DNA Barcoding of Plants: *matK* primers for angiosperms

Royal Botanic Garden Edinburgh, 20a Inverleith Row, Edinburgh EH3 5LR, U.K.

Funded by The Gordon & Betty Moore Foundation

v.1.0 February 2012: If you use these primers please send an email to barcoding@rbge.ac.uk. This is to enable us to provide protocol updates and to solicit feedback on how well the primers perform.

Summary:

Following *in silico* and laboratory tests of *matK* barcoding primer combinations, a two-round amplification and sequencing protocol is recommended using the following primer sets:

1st round:

This primer pair routinely recovered the highest number of high quality bi-directional sequences.

xF[1] 5′-TAATTTACGATCAATTCATTC-3′

MALPR1[2] 5′-ACAAGAAAGTCGAAGTAT-3′

2nd round:

Two primer combinations gave similar results on samples not successful in the first round :

matK1R+matK3F:

1R[3] 5′-ACCCAGTCCATCTGGAAATCTTGGTTC-3′

3F[3] 5′-CGTACAGTACTTTTGTGTTTACGAG-3′

472F+1248R:

472F[4] 5′-CCCRTYCATCTGGAAATCTTGGTTC-3′

1248R[4] 5′-GCTRTRATAATGAGAAAGATTTCTGC-3′

These primer pairs increased overall recovery rates to 94% and 96% respectively. While the latter primer pair gives slightly increased overall recovery rates, it is a shorter amplicon resulting in the loss of over 100bp.

Laboratory test details:

Primer pairs were tested on a sample set comprising 470 accessions (representing 50/61 orders and 172 families of angiosperms *sensu* APG3). These samples were selected to include 40% which had previously been successfully sequenced using the Kim 1R+3F primers, and 60% which had failed with those primers. The test set is thus biased towards 'difficult to recover' samples.

In the first-round, 418 of these samples were sequenced successfully (ca. 89%), with the second-round increasing this to 442 and 450 (94% and 96% respectively for the primers outlined above). Of these, 15 were contaminants (e.g. good quality sequences obtained for the incorrect taxon) of which about half were due to incorrectly supplied samples. Some poor quality sequences counted here as fails (often containing mononucleotide repeat regions) were recovered using Phusion *Taq* polymerase with primers xF+ERIR[2] (5′-GCACAAGAAAGTCGAAGTAT-3′) with a modified protocol (given at the end of this document).

Protocols:

PCR (final concentrations in total volume 10µl): 1x PCR buffer, 0.2mM each dNTP, 2.5mM $MgCl^2$, 1M betaine, 0.2M trehalose, 0.5µM each primer, 0.5U Platinum *Taq* (Invitrogen). Template: 1ng.

PCR thermocycling parameters: 94°C for 4 mins; 10 cycles of 94°C for 30 secs, 52°C for 30 secs, 72°C for 1 min; 25 cycles of 88°C for 30 secs, 48°C for 30 secs, 72°C for 1 min; 72°C for 10 mins; storage at 8°C.

PCR clean-up: add 2µl of ExoSAP-IT (diluted 1:10) to 5µl of PCR product.

PCR clean-up thermocycling parameters: 37°C for 30 mins, followed by 80°C for 15 mins then storage at 8°C.

Sequencing PCR (final concentrations in total volume 10ul): 1x sequencing buffer, 0.5µl BigDye, 0.32µM primer, 0.2M trehalose, 1µl template.

Sequencing thermocycling parameters: 25 cycles of 95°C for 30 secs, 50°C for 20 secs, 60°C for 4 mins; storage at 8°C.

Phusion *Taq* (primer pair xF + ERI-R):

PCR (final concentrations in total volume 10µl): 1x PCR buffer, 0.2mM each dNTP, 1M betaine, 0.2M trehalose, 0.5µM each primer, 0.5U Phusion *Taq* (Finnzymes) and 1ng template DNA.

PCR thermocycling parameters: 98°C for 45 secs; 35 cycles of 98°C for 10 secs, 58°C for 30 secs, 72°C for 40 secs; 72°C for 10 mins; storage at 8°C.

References:

[1] Ford *et al* (2009) Botanical Journal of the Linnaean Society 159, 1-11.
[2] Dunning & Savolainen (2010) Botanical Journal of the Linnaean Society 164, 1-9.
[3] KJ Kim, unpublished.
[4] Yu et al (2011) Journal of Systematics and Evolution 49, 1-6.

Technical notes:

A range of different PCR additives, component concentrations and thermocycling conditions were tested. The inclusion of 1M betaine and 0.2M trehalose gave increased amplification success compared to all other additives assayed. Dilution of DNA template can also improve amplification. These tests diluted all samples to ca 1ng per 10µl reaction.

Amplicon clean-up protocols tested did not differ significantly, therefore dilute ExoSAP-IT was used for economic reasons. Lack of an amplicon clean-up step resulted in poor quality unreadable sequences in many cases.

Sequencing reactions routinely included 0.2M trehalose as this increased read length by up to 150bp.