Engineered DNA Polymerase Improves PCR Results for Plastid DNA

Authors: Schori, Melanie, Appel, Maryke, Kitko, Alexarae, and Showalter, Allan M.

Source: Applications in Plant Sciences, 1(2)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1200519

Applications in Plant Sciences

Applications in Plant Sciences 2013 1(2): 1200519

Protocol Note

Engineered DNA polymerase improves PCR results for plastid DNA[1]

Melanie Schori[2,5], Maryke Appel[3], AlexaRae Kitko[2,4], and Allan M. Showalter[2]

[2]Department of Environmental and Plant Biology, Molecular and Cell Biology Program, Porter Hall 315, Ohio University, Athens, Ohio 45701 USA; [3]Kapa Biosystems, 600 W. Cummings Park, Suite 2250, Woburn, Massachusetts 01801 USA; and [4]5833 Stearns Road, North Olmsted, Ohio 44070 USA

- *Premise of the study:* Secondary metabolites often inhibit PCR and sequencing reactions in extractions from plant material, especially from silica-dried and herbarium material. A DNA polymerase that is tolerant to inhibitors improves PCR results.
- *Methods and Results:* A novel DNA amplification system, including a DNA polymerase engineered via directed evolution for improved tolerance to common plant-derived PCR inhibitors, was evaluated and PCR parameters optimized for three species. An additional 31 species were then tested with the engineered enzyme and optimized protocol, as well as with regular *Taq* polymerase.
- *Conclusions:* PCR products and high-quality sequence data were obtained for 96% of samples for *rbcL* and 79% for *matK*, compared to 29% and 21% with regular *Taq* polymerase.

Key words: directed evolution; engineered KAPA3G DNA Polymerase; *matK*; PCR inhibition; *rbcL*.

Plants contain many secondary metabolites, including phenolics, polysaccharides, and glycoproteins, that can interfere with DNA extraction, PCR, and cycle sequencing. Multiple extraction protocols aimed at reducing or removing inhibitory compounds (e.g., Olmstead and Palmer, 1994; Setoguchi and Ohba, 1995; Hughey et al., 2001; Drábková et al., 2002; Malvick and Grunden, 2005), or attenuating their effects on PCR efficiency (Saunders, 1993; De Boer et al., 1995), have been developed, but these are often not effective. Certain taxa pose significant challenges to successful PCR and sequencing, even in cases where purified genomic DNA is used. While *rbcL* is generally considered an easy region to amplify and sequence with standard primers (Kress and Erickson, 2007; Hollingsworth et al., 2009), the first author had problems amplifying or sequencing the gene from multiple medicinal plants from Pakistan, including *Amaranthus* sp. (Chenopodiaceae), *Anethum graveolens* L. (Apiaceae), *Butea monosperma* (Lam.) Taub. (Fabaceae), *Fagonia indica* Burm. f. (Zygophyllaceae), *Senna* sp. (Fabaceae), and *Trachyspermum ammi* (L.) Sprague (Apiaceae). DNA was extracted from silica-dried or air-dried samples before PCR was attempted with regular *Taq* polymerase. An extraction of *Anethum*, prepared from fresh material, amplified and sequenced cleanly, suggesting that secondary metabolites in the dried material (which were not effectively removed with a commercial DNA purification kit), rather than suboptimal PCR parameters (identical for both samples), reduced amplification efficiency and inhibited sequencing. Inhibitors in dried material pose a serious challenge because fresh tissue is often not available.

The problem of poor PCR efficiency can be addressed at the polymerase level. Kapa Biosystems (Woburn, Massachusetts, USA) recently developed an enzyme with specific tolerance to common plant inhibitors. "KAPA3G" DNA Polymerase was derived from a previously engineered, more processive variant of *Taq* DNA polymerase (processivity reflects the average number of nucleotides added by a DNA polymerase per association/dissociation event with the template; processive enzymes synthesize DNA more quickly and are more efficient in the presence of inhibitors). In short, a randomized gene library of the parental "KAPA2G" DNA polymerase gene was generated and expressed in *E. coli.* Individual bacterial cells, each containing both the expressed, mutant DNA polymerase protein, as well as the gene encoding that variant, were compartmentalized in a water-in-oil emulsion (Griffiths and Tawfik, 2006). In this system, each mutant enzyme was required to amplify its own gene in the presence of secondary metabolites derived from several different plant species. After several rounds of selection with increasing levels of inhibition pressure, gene variants coding for polymerases with improved tolerance to plant inhibitors were exponentially enriched over variants with no advantage. The KAPA3G DNA Polymerase that was evolved in this manner was blended with a small quantity of an engineered high-fidelity enzyme, to allow for the efficient amplification of DNA fragments >5 kb from plant samples. KAPA3G was effectively tested for PCR with purified plant DNA, crude plant extracts, and in direct PCR from leaf discs or seeds of a variety of crop plants (Appendix S1) before the KAPA3G Plant PCR Kit was released. This report constitutes the first study of the effectiveness of KAPA3G DNA Polymerase using dried

[1]Manuscript received 27 September 2012; revision accepted 15 November 2012.

The authors thank V. Nadella and R. Yoho at the Ohio University Genomics Facility for technical assistance. Sample kits for optimization purposes were supplied by Kapa Biosystems. This research was funded by grant PGA-P210852 from the National Academy of Sciences to A. M. Showalter. Maryke Appel works for Kapa Biosystems.

[5]Author for correspondence: schori@ohio.edu

doi:10.3732/apps.1200519

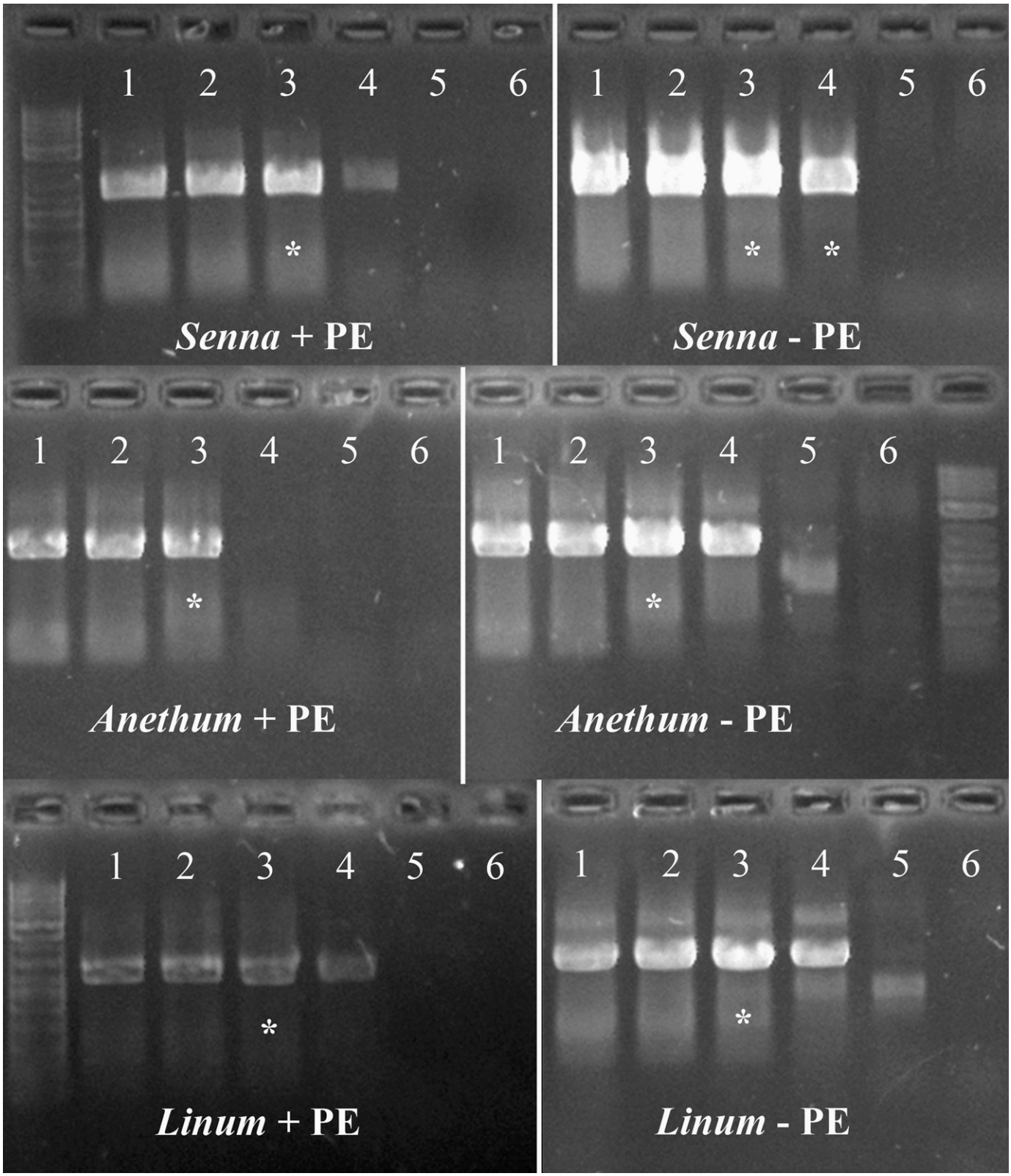

Fig. 1. Annealing temperature optimization by gradient PCR for the KAPA3G Plant PCR Kit, in the presence (+ PE) or absence (– PE) of the Plant Enhancer. Overall amplification was greater without Plant Enhancer. Numbers correspond to different annealing temperatures over a 20°C gradient: 1 = 50°C, 2 = 54°C, 3 = 58°C, 4 = 62°C, 5 = 66°C, 6 = 70°C. The highest temperature that resulted in successful product for all samples was 58°C. Marked PCR products (*) were submitted for sequencing with *rbcL* 1F, 636F, 724R, and 1460R primers.

TABLE 1. Comparison of *rbcL* sequencing data quality for *Linum usitatissimum*, *Anethum graveolens*, and *Senna* sp. using *Taq* polymerase and the KAPA3G PCR Kit with and without Plant Enhancer.

Species	*rbcL* Primer	*Taq* Phred Q20	KAPA3G + Enhancer Phred Q20	KAPA3G – Enhancer Phred Q20
Linum	1F	**956**	464	942
Linum	636F	**866**	864	837
Linum	724R	579	0	**776**
Linum	1460R	617	777	**983**
Anethum	1F	0	838	**989**
Anethum	636F	0	0	**792**
Anethum	724R	0	650	**718**
Anethum	1460R	0	**870**	146
Senna	1F	—	12	**948**
Senna	636F	—	0	**797**
Senna	724R	—	**837**	683
Senna	1460R	—	871	**895**
Senna 62°C	1F			598
Senna 62°C	636F			**909**
Senna 62°C	724R			**877**
Senna 62°C	1460R			466

Note: An annealing temperature of 48°C was used with regular *Taq*, whereas the optimal annealing temperature with the KAPA3G chemistry was 58°C for all samples except *Senna* (62°C). *Senna* did not amplify with *Taq* polymerase. The highest Q20 values for each primer given the three different PCR master mixes are in bold.

material from noncrop plants, and documents the potential advantages of the KAPA3G Plant PCR Kit for a wide range of species. A PCR optimization is presented to aid researchers in selecting the appropriate annealing temperature and $MgCl_2$ concentration for their specific assays. PCR results with the KAPA3G enzyme are compared to those with regular *Taq* polymerase.

METHODS AND RESULTS

Three DNA extracts that produced varying degrees of amplification and sequencing success with regular *Taq* polymerase were chosen for initial *rbcL* optimization (primers 1F [Fay et al., 1997] and 1460R [Fay et al., 1998; Cuénoud et al., 2002]) with the KAPA3G Plant PCR Kit. See Appendix 1 for voucher information for all species included in the study. Wild collections were not georeferenced at the time of collection. Mini-extractions for *Linum usitatissimum* L. (Linaceae) and *Anethum graveolens* (Apiaceae) (both silica-dried) and *Senna* sp. (Fabaceae) (air-dried) were prepared using a standard cetyltrimethylammonium bromide (CTAB) protocol (Doyle, 1991) and purified using the UltraClean 15 kit (MO BIO, Carlsbad, California, USA). PCR for *rbcL*, *matK*, and *psbA-trnH* had been attempted using ReadyMix PCR master mix with *Taq* polymerase (Sigma, St. Louis, Missouri, USA). The following thermal cycler program was used for *rbcL* and *matK* PCR with *Taq* polymerase: 94°C 5 min; 30 cycles: 94°C 1 min, 48°C 1 min, 72°C 1 min; 72°C 7 min. All three regions were successfully amplified and sequenced for *Linum*, but *rbcL* failed to sequence for *Anethum* and did not amplify for *Senna*, although the *psbA-trnH* spacer was sequenced for both. *Linum* was selected for the KAPA3G Plant PCR Kit evaluation as it had amplified and sequenced with *Taq* polymerase, while *Anethum* was chosen because it had amplified but failed to sequence, and *Senna* as it had not amplified at all. The KAPA3G Plant PCR Kit includes an optional Plant Enhancer, a reducing agent that improves amplification efficiency for some types of samples through an unknown mechanism. Two sets of reactions were run for each taxon, one with 0.5 µL (1×) Enhancer and one without. Each reaction contained the KAPA3G Plant Buffer (1× final concentration, includes dNTPs at 0.2 mM each), $MgCl_2$ (2 mM final concentration), 1 unit DNA polymerase, primers at a final concentration of 0.3 µM each, and PCR-grade water to bring the volume to 50 µL. An annealing temperature gradient PCR was performed, in increments of 4°C from 50°C to 70°C, using a Veriti Thermal Cycler (Applied Biosystems, Carlsbad, California, USA) and the following cycling parameters: 95°C 10 min; 40 cycles: 95°C 20 s, 50–70°C [gradient] 15 s, 72°C 90 s; 72°C 90 s. The gradient PCR identified the highest temperature at which amplification was successful for all samples (58°C). To test the amplification quality, six of the best PCR products (corresponding to the brightest bands in a 1% agarose gel) were selected for sequencing: *Linum*, *Anethum*, and *Senna* generated with an annealing temperature of 58°C, with and without Enhancer. The best overall PCR product (*Senna* without Enhancer, generated with an annealing temperature of 62°C) was also sequenced for

TABLE 2. PCR and sequencing success of 31 species for *rbcL* and *matK* using *Taq* polymerase or the KAPA3G Plant PCR Kit.

	rbcL		*matK* 390/1360	
Species	Regular *Taq*	KAPA3G Plant PCR Kit	Regular *Taq*	KAPA3G Plant PCR Kit
Acacia nilotica (L.) Willd. ex Delile	–	+	–	+[a]
Achyranthes aspera L.	–	+	–	+[a]
Argemone mexicana L.	+		+[b]	
Artemisia absinthium L.	–	+	–	+
Asparagus racemosus Willd.	–	+	–	+[a]
Buxus papillosa C. K. Schneid.	+		+	+
Convolvulus arvensis L.	–	+	–	+
Crocus sativus L.	+		+	
Cuminum cyminum L.	–	+	+	+
Euphorbia helioscopia L.	+			+
Fumaria indica (Hausskn.) Pugsley	–	+	–	–
Fumaria indica	+		+[b]	
Galium aparine L.	–	+		+
Hygrophila auriculata (Schumach.) Heine	–	–	–	–
Justicia adhatoda L.	–	+	+[b]	+[a]
Lathyrus aphaca L.	+		–	+[c]
Launaea nudicaulis Hook. f.	–	+	–	+
Lawsonia inermis L.	–	+	–	+
Lepidium didymum L.	–	+	–	+[c]
Matricaria chamomilla L. var. *recutita* (L.) Fiori	–	+	–	+[a,c]
Mucuna pruriens (L.) DC.	–	+	–	+[d]
Plumbago auriculata Lam.	–	+	–	+[a]
Schinus molle L.	–	+	–	+[d]
Solanum surattense Burm. f.	–	+	–	+[a]
Taraxacum officinale F. H. Wigg.	–	+		+[a,d]
Trichodesma indicum (L.) Sm.	–	+	–	+
Urtica dioica L.	–	+[f]	–	–
Veronica polita Fr.	+		–	+
Vicia faba L.	+		–	–
Vitex negundo L.	–	+	–	+
Withania somnifera (L.) Dunal	+		+[e]	+
No. of PCR products obtained	9/31 (29%)	21/22 (95%)	7/29 (24%)	24/28 (86%)

Note: Unless otherwise noted, a + indicates both successful PCR and sequencing. PCR for *matK* was not attempted with *Taq* for *Euphorbia helioscopia*, *Galium aparine*, or *Taraxacum officinale*. Gene regions that were successfully sequenced after PCR with *Taq* were generally not tried with the KAPA3G Plant PCR Kit.

[a] 2 mM $MgCl_2$; [b] *matK* 1F/3R (Sang et al., 1997); [c] PCR product failed to sequence; [d] Faint band, not submitted for sequencing; [e] Poor quality sequence; [f] CTAB extract amplified with 2 mM $MgCl_2$, PowerPlant Pro extract amplified with 1.5 mM $MgCl_2$.

comparison. PCR products were cleaned with the Wizard SV Gel and PCR Clean-Up System (Promega Corporation, Madison, Wisconsin, USA). DNA sequences were generated at Ohio University's Genomics Facility and analyzed using an ABI 3130xl Genetic Analyzer (Applied Biosystems, Carlsbad, California, USA). Each sequencing reaction included 2 µL 5× buffer (Applied Biosystems), 0.5 µL dimethyl sulfoxide (DMSO; Sigma), 0.5 µL BigDye (Applied Biosystems), 0.1 µL ThermoFidelase (Fidelity Systems, Gaithersburg, Maryland, USA), 10–40 ng template DNA, and PCR-grade water for a total volume of 8 µL. Cycle sequencing products were cleaned with the BigDye XTerminator Purification Kit (Applied Biosystems). Phred Q20 values (Ewing et al., 1998) were used as an initial indication of sequence quality. External *rbcL* primers 1F and 1460R, and internal primers 636F and 724R (Fay et al., 1997), were used for sequencing.

Results of the PCR optimization are shown in Fig. 1. Amplification was successful at annealing temperatures from 50–62°C, although amplification at 62°C was reduced or failed when Enhancer was present. More product was produced without Enhancer, but more nonspecific amplification occurred. *Senna*, which did not amplify for *rbcL* using *Taq* polymerase, amplified strongly using the KAPA3G enzyme. Sequencing results for *rbcL* primers 1F, 636F, 724R, and 1460R are presented in Table 1, with partial *rbcL* 1F chromatograms in Appendices S2 and S3. Sequence data for *Linum* 1F and 636F were of higher quality from PCR products using *Taq* polymerase, whereas sequence data for all other taxa were of higher quality with the KAPA3G enzyme. With the exception of *Anethum* 1460R and *Senna* 724R at an annealing temperature of 58°C, sequence data were of a higher quality from samples without Enhancer. This suggested that residual Enhancer (carried through PCR clean-up) may have inhibited the cycle sequencing reaction. However, the results of a second optimization did not support this conclusion. A second round of optimization for *rbcL* was performed with the *Linum*, *Anethum*, and *Senna* extracts to reduce nonspecific amplification, although no significant improvements were observed for these particular species. See Appendices S4 and S5 for the protocol and results, which tested the effects of different thermal cycling programs, $MgCl_2$ concentrations, and the presence/absence of Enhancer.

Extracts of an additional 31 species from 23 different families, prepared with the same methods outlined above, were tested for *rbcL* (Table 2), first with *Taq* polymerase, and then with the KAPA3G enzyme using the optimized cycling program with an annealing temperature of 58°C. Nine out of 31 samples (29%) amplified and sequenced for *rbcL* with *Taq*, whereas 21 out of 22 samples (95%) that failed with *Taq* amplified and sequenced with the KAPA3G enzyme (1.5 mM $MgCl_2$, no Enhancer). This success rate is much higher than the best *rbcL* PCR rate (26%) reported by Särkinen et al. (2012) for several different DNA polymerase enzymes, although extracts from much older herbarium specimens were used in their study. The same initial optimization outlined above was performed for *matK* 390F/1360R (Cuénoud et al., 2002), and an annealing temperature of 54°C was selected for this assay. Three out of 26 samples (12%) amplified and sequenced with *Taq* polymerase for *matK* while 21 out of 28 samples (75%) amplified and sequenced with the KAPA3G enzyme (Table 2). A higher concentration (2 mM) of $MgCl_2$ was required for successful PCR of eight of these species. A few samples (e.g., *Hygrophila*, *Urtica*) did not amplify for one or both gene regions with the KAPA3G enzyme. These samples were characterized by abundant mucilage during the extraction process, and purifying the genomic DNA did not remove all the mucilage. A nonmucilaginous extract prepared from seeds (market sample) of *Hygrophila* did amplify successfully for *matK* (but not *rbcL*) with the KAPA3G enzyme after PCR with *Taq* polymerase failed. An *Urtica* extract prepared with the PowerPlant Pro DNA Isolation Kit (MO BIO) amplified readily for *rbcL* (but not *matK*) at 1.5 mM $MgCl_2$ with the KAPA3G enzyme (Table 2). For certain species of Lamiaceae (*Ajuga*, *Mentha*, *Ocimum*, results not shown here), successful amplification of *rbcL* and *matK* was achieved with the KAPA3G enzyme from dirty pellets (not purified after CTAB extraction), while other species (*Lycopus*, *Nepeta*, *Origanum*) failed to amplify until genomic DNA had been purified or extracted with the PowerPlant Pro DNA Isolation Kit. Taken together, these results suggest that while the KAPA3G enzyme offers much higher success rates than *Taq* polymerase, PCR from plant samples remains challenging in the presence of high levels of inhibitors, particularly when primers are not perfectly matched to target sequences.

CONCLUSIONS

This study demonstrated that the KAPA3G Plant PCR Kit successfully amplified DNA from extracts that failed with *Taq*. Quality sequence data were obtained from species from 24 different families. The variable results obtained with *Taq* polymerase and the KAPA3G Plant PCR Kit indicate that PCR success and sequence quality may be as much a function of the taxon as the methodologies used. Differences in secondary metabolites presumably account for some of this variation. Although the KAPA3G Plant PCR Kit did not always lead to high-quality sequence data, it effectively amplified DNA that failed to amplify with *Taq* polymerase. The KAPA3G Plant PCR Kit can therefore be a very useful tool for plant biologists working with difficult taxa that have failed to amplify with *Taq* polymerase. We recommend using the optimization protocol (Appendix 2) to select the best annealing temperature for a specific assay, and then performing the PCR with 1.5 mM $MgCl_2$ and no Enhancer. If the PCR fails, increasing the $MgCl_2$ concentration (2 mM) and/or adding Enhancer should be tried as these proved to be critical for successful PCR for certain taxa.

LITERATURE CITED

Cuénoud, P., V. Savolainen, L. W. Chatrou, M. Powell, R. J. Grayer, and M. W. Chase. 2002. Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL*, *atpB*, and *matK* DNA sequences. *American Journal of Botany* 89: 132–144.

De Boer, S. H., L. J. Ward, X. Li, and S. Chittaranjan. 1995. Attenuation of PCR inhibition in the presence of plant compounds by addition of BLOTTO. *Nucleic Acids Research* 23: 2567–2568.

Doyle, J. J. 1991. DNA protocols for plants. *In* G. Hewitt, A. W. B. Johnson, and J. P. W. Young [eds.], Molecular techniques in taxonomy, 283–293. NATO ASI Series H, Cell Biology 57. Springer-Verlag, Berlin, Germany.

Drábková, L., J. Kirschner, and Č. Vlček. 2002. Comparison of seven DNA extraction and amplification protocols in historical herbarium specimens of Juncaceae. *Plant Molecular Biology Reporter* 20: 161–175.

Ewing, B., L. Hillier, M. C. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using *phred*. I. Accuracy assessment. *Genome Research* 8: 175–185.

Fay, M. F., S. M. Swensen, and M. W. Chase. 1997. Taxonomic affinities of *Medusagyne oppositifolia* (Medusagynaceae). *Kew Bulletin* 52: 111–120.

Fay, M. F., C. Bayer, W. S. Alverson, A. Y. de Bruijn, and M. W. Chase. 1998. Plastid *rbcL* sequence data indicate a close affinity between *Digodendron* and *Bixa*. *Taxon* 47: 43–50.

Griffiths, A. D., and D. S. Tawfik. 2006. Miniaturising the laboratory in emulsion droplets. *Trends in Biotechnology* 24: 395–402.

Hollingsworth, P. M., L. L. Forrest, S. L. Spouge, M. Hajibabaei, S. Ratnasingham, M. van der Bank, et al. 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences, USA* 106: 12794–12797.

Hughey, J. R., P. C. Silva, and M. H. Hommersand. 2001. Solving taxonomic and nomenclatural problems in Pacific Gigartinaceae (Rhodophyta) using DNA from type material. *Journal of Phycology* 37: 1091–1109.

Kress, J. W., and D. L. Erickson. 2007. A two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the noncoding *trnH-psbA* spacer region. *PLoS ONE* 2: e508.

Malvick, D. K., and E. Grunden. 2005. Isolation of fungal DNA from plant tissues and removal of DNA amplification inhibitors. *Molecular Ecology Notes* 5: 958–960.

Olmstead, R. G., and J. D. Palmer. 1994. Chloroplast DNA systematics: A review of methods and data analysis. *American Journal of Botany* 81: 1205–1224.

Sang, T., D. J. Crawford, and T. F. Stuessy. 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany* 84: 1120–1136.

Särkinen, T., M. Staats, J. E. Richardson, R. S. Cowan, and F. T. Bakker. 2012. How to open the treasure chest? Optimising DNA extraction from herbarium specimens. *PLoS ONE* 7: e43808.

Saunders, G. W. 1993. Gel purification of red algal genomic DNA: An inexpensive and rapid method for the isolation of polymerase chain reaction–friendly DNA. *Journal of Phycology* 29: 251–254.

Setoguchi, H., and H. Ohba. 1995. Phylogenetic relationships in *Crossostylis* (Rhizophoraceae) inferred from restriction site variation of chloroplast DNA. *Journal of Plant Research* 108: 87–92.

Appendix 1. Voucher specimens of medicinal plant species from Pakistan used in this study.

Species	Voucher specimen accession no.[a]	Collection locality[b]	Geographic coordinates
Acacia nilotica	*M. N. Badshah MSAE-1*	Islamabad, Pakistan	
Achyranthes aspera	*M. N. Badshah MSAE-4*	Islamabad, Pakistan	
Anethum graveolens	*M. N. Badshah MSAE-10*	Islamabad, Pakistan	
Argemone mexicana	*M. N. Badshah 206*	Islamabad, Pakistan	
Artemisia absinthium	*Farooq 9(1)-1.absinthium*	Qarshi Herb Garden, Hattar, Pakistan	33°53′50″N, 72°51′43″E
Asparagus racemosus	*Nazir 49(3)-3.racemosus*	Qarshi Herb Garden, Hattar, Pakistan	33°53′50″N, 72°51′43″E
Buxus papillosa	*M. N. Badshah 216*	Islamabad, Pakistan	
Convolvulus arvensis	*M. N. Badshah 218*	Islamabad, Pakistan	
Crocus sativus	*Nazir 45(2)-1.sativus*	Qarshi Herb Garden, Hattar, Pakistan	33°53′50″N, 72°51′43″E
Cuminum cyminum	*Khan 89(2)-1.cyminum*	Qarshi Herb Garden, Hattar, Pakistan	33°53′50″N, 72°51′43″E
Euphorbia helioscopia	*M. N. Badshah 211*	Islamabad, Pakistan	
Fumaria indica	*M. N. Badshah 203*	Islamabad, Pakistan	
Fumaria indica	*Farooq 37(1)-1.indica*	Qarshi Herb Garden, Hattar, Pakistan	33°53′50″N, 72°51′43″E
Galium aparine	*M. N. Badshah 210*	Islamabad, Pakistan	
Hygrophila auriculata	*Farooq 4(2)-1.longifolia*	Qarshi Herb Garden, Hattar, Pakistan	33°53′50″N, 72°51′43″E
Justicia adhatoda	*M. N. Badshah 202*	Islamabad, Pakistan	
Lathyrus aphaca	*M. N. Badshah 217*	Islamabad, Pakistan	
Launaea nudicaulis	*M. N. Badshah 209*	Rawalpindi, Pakistan	
Lawsonia inermis	*Nazir 47(1)-1.inermis*	Qarshi Herb Garden, Hattar, Pakistan	33°53′50″N, 72°51′43″E
Lepidium didymium	*M. N. Badshah 207*	Islamabad, Pakistan	
Linum usitatissimum	*N. Allam MSAE-8*	Islamabad, Pakistan	
Matricaria chamomilla var. *recutita*	*Farooq 9(14)-1.recutita*	Qarshi Herb Garden, Hattar, Pakistan	33°53′50″N, 72°51′43″E
Mucuna pruriens	*Nazir 9(14)-1.recutita*	Qarshi Herb Garden, Hattar, Pakistan	33°53′50″N, 72°51′43″E
Plumbago auriculata	*Farooq 73(2)-2.auriculata*	Qarshi Herb Garden, Hattar, Pakistan	33°53′50″N, 72°51′43″E
Schinus molle	*M. N. Badshah 215*	Islamabad, Pakistan	
Senna sp. (cf. *auriculata*)	*50945*[c]	Market sample from Sawat Pansar Store and Dawakhana, Aabpara Market, Islamabad	
Solanum surattense	*N. Allam MSAE-13*	Islamabad, Pakistan	
Taraxacum officinale	*M. N. Badshah 201*	Islamabad, Pakistan	
Trichodesma indicum	*M. N. Badshah 212*	Islamabad, Pakistan	
Urtica dioica	*M. N. Badshah 204*	Islamabad, Pakistan	
Veronica polita	*M. N. Badshah 208*	Islamabad, Pakistan	
Vicia faba	*M. N. Badshah 213*	Islamabad, Pakistan	
Vitex negundo	*M. N. Badshah 214*	Islamabad, Pakistan	
Withania somnifera	*M. N. Badshah 205*	Mirpur, Pakistan	

[a] Badshah and Allam collections are at ISL; Qarshi collections are privately held at the company herbarium in Hattar.

[b] Qarshi collections are from cultivated plants in a demonstration herb garden at the company headquarters in Hattar.

[c] *Senna* sample was sold as "aak" (*Calotropis procera*, Apocynaceae) but is a collection of *Senna* leaflets and buds. Voucher is at BHO; leaflets were used for DNA extraction.

Appendix 2. KAPA3G optimization protocol.

This optimization procedure should be performed for each new primer set to select the best (highest effective) annealing temperature with the KAPA3G Plant PCR Kit (Kapa Biosystems, Woburn, Massachusetts, USA). Once the temperature is selected, it may be necessary to increase the final $MgCl_2$ concentration (from 1.5 mM at 1× in the KAPA3G Plant PCR Buffer) and/or add Plant Enhancer to PCR reactions for successful amplification of specific taxa. Four different PCR reaction mixtures (A–D) are recommended for the initial optimization. Certain taxa (e.g., some Clusiaceae) may fail to amplify without Enhancer. See Note 1 for more details.

Gradient PCR program: 95°C 10 min; 40 cycles: 95°C 20 s, 50–70°C (gradient) 15 s, 72°C 90 s; 72°C 90 s; 10°C hold. The gradient was programmed at 4°C intervals on a Veriti 96-well, 0.2 mL Thermal Cycler (Applied Biosystems, Carlsbad, California, USA).

Mix A—1.5 mM $MgCl_2$, with Enhancer (50 µL reaction volume)	
PCR-grade water	20.1 µL
KAPA3G Plant PCR Buffer (2×)†	25 µL
Forward primer (10 µM)*	1.5 µL
Reverse primer (10 µM)*	1.5 µL
KAPA3G Plant DNA Polymerase (2.5 U/µL)	0.4 µL
KAPA Plant PCR Enhancer (100×)	0.5 µL
Template	1.0 µL
Mix B—1.5 mM $MgCl_2$, without Enhancer (50 µL reaction volume)	
PCR-grade water	20.6 µL
KAPA3G Plant PCR Buffer (2×)†	25 µL
Forward primer (10 µM)*	1.5 µL
Reverse primer (10 µM)*	1.5 µL
KAPA3G Plant DNA Polymerase (2.5 U/µL)	0.4 µL
Template	1.0 µL
Mix C—2.0 mM $MgCl_2$, with Enhancer (50 µL reaction volume)	
PCR-grade water	19.1 µL
KAPA3G Plant PCR Buffer (2×)†	25 µL
Forward primer (10 µM)*	1.5 µL
Reverse primer (10 µM)*	1.5 µL
KAPA3G Plant DNA Polymerase (2.5 U/µL)	0.4 µL
$MgCl_2$ (25 mM)	1.0 µL
KAPA Plant PCR Enhancer (100×)	0.5 µL
Template	1.0 µL
Mix D—2.0 mM $MgCl_2$, without Enhancer (50 µL reaction volume)	
PCR-grade water	19.6 µL
KAPA3G Plant PCR Buffer (2×)†	25 µL
Forward primer (10 µM)*	1.5 µL
Reverse primer (10 µM)*	1.5 µL
KAPA3G Plant DNA Polymerase (2.5 U/µL)	0.4 µL
$MgCl_2$ (25 mM)	1.0 µL
Template	1.0 µL

† Includes dNTPs at a final concentration of 0.2 mM each.

* If primer stocks are at a different concentration than 10 µM, include the appropriate volume of each primer for a final concentration of 0.3 µM each, and adjust the volume of water accordingly (for a reaction volume of 50 µL).

NOTES

1. For initial optimization, the following PCR schedule is recommended: first use Mix B (1.5 mM $MgCl_2$, no Enhancer). Use Mix D (2 mM $MgCl_2$, no Enhancer) for samples that did not amplify well with Mix B, then try Mixes A and C at the same time. Systematicists working on one particular group of plants may be able to select an optimal mix for that group, while those working with a broader range of genera or families may need to identify the optimal mix for each taxon. PCR efficiency tends to be lower if Enhancer is added, so if Enhancer is required with any samples, the highest temperature that is effective with Enhancer should be selected.

2. A reaction volume of 50 µL is necessary for the appropriate PCR chemistry. We do not recommend reducing the volume to 25 µL or increasing it to 100 µL as this may adversely affect PCR efficiency.

3. Addition of Enhancer results in PCR products that appear cloudy. This is normal and does not affect the product or sequencing.

4. Clean PCR products with a spin-filter system to ensure removal of all dNTPs, $MgCl_2$, and enzyme prior to sequencing.

Appendix 2. Continued.

TROUBLESHOOTING

Non-specific amplification—Non-specific amplification may occur and appear as discrete bands or a smear above or below the targeted amplification product. The KAPA3G DNA polymerase is very active and will amplify fragments of DNA from spurious annealing events to a greater extent than *Taq* polymerase. Post-PCR clean-up sometimes reduces higher-molecular-weight smears that initially appear to be non-specific amplification products, but could be an artifact from the electrophoretic analysis. Non-specific amplification can be reduced by further optimization of annealing temperature, reducing extension time per cycle, or adding Plant Enhancer. The addition of polyvinylpyrrolidone (PVP), dimethyl sulfoxide (DMSO), or 2-mercaptoethanol (BME) (not used in this study) may also increase the yield of specific product. Try different strategies or combinations of the strategies outlined above for taxa that prove to be particularly problematic (see Appendix S4). Plant Enhancer generally reduces the overall amount of both specific and non-specific amplification products. If all of the above strategies fail, primers may have to be redesigned.

No bands/faint bands on gel—PCR may fail for a variety of reasons. Check the quality of genomic DNA by running 10 μL in a 1% agarose gel. Older extractions that were eluted in water may have degraded over time—always elute, store, and dilute DNA (and primers) in 10 mM Tris-HCl, pH 8.5. Primer mismatches, especially at the 3′ end, will reduce yield and specificity, and can affect the quality of sequence data. Note that "universal" primers may have one or more mismatches for particular taxa. For some species, the final $MgCl_2$ concentration in the reaction could affect the yield of the specific product significantly. If the PCR produced a "clean" band (i.e., non-specific amplification is not an issue), the yield of specific product may be increased by extending the PCR with five or 10 more cycles.