

Plant barcode region proposals — Review report

In September 2009 two proposals for plant barcode region selection were submitted to CBOL's executive secretary, and then passed on to an ad-hoc review panel consisting of three botanists with a strong interest in, yet no direct involvement with the proposals. The proposals were structured according to the requirements outlined in CBOL's non-COI region selection protocol.

In brief, the proposal by Hollingsworth & al. involves the entire CBOL Plant Working Group and reflects the collaborative outcome from at least three international research projects as well as individual efforts and contributions, for instance from the CCDB in Guelph. Hollingsworth & al. propose to adopt the plastid-encoded regions *rbcl* and *matK* to be the plant barcode, based on findings which are now published in PNAS 2009.

Kress & al. state in their proposal that a third region, *trnH-psbA*, is added to the core-two in the Hollingsworth & al. proposal. They emphasize that, given the sub-optimal reputation of *matK* in terms of PCR success, adding *trnH-psbA* is a good insurance in case *matK* can't be generated for every specimen, even if adding this 3rd region does not seem to add significantly to discriminating power.

The committee considered the proposals to comply with CBOL's non-COI selection criteria, held two telephone conferences and exchanged emails. It was felt that a recommendation containing both a consensus part and, in addition, personal views of committee members would be produced and submitted to CBOL's board for consideration.

The committee reached consensus on the following points:

- A multi-locus barcode region is necessary for plants in order to reach the objective of a species identification system. The cost factor is expected to be marginalized in the light of rapidly-advancing DNA PCR and sequencing technology.
- Based on the proposed regions at present the identification success rate is seen as ~70% in plants; in order to get this improved to insect/bird standards (i.e. >95%) a nuclear region needs to be added eventually. Assigning barcode status to further regions at a later point must always be an option if increased resolution/performance has been demonstrated.
- For a multi-locus barcode, every locus should be sequenced, and all are subject to requirements outlined in CBOL's data quality standards. However, there are technical and other reasons why this might not be possible.
- Better have a larger set of barcode regions later on, than now deciding for a small core, and have 'local communities' decide on additional clade-specific regions that shall not be named 'barcode regions'
- In hind-sight, criteria for candidate region selection should have been more inclusive and included chloroplast and nuclear (ITS) non-coding regions; 'barcode alignability' is not a critical issue, although in itself the ability to check reading frames in coding sequences does indeed present an advantage
- *matK* primer universality may be a more serious concern than indicated by the PWG

- trnH-psbA sequence quality may be less of a concern than indicated by the PWG
- rDNA ITS: CBOL PWG should expedite/inspire the compilation of rDNA ITS test alignments for, possibly, the same set of sister species used by the PWG, in addition to all other properly documented & vouchered ITS sequences available elsewhere, in order to test their performance as barcode region, which is claimed by some could reach up to 95% success in species discrimination. The committee is aware that because of the paralogy problem documented for rDNA ITS regions in various plant groups, using this region as plant barcode might not be suitable universally.
- For now the best way ahead is to select the three regions collectively proposed by Hollingsworth & al., and Kress & al. and include an 18 month trial period in order to assess the success of both matK amplification and psbA-trnH sequencing.