Anales del Jardín Botánico de Madrid
Vol. 66(2): 285-288
julio-diciembre 2009
ISSN: 0211-1322
doi: 10.3989/ajbm.2231

Universal primers for the amplification of the plastid *trnK/matK* region in land plants

by

Susann Wicke[1] & Dietmar Quandt[2]

[1] Department of Biogeography and Botanical Garden, University of Vienna, Rennweg 14, A-1030 Vienna, Austria
[2] Nees Institute for Plant Biodiversity, Rheinische Friedrich-Wilhelms-Universität Bonn,
Meckenheimer Allee 170, D-53115 Bonn, Germany, quandt@uni-bonn.de

Wicke, S. & Quandt, D. 2009. Universal primers for the amplification of the plastid trnK/matK region in land plants. *Anales Jard. Bot. Madrid* 66(2): 285-288.

Wicke, S. & Quandt, D. 2009. Cebadores universales para la amplificación de la región del plasto trnK/matK en plantas terrestres. *Anales Jard. Bot. Madrid* 66(2): 285-288 (en inglés).

The incorporation of DNA sequence data in order to infer relationships among green plants has revolutionised systematic botany considerably. Although sequence data from all three genomes are available now for innumerable plant species, thus facilitating primer design, numerous plastid gene regions that serve as prominent markers in angiosperm phylogenetics are rarely used among early land plants. Reasons are generally twofold, either genomic reorganisations or mutations in the priming site hamper amplification of the target region with known primers.

One promising candidate region for phylogenetic reconstructions among early diverging land plants that has already been successfully used among various angiosperm lineages is the plastid gene region *trnK*$_{UUU}$-*matK*. This region is universally present in land plants and only few exceptions of a secondary loss or reorganisations are known to date (Kelchner, 2002; Sanderson, 2003; Wolf & al., 2003; Funk & al., 2007; Tsuji & al., 2007, own data – unpublished). The genic region coding for the Lysine transfer RNA is divided into two exons separated by a group II-intron. As a synapomorphy of Embryophytes, an open reading frame, commonly referred to as *matK*, is inserted into domain IV of the intron. It encodes a protein displa-ying high structural similarity to other group-II intron ORFs (Hausner & al., 2006). The *matK* gene is the only intact ORF within plastid group II introns. Therefore, it has been speculated to play an essential role in RNA processing by acting as a putative general maturase for plastid introns (Ems & al., 1995; Vogel & al., 1997), though its exact function remains unclear.

The *matK* coding region exhibits an exceptional feature by showing a nearly equal distribution of nucleotide substitutions of the first, second and third codon position (Hilu & Liang, 1997; Müller & al., 2006). Thus, the *matK* gene evolves more rapidly in contrast to other plastid genes, despite underlying transcriptional and functional constraints. In combination with the fast evolving but also constrained *trnK*$_{UUU}$-intron, this region offers, due to the mosaic structure of different variability degrees across the region, a suitable marker to address both deep and low level phylogenetic questions. The entire region ranges from approximately 2.2 kb (liverworts) to 2.6 kb (seed plants) in size (own data – not shown).

On the one hand, slow evolving coding regions (e.g. genes for photosynthesis, ribosomal subunits, etc.) have a high likelihood of alignability across a wide set of land plant lineages, but often fail to confidently resolve deep nodes and generally achieve no resolution within rapidly radiated lineages (compare e.g., Buck & al., 2000; Wanke & al., 2007). On the other hand, non-coding and fast evolving coding markers are more troublesome in terms of primary

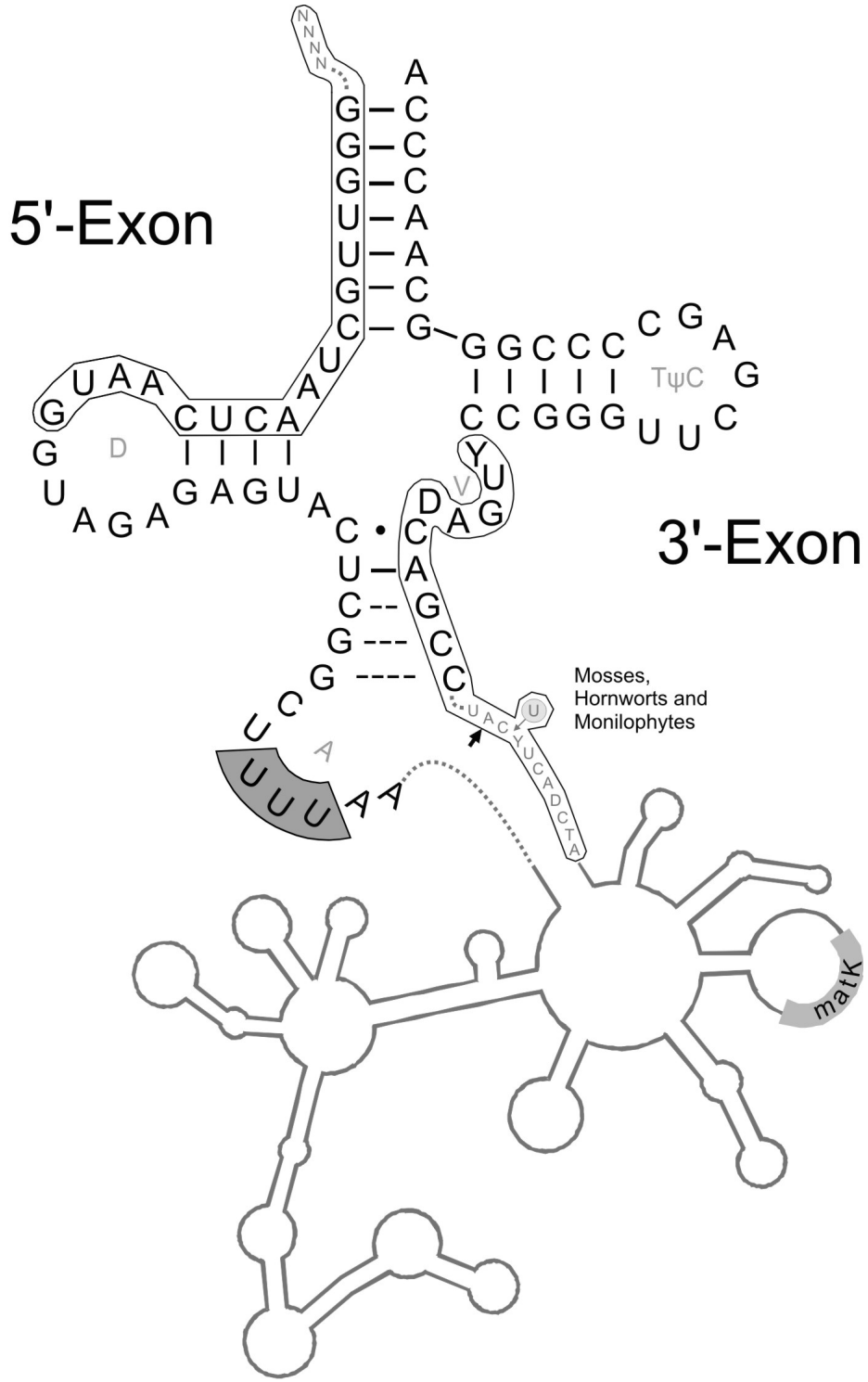

Fig. 1. Land plant consensus secondary structure of the Lysine (UUU) tRNA containing the *matK*-open reading frame. The $trnK_{UUU}$ coding region is widely conserved with only a few substitutions observed within the short variable loop (v-loop). Wobble bases represent nucleotide substitutions observed within different land plant lineages. The group II-intron, which separates both *trnK*-exons is shown in grey. The intron-exon as well as exon-spacer junctions are indicated by dotted lines. Nucleotide sequences of the adjacent non-coding regions (intron and spacer) are shown in smaller grey letters. Priming sites of the MG15 (located in the 5'exon), as well as the MG1 (derived from trnK2R and located in the 3'exon) of previous studies are framed. Four Ns in the MG15 tail indicate that the exon-spacer junction is highly variable, resulting in a mismatch of the primer across land plants. Similarly, an additional nucleotide in the closing helix of the group II intron hampers either MG1 or trnK2R annealing. Thus primers universal for land plants necessarily need to be located behind the mononucleotide indel as indicated by a black arrow. The secondary structure representation of the group II intron was adopted from Kelchner (2002).

homology assessment and thus less frequently used at deeper levels. However, once the rather time consuming alignment has been performed, fast evolving regions – especially *matK* – not only tend to provide the highest phylogenetic structure, they also offer the desired phylogenetic information even at deeper nodes (Borsch & al., 2003; Müller & al., 2006; Borsch & Quandt, 2009). For flowering plants, this marker is already well established and easily accessible with standard PCR primers published by Liang & Hilu (1996) and Hilu & Liang (1997) and by Johnson & Soltis (1995). The use of these primers for early diverging land plant lineages, however, has been problematic. Long and colleagues (2000) suggested a new set of primers suitable for bryophytes. But still, those are not widely applicable for other early land plant lineages. The reason for this is the localisation of the primers displayed in Fig. 1. The forward MG15 primer (Liang & Hilu, 1996) spans the spacer-exon junction of the 5' *trnK*-exon. The short spacer sequence shows a large variability among land plant lineages and reveals numerous mismatches. However, the conserved structure at the hydroxyl end in combination with a lower annealing temperature still allows proper PCR amplification. The available reverse primers (Johnson & Soltis, 1995; Liang & Hilu, 1996) are located at the intron-exon junction of the 3' *trnK*-exon. Mosses, hornworts and ferns share a thymidin residue at position -4 that is lacking in the remaining land plant lineages. The primer heads of the reverse trn2RB primer and its derived MG1 primer span this region and cannot anneal properly. As a result, synthesis activity of polymerase activity is strongly impaired by mismatches affecting the 3'-primerhead (Thweatt & al., 1990; Liang & Pardee, 1992; Ayyadevara & al., 2000).

Based on these facts, we have designed new universal primers for the $trnK_{UUU}$-*matK* region. The forward primer (trnK-F: 5' – GGG TTG CTA ACT CAA TGG TAG AG – 3') spans the 5'-acceptor stem of the tRNA and, thus, provides sufficient specificity. A variety of reverse primers have been designed for both direct amplification and nested PCR to help overcome difficult amplification. Three out of four new reverse primers span or reach into the V-loop of the tRNA in order to avoid mispriming (trnK-R1: 5' – GAA CCC GGA ACT HGT CGG AT – 3'; trnK-R2: 5' – TCG AAC CCG GAA CTH GTC GG – 3'; trnK-R3: 5' – CGG GGC TCG AAC CCG GA – 3'). The fourth is located across the 3'acceptor stem region (trnK-R4: 5' – TGG GTT GCC CGG GGC TCG AAC – 3'). All primer combinations have proved to work well in all land plant lineages (data not shown). However, trnK-R1 and 3' trnK-R3 turned out to work best as a nested primer, whereas trnK-R2 and trnK-R4 generate high yields of *trnK*/*matK*-amplicons.

Methods

Nucleotide data of $trnK_{UUU}$-Exons and adjacent spacer/intron sequences were downloaded from Genbank and manually aligned using the *Phylogenetic Data Editor* (PhyDE, www.phyde.de). Primer design was carried out employing the *PhyDE*-plugin *Seqstate* (Müller, 2005). Primer functionality and applicability was tested on more than 100 species representing the diversity of liverworts, mosses, hornworts, lycophytes, and monilophytes. In addition, the primers were already successfully used in various studies dealing with flowering plants (Worberg & al., 2007; Kassahun & al., 2007). PCR products could be generated easily using peQlab SAWADY *Taq*-polymerase (25 µl reaction containing 1.0 U *Taq* DNA polymerase, 1 mM dNTPs-Mix, 1 × *Taq* buffer, 3.0 mM $MgCl_2$, 20 mM of each amplification primer and between 10-50 ng of template DNA) in combination with a one-step touchdown PCR-program (1 cycle at 90 sec at 96 °C, 60 sec at 50 °C, 120 sec at 68 °C, 35 cycles at 30 sec at 95 °C, 60 sec at 48 °C, 120 sec at 68 °C, subsequent final elongation of 20 min at 68 °C).

Acknowledgements

Research was initially funded by the European Commission's BIODIBERIA HUMAN POTENTIAL PROGRAMME (DQ), and later by SYNTHESYS (SW, DQ), which is financed by the European Community Research Infrastructure Action under the FP6 "Structuring the European Research Area" Programme (http://www.sysnthesys.info). Additional funds were received from the Deutsche Forschungsgemeinschaft (DFG QU 153/2 & 153/3). We are grateful for the generous support of the people at the Royal Botanical Garden Madrid. Special thanks are due to our host Jesús Muñoz and the former lab manager María-Paz Martín Esteban.

References

Ayyadevara, S., Thaden, J.J. & Shmookler Reis, R.J. 2000. Discrimination of Primer 3'-Nucleotide Mismatch by Taq DNA Polymerase during Polymerase Chain Reaction. *Analytical Biochemistry* 284: 11-18.

Borsch, T., Hilu, K.W., Quandt, D., Wilde, V., Neinhuis, C. & Barthlott, W. 2003. Noncoding plastid *trn*T-*trn*F sequences reveal a well resolved phylogeny of basal angiosperms. *Journal of Evolutionary Biology* 16: 558-576.

Borsch, T. & Quandt, D. 2009. Mutational dynamics and phylogenetic utility of non-coding chloroplast DNA. *Plant Systematics and Evolution* 282: 169-199.

Buck, W.R., Goffinet, B. & Shaw, A.J. 2000. Testing morphological concepts of orders of pleurocarpous mosses (Bryophyta) using phylogenetic reconstructions based on *trnL-trnF* and *rps4* sequences. *Molecular Phylogenetics and Evolution* 16: 180-198.

Ems, S.C., Morden, C.W., Dixon, C.K., Wolfe, K.H., de Pamphilis, C.W. & Palmer, J.D. 1995. Transcription, splicing and

editing of plastid RNAs in the nonphotosynthetic plant Epifagus virginiana. *Plant Molecular Biology* 29: 721-733.

Funk, H.T., Berg, S., Krupinska, K., Maier, U.G. & Krause, K. 2007. Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, Cuscuta reflexa and Cuscuta gronovii. *BMC Plant Biology* 7:45 doi:10.1186/1471-2229-7-45.

Hausner, G., Olson, R., Simon, D., Johnson, I., Sanders, E.R., Karol, K.G., McCourt, R.M. & Zimmerly, S. 2006. Origin and Evolution of the Chloroplast trnK (matK) Intron: A Model for Evolution of Group II Intron RNA Structures. *Molecular Biology and Evolution* 23(2): 380-391.

Hilu, K.W. & Liang, H. 1997. The *mat*K gene: sequence variation and application in plant systematics. *American Journal of Botany* 84: 830-839.

Johnson, L.A. & Soltis, D.E. 1995. Phylogenetic inference using *mat*K sequences. *Annals of the Missouri Botanical Garden* 82: 149-175.

Kassahun, T., Borsch, T., Govers, K. & Endashaw, B. 2007. Characterization of Coffea chloroplast microsatellites and evidence for the recent divergence of C. arabica and C. eugenioides chloroplast genomes. *Genome* 50: 1112-1129.

Kelchner, S.A. 2002. Group II introns as phylogenetic tools: structure, function and evolutionary constraints. *American Journal of Botany* 89: 1651-1669.

Liang, H. & Hilu, K.W. 1996. Application of the *mat*K gene sequences to grass systematics. *Canadian Journal of Botany* 74: 125-134

Liang, P. & Pardee, B.A. 1992. Differential display of eukaryotic mRNA by means of polymerase chain reaction. *Science* 257: 967-971.

Long, D.G., Möller, M. & Preston J. 2000. Phylogenetic Relationships of Asterella (Aytoniaceae, Marchantiopsida) Inferred from Chloroplast DNA Sequences. *The Bryologist* 103(4): 625-644.

Müller, K.F. 2005. SeqState - primer design and sequence statistics for phylogenetic DNA data sets. *Applied Bioinformatics* 4: 65-69

Müller, K.F., Borsch, T. & Hilu, K.W. 2006. Phylogenetic utility of rapidly evolving DNA at high taxonomical levels: contrasting *matK*, *trn*T-F and *rbc*L in basal angiosperms. *Molecular Phylogeny and Evolution* 41: 99-117

Sanderson, M.J. 2003. Molecular data from 27 proteins do not support a Precambrian origin of land plants. *American Journal of Botany* 90: 954-956.

Thweatt, R., Goldstein, S. & Shmookler Reis, R.J. 1990. A universal primer mixture for sequence determination at the 3'ends of cDNAs. *Analytical biochemistry* 190: 314-316.

Tsuji, S., Ueda, K., Nishiyama, T., Hasebe, M., Yoshikawa, S., Konagaya, A., Nishiuchi T. & Yamaguchi, K. 2007. The chloroplast genome from a lycophyte (microphyllophyte), Selaginella uncinata, has a unique inversion, transpositions and many gene losses. *Journal of Plant Research* 120(2): 281-290.

Vogel, J., Hübschmann, T., Börner, T. & Hess, W.R. 1997. Splicing and Intron-internal RNA Editing of *trn*K-*mat*K Transcripts in Barley Plastids: Support for *mat*K as an Essential Splice Factor. *Journal of Molecular Biology* 270: 179-187.

Wanke, S., Jaramillo, M.A., Borsch, T., Samain, M.S., Quandt, D. & Neinhuis, C. 2007. Evolution of the Piperales – *mat*K and *trn*K intron sequence data reveals a lineage specific resolution contrast. *Molecular Phylogenetics and Evolution* 42: 477-497

Wolf, P.G., Rowe, C.A., Sinclair, R.B. & Hasebe, M. 2003. Complete nucleotide sequence of the chloroplast genome from a leptosporangiate fern, Adiantum capillus-veneris L. *DNA Research* 10(2): 59-65.

Worberg, A., Quandt, D., Barniske, A.-M., Löhne, C., Hilu, K.W. & Borsch, T. 2007. Phylogeny of basal eudicots: insights from non-coding and rapidly evolving DNA. *Organisms, Diversity and Evolution* 7: 55-77.

Associate Editor: P. Catalán
Received: 8-IV-2009
Accepted: 28-VI-2009