F.M.; Pandey, A.K.; Lee, J. The changing epitome of species identification–DNA barcoding. Saudi J. Biol. Sci. 2014, 21, 204–231.
- 7. Newmaster, S.G.; Fazekas, A.J.; Ragupathy, S. DNA barcoding in land plants: Evaluation of rbcL in a multigene tiered approach. Botany 2006, 84, 335–341.
- Kress, W.J.; Erickson, D.L. 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, e508.
- 9. Kowalska, Z.; Pniewski, F.; Latała, A. DNA barcoding–A new device in phycolog'st's toolbox. Ecohydrol. Hydrobiol. 2019, 19, 417–427.
- 10. Li, X.; Yang, Y.; Henry, R.J.; Rossetto, M.; Wang, Y.; Chen, S. Plant DNA barcoding: From gene to genome. Biol. Rev. 2015, 90, 157–166.
- 11. Yu, J.; Wu, X.; Liu, C.; Newmaster, S.; Ragupathy, S.; Kress, W.J. Progress in the use of DNA barcodes in the identification and classification of medicinal plants. Ecotoxicol. Environ. Saf. 2021, 208, 111691.
- 12. Hollingsworth, P.M.; Graham, S.W.; Little, D.P. Choosing and using a plant DNA barcode. PLoS ONE 2011, 6, e19254.
- Liu, M.; Li, X.-W.; Liao, B.-S.; Luo, L.; Ren, Y.-Y. Species identification of poisonous medicinal plant using DNA barcoding. Chin. J. Nat. Med. 2019, 17, 585–590.
- 14. Hilu, K.W.; Liang, G. The matK gene: Sequence variation and application in plant systematics. Am. J. Bot. 1997, 84, 830–839.
- 15. Dong, W.; Liu, J.; Yu, J.; Wang, L.; Zhou, S. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PLoS ONE 2012, 7, e35071.
- Chase, M.W.; Cowan, R.S.; Hollingsworth, P.M.; Van Den Berg, C.; Madriñán, S.; Petersen, G.; Seberg, O.; Jørgsensen, T.; Cameron, K.M.; Carine, M.; et al. A proposal for a standardised protocol to barcode all land plants. Taxon 2007, 56, 295–299.
- 17. Yu, J.; Xue, J.H.; Zhou, S.L. New universal matK primers for DNA barcoding angiosperms. J. Syst. Evol. 2011, 49, 176–181.
- Fazekas, A.J.; Burgess, K.S.; Kesanakurti, P.R.; Graham, S.W.; Newmaster, S.G.; Husband, B.C.; Percy, D.M.; Hajibabaei, M.; Barrett, S.C.H. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS ONE 2008, 3, e2802.
- 19. Zurawski, G.; Perrot, B.; Bottomley, W.; Whitfeld, P.R. The structure of the gene for the large subunit of ribulose 1, 5bisphosphate carboxylase from spinach chloroplast DNA. Nucleic Acids Res. 1981, 9, 3251–3270.
- 20. Tnah, L.H.; Lee, S.L.; Tan, A.L.; Lee, C.T.; Ng, K.K.S.; Ng, C.H.; Farhanah, Z.N. DNA barcode database of common herbal plants in the tropics: A resource for herbal product authentication. Food Control. 2019, 95, 318–326.
- 21. Saddhe, A.A.; Kumar, K. DNA barcoding of plants: Selection of core markers for taxonomic groups. Plant Sci. Today 2018, 5, 9–13.
- 22. Serino, G.; Maliga, P. RNA polymerase subunits encoded by the plastid rpo genes are not shared with the nucleusencoded plastid enzyme. Plant Physiol. 1998, 117, 1165–1170.
- 23. Liu, Y.; Yan, H.F.; Cao, T.; Ge, X.J. Evaluation of 10 plant barcodes in Bryophyta (Mosses). J. Syst. Evol. 2010, 48, 36–46.
- 24. Park, E.; Kim, J.; Lee, H. Plant DNA barcoding system for forensic application. Forensic Sci. Int. Genet. Suppl. Ser. 2017, 6, e282–e283.
- 25. 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, 1105–1109.

- 26. Meng, B.Y.; Wakasugi, T.; Sugiura, M. Two promoters within the psbK-psbl-trnG gene cluster in tobacco chloroplast DNA. Curr. Genet. 1991, 20, 259–264.
- Lee, H.L.; Yi, D.K.; Kim, J.S.; Kim, K.J. Development of plant DNA barcoding markers from the variable noncoding regions of chloroplast genome. In Abstract Presented at the Second International Barcode of Life Conference; Academia Sinica: Taipei, Taiwan, 2007.
- 28. Lahaye, R.; Savolainen, V.; Duthoit, S.; Maurin, O.; van der Bank, M. A test of psbK-psbI and atpF-atpH as potential plant DNA barcodes using the flora of the Kruger National Park (South Africa) as a model system. Nat. Preced. 2008, 1.
- 29. Vijayan, K.; Tsou, C.H. DNA barcoding in plants: Taxonomy in a new perspective. Curr. Sci. 2010, 99, 1530–1541.
- 30. Wang, W.; Wu, Y.; Yan, Y.; Ermakova, M.; Kerstetter, R.; Messing, J. DNA barcoding of the Lemnaceae, a family of aquatic monocots. BMC Plant Biol. 2010, 10, 205.
- 31. Neto, A.B.; Morais, M.B.; Dutra, E.D.; Junior, T.C. Biological diversity of Lemna aequinoctialis Welw. isolates influence biomass production and wastewater phytoremediation. Bioresour. Technol. Rep. 2019, 6, 251–259.

Retrieved from https://encyclopedia.pub/entry/history/show/124368