Present Status of DNA Barcoding in Plants

The focus of barcoding studies in plants, thus far, was mostly on assessing the relative efficiency of molecular markers that had been used in various phylogenetic studies. From this relative assessment it became clear that none of the DNA segments tested so far had all the qualities essential for a standard barcode for plants. Although some of the loci tested had many promising characters, they had several limitations as well.

- For instance, *rbcL* and *trnL* (UAA)–*trnF*(GAA) have higher **universality**, but they **lack adequate species discriminatory power**.
- *matK* and *trnH–psbA* have higher species resolution, but problems remain **with PCR amplification and sequencing**.
- *rpoC1*, *rpoB*, *atpF–atpH* and *psbK–psbI* have problems on either species discrimination or PCR amplification across all major plant groups.
- This promotes the proposal of using locus combinations, which can complement each other, for designation as a standard barcode. CBOL–Plant Working Group5preferred a two-locus barcode combination consisting of ***rbcL* and *matK*** genes as the standard barcode for land plants.
- *rbcL* also performs well in discrimination tests in combination with other loci. Likewise, the *matK* gene sequence, as stated above, has the highest evolution rate among plastid genes, and thus has high species discriminatory power.

Recently developed primers have improved its PCR amplification and sequencing in a wide range of angiosperms. Thus, the CBOL–Plant Working Group considers that '*rbcL+matK* combination represents a pragmatic solution to a complex trade-off between universality, sequence quality, discrimination and cost'.

Primers	5′ sequence 3′	Usage	Designed by
MG15	'ATCTGGGTTGCTAACTCAATG'	PCR Amplification	Hilu and Liang (1997)
MG1	'CTACTGCAGAACTAGTCGGATGGAGTAGAT'	PCR Amplification	Hilu and Liang (1997)
MS4R	'TATCTGACATAATGCATGAA'	Sequencing	Shi et al. (this study)
MS5R	'TTGAATGAATAGATCGTRA'	Sequencing	Shi et al. (this study)
Mat*K*-1470R	'AAGATGTTGAT(T/C)GTAAATGA'	Sequencing	Johnson and Soltis (1994)
MS2F	'CTATATAATTCTCATGTAT'	Sequencing	Shi et al. (2000)
Mat*K* 5	'CGATCCTTTCATGCATT'	Sequencing	Hilu and Liang (1997)
Mat*K* F6	'TCAGTGGTACGGAATCAAAT'	Sequencing	Sang et al. (1997a)

Spooner, after investigating the efficacy of *psbA–trnH*, *matK* and *nrITS* as barcodes on 104 accessions from 63 species of wild potatoes, reported that sequences of *psbA–trnH*, *matK* and *nrITS* failed to provide species-specific markers. The plastid genes failed to provide adequate differentiation, whereas the *nrITS* sequences exhibited high intraspecific variations.

Similar difficulties were also observed earlier in many genera of the subfamily Magnolioideae, family Magnoliaceae, and in another family Lauraceae with *matK* gene.

A combination of plastid genomic loci such as *rbcL* + *rpoC1* + *matK* + *trnH–psbA* is useful as a barcoding system only for identification of broad groups of species. Thus, for projects like identification or circumscription of species which require high resolution, the presently proposed two-locus standard barcode is not sufficient. Furthermore, it is also a known fact that extensive intraspecific and intrapopulation variations were reported from many angiosperms.

For instance, using *trnH–psbA* as barcodes, floating pennywort (*Hydrocotyle ranunculoides* L.f.) was distinguished from its most closely related congeners. In another study using *matK* and *trnH–psbA* as DNA barcodes, Raghupathy *et al*., discriminated a new cryptic species of grass *Tripogon cope*, as deciphered by the hill tribes, from its close relatives in the Western Ghats and part of the Nilgiri Biosphere Reserve in India.

AVAILABILITY OF BIOINFORMATICS RESOURCES

- Parallel to the efforts in finding a standard barcoding system for all organisms in this planet, efforts have also been made to develop adequate bioinformatics resources to support the barcoding of life.
- The Barcode of Life Data System (BOLD; http://www.barcodinglife.org/views/login.php) was the result of such efforts made by CBOL to facilitate easy deposition and retrieval of data on barcodes.
- BOLD provides an integrated bioinformatics platform for all phases of the analytical pathway from specimen collection to tightly validated barcode library.

- NCBI has also provided a web-based barcode submission tool (BarSTool) for submitting sequences of barcodes, but currently it accepts sequences of CO1 only and other sequences should be submitted through Bankit or Sequin.

CONCLUSION

- Based on the available information, the CBOL–Plant Working Group proposed a two-locus (*rbcL* + *matK)* standard barcode for all land plants.
- Although empirical data have shown the inadequacies of this as a universal barcode for all land plants, the proposal of such a barcode for all plants is an important step towards establishing a centralized plant barcode database, just like the DNA database in GenBank, for its effective and easy use:
    - in **taxonomy**, **biodiversity assessment** and **conservation**, **prevention** of **illegal plant exports and imports**, **identification** of **endemic** and **endangered plant species** and other activities where identification of plant species is essential.