CBOL approves *matK* and *rbcL* as the BARCODE regions for Land Plants

Statement by the Executive Committee, Consortium for the Barcode of Life

16 November 2009

The Consortium for the Barcode of Life (CBOL) received two well-documented proposals for the barcode regions for land plants. The first proposed rbcL and matK and referred to an August 2009 publication in the Proceedings of the National Academy of Science. This article pointed out that *matK* was not easy to amplify in some groups and that additional work on primer development is needed. The article argued that non-coding regions, especially *trnH-psbA*, have strong potential as DNA barcodes, but suffered from technical problems that make automated sequence assembly difficult. The PNAS paper argued that progress with a two-locus barcode region would be cheaper and faster than with a three-locus region, especially if the third region is non-coding and requires manual sequence editing. The Plant Working Group concluded that success in solving primer problems with matK is more likely than solving sequence quality problems for *trnH-psbA* and they recommended rbcL and matK as the two-locus barcode.

The second proposal to CBOL proposed a three-locus barcode consisting of *rbcL*, *matK*, and *trnH-psbA*. The proposal argued that breakthroughs on the *matK* and *trnH-psbA* problems are difficult to predict, and that a three-locus barcode would provide greater probability of having two barcode sequences for all species.

CBOL appointed an ad hoc panel of three independent reviewers for an evaluation of both proposals. The panel questioned the assertion in the PNAS paper that a solution to the matK primer problem is more likely than automated sequence processing for a non-coding region such as *trnH-psbA*. The panel recommended approval of the three-locus barcode with a reassessment after 18 months. If significant progress on either the *matK* or *trnH-psbA* problem is made by then, they said that the still-problematic barcode region could be eliminated.

CBOL's Executive Committee considered both proposals and the recommendations of the review panel. The Committee was convinced that the proposals provided solid factual basis on which to make a decision. Like the proposers and review panel, the Executive Committee noted that the 70-75% success rate of the proposed plant barcodes is significantly lower than the success of COI among animals. Nevertheless, the Committee agreed that further delay is unwarranted, unwise, and unlikely to produce a better solution.

The Executive Committee viewed the review panel's recommendation for a three-locus barcode region as scientifically defensible, but a more conservative and costly solution. The Committee was not convinced of the advantages of a three-locus barcode over a two-locus standard. The Committee therefore gave more weight to the findings of CBOL's Plant Working Group in its August 2009 PNAS paper. In the Committee's view, requiring a third region for the very large sample sizes involved in plant barcoding would add significant cost and delay to the plant barcoding initiative without adding resolving power in cases where effective *matK* primers have been developed.

The Executive Committee therefore concludes that only *rbcL* and *matK* are approved and required barcode regions for land plants. CBOL will inform GenBank that sequence records submitted to the International Nucleotide Sequence Database Collaborative are eligible to have the reserved keyword "BARCODE" as stipulated in the barcode data standards.

However, the Executive Committee accepted the review panel's recommendation to reassess the situation in 18 months. The current inability of the proposed plant barcode to resolve more than ~70% of species indicates that improvement in the approach is needed, along with more *rbcL* and *matK* data. A reassessment in 18 months would evaluate progress being made on *matK* primers and sequence assembly techniques for non-coding regions such as *trnH-psbA*.

As stated in the 2009 PNAS paper by the Plant Working Group, "In the short term, where further resolution and universality are required, we envisage that the core *rbcL-matK* barcode will be augmented in individual projects from a flexible short-list of supplementary loci including the noncoding plastid regions examined here (*trnH-psbA, atpF-atpH*, and *psbK-psbI*), and the *trnL* intron which has been advocated for situations involving highly degraded tissue (19). The rapidly evolving internal transcribed spacers of nuclear ribosomal DNA also represent a useful supplementary barcode in taxonomic groups in which direct sequencing of this locus is possible." For this reason, CBOL's Executive Committee encourages the community to collect data on *trnH-psbA* and other non-coding regions as a back-up to *matK* and to enhance protocols for the use of non-coding regions for DNA barcoding.