BioOne COMPLETE

Plastid Primers for Angiosperm Phylogenetics and Phylogeography

Author: Prince, Linda M.

Source: Applications in Plant Sciences, 3(6)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1400085

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

Downloaded From: https://complete.bioone.org/journals/Applications-in-Plant-Sciences on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Applications in Plant Sciences

Applications in Plant Sciences 2015 3(6): 1400085

Application Article

Plastid primers for angiosperm phylogenetics and phylogeography[1]

Linda M. Prince[2,3]

[2]Department of Botany, The Field Museum, 1400 S. Lake Shore Drive, Chicago, Illinois 60605 USA

- *Premise of the study:* PCR primers are available for virtually every region of the plastid genome. Selection of which primer pairs to use is second only to selection of the genic region. This is particularly true for research at the species/population interface.
- *Methods:* Primer pairs for 130 regions of the chloroplast genome were evaluated in 12 species distributed across the angiosperms. Likelihood of amplification success was inferred based upon number and location of mismatches to target sequence. Intraspecific sequence variability was evaluated under three different criteria in four species.
- *Results:* Many published primer pairs should work across all taxa sampled, with the exception of failure due to genomic reorganization events. Universal barcoding primers were the least likely to work (65% success). The list of most variable regions for use within species has little in common with the lists identified in prior studies.
- *Discussion:* Published primer sequences should amplify a diversity of flowering plant DNAs, even those designed for specific taxonomic groups. "Universal" primers may have extremely limited utility. There was little consistency in likelihood of amplification success for any given publication across lineages or within lineage across publications.

Key words: comparative sequencing; complete chloroplast genome; cpDNA.

Whole genome sequencing is more available and less expensive than ever before, yet most scientists continue to rely on targeted, comparative sequencing for phylogenetics and phylogeography. Identifying the most appropriate markers to employ has been challenging. Information for model organisms abounds (e.g., grasses; Saski et al., 2007; Bortiri et al., 2008; Leseberg and Duvall, 2009), and a few studies have specifically screened the same set of markers across a diversity of plant groups, ranking the utility of these markers either explicitly or implicitly (Shaw et al., 2005, 2007, 2014). These studies are exceedingly valuable, demonstrating there is no one-size-fits-all answer to the question "which markers?". The second critical question to "which markers" is "which primers?". Hundreds of primer sequences have been published, many designed for specific taxonomic groups. The work presented here was inspired by "The Tortoise and the Hare II" (Shaw et al., 2005), which was the first study to pull together information on a large number of regions commonly in use (at that time) for plant phylogenetics. Our laboratory was also compiling such information, as were many others.

[1]Manuscript received 8 September 2014; revision accepted 24 March 2015.

This work was inspired by phylogenetic combs obtained for *Guzmania* (Bromeliaceae) and *Monardella* (Lamiaceae), and by the work of Joey Shaw. Some plant material or DNA samples were provided by The Desert Botanic Garden, S. Eliason, E. A. Friar, Y.-L. Qiu, S. Vanderplank, and G. Wallace. Cris Martinez and Bill Waggoner assisted with primer screening. Harith Alappat assisted with whole chloroplast genome and primer alignments. Rancho Santa Ana Botanic Garden provided financial support.

[3]Author for correspondence: lprince@fieldmuseum.org

doi:10.3732/apps.1400085

The Tortoise and the Hare II paper was revolutionary in assessing sequence variability for all regions studied across a broad diversity of flowering plants, and providing a ranking of that variability. In the mid-2000s, a small number of complete chloroplast genome sequences were available for land plants and some of those were not annotated (e.g., *Medicago truncatula* Gaertn. [GenBank NC_003119]; Saski et al., 2005). Grivet et al. (2001) were visionary when they moved beyond analyzing regions commonly being used to design primers for lesser-known and potentially faster-evolving regions of the chloroplast genome. They were the first to take advantage of the new genomic data boom, providing a set of 20 universal chloroplast primers designed around the complete chloroplast data from seven flowering plant species. Around the same time, I developed nondegenerate primers for 36 noncoding regions in the large and small single-copy regions of the chloroplast genome (published here). These near-universal primers were designed based on the complete chloroplast genome sequences of 16 flowering plant species (see Appendix 1).

Grivet et al. (2001) and I designed primers, but Shaw et al. (2007) took an even more applied approach when they examined sequences for three different taxon pairs (*Atropa/Nicotiana*, *Lotus/Medicago*, and *Saccharum/Oryza*), specifically searching for faster-evolving regions. Shaw et al. (2014) go one step further, comparing complete chloroplast genome sequences for 25 (primarily congeneric) sister species pairs. They examined sequence diversity for 107 single-copy noncoding regions, providing the most comprehensive analysis to date.

There are now at least 150 primer pairs available to amplify almost every intergenic, intron, and exon region of the chloroplast genome, including portions of the inverted repeats, thanks to the efforts of Shaw et al. (2005, 2007, 2014) and others (Ebert and Peakall, 2009; Scarcelli et al., 2011; Dong et al., 2012,

 © 2015 Prince. Published by the Botanical Society of America. This work is licensed under a Creative Commons Attribution License (CC-BY-NC-SA).

Downloaded From: https://complete.bioone.org/journals/Applications-in-Plant-Sciences on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

2013). Not surprisingly, although all worked independently, many of the same regions were explored (Appendix 2) and, in some cases, identical or nearly identical primers were designed. The push to identify faster-evolving regions was, in part, spurred by groups of organisms with exceptionally slowly evolving chloroplast genomes such as Bromeliaceae (Gaut et al., 1992) and Arecaceae (Asmussen and Chase, 2001). Heinze provided access to a comprehensive database of chloroplast primers in 2007 (Heinze, 2007). The database is periodically updated (last update 18 March 2014) and is available at http://bfw.ac.at/200/2043.html.

In the absence of taxon-specific complete chloroplast genome data, it is possible to mine the wealth of genomic data available in international databases such as GenBank (National Center for Biotechnology Information), EMBL-Bank (European Molecular Biology Laboratory), and DDBJ (DNA Data Bank of Japan). Primer pairs for 130 regions of the chloroplast genome were evaluated relative to representatives of 12 genera, spanning the diversity of flowering plants. Exon regions were avoided because they generally evolve more slowly than intron and intergenic spacer regions. The primers of Shaw et al. (2005, 2007), Scarcelli et al. (2011), and Dong et al. (2012), as well as the primers provided here, were evaluated. Many of the Shaw et al. (2005, 2007) and Scarcelli et al. (2011) primers are degenerate, improving the breadth of taxa they can be used on, but reducing their efficiency during the amplification process. The Dong et al. (2012) primers are primarily used for barcoding, thus amplify a diversity of taxa, but may not target the most quickly evolving regions of the genome. The likelihood of amplification success was estimated based upon the number and position of mismatches between the primer and the target sequence. These data were then evaluated in the context of Shaw et al. (2014) to provide generalizations, by taxonomic group, for primer utility in conjunction with sequence variability.

Finally, a small number of plant species have sequences available for multiple accessions or different subspecific taxa including *Fragaria vesca* L. (Rosaceae, $N = 2$), *Gossypium herbaceum* L. (Malvaceae, $N = 2$), *Olea europaea* L. (Oleaceae, $N = 4$), and *Oryza sativa* L. (Poaceae, $N = 3$). Shaw et al. (2014) specifically excluded species pairs with very low and very high levels of sequence divergence. Very high levels of divergence made alignment difficult, and very low levels provide too few characters for reasonable comparison across all flowering plants. Here I compare the variation at the subspecific level to that of higher-level relationships to determine if the same regions are useful at multiple taxonomic levels.

METHODS

***Primers designed here*—**Sixteen chloroplast genomes, representing a diversity of flowering plants, were downloaded from GenBank (see Appendix 1). Homologous gene sequences were aligned in Se-Al version 2.0a11 (Rambaut, 1996). Primers were designed based on simultaneous viewing of the Se-Al file and an Oligo 4.02 (Rychlik, 2002) file, using a single sequence from the pool. Primers were anchored in coding regions and were designed to have a minimum number of hair-pins and primer-primer interactions, annealing temperatures between 50°C and 64°C, and a 3′ GC clamp if possible, targeting regions 400–1800 bp in length. Primer details are provided in Table 1, and are provided in the order of appearance in the tobacco genome (*Nicotiana tabacum* L. [GenBank Z00044.1]). The tobacco genome was the genome of choice for describing the location of primers prior to the recent accumulation of genomic data. A total of three different *trnS* primers were designed, corresponding to the three *trnS* genes encoded by the chloroplast genome (*trnS*-GCU, *trnS*-UGA, and *trnS*-GGA). Gene order is highly conserved on the chloroplast genome of flowering plants, but does vary and can be highly informative, for example, as in the 22-kb inversion in almost all Asteraceae (Jansen and Palmer, 1987a, 1987b) and the 78-kb inversion in Fabaceae subtribe Phaseolinae (Bruneau et al., 1990). Some primer combinations are not useful in particular groups of plants due to structural rearrangements. In some cases, the downloaded genomes differ in the identification of specific genes.

***Primer utility*—**The chloroplast genomes for species of eight genera (*Acorus* L., *Amborella* Baill., *Canna* L., *Ceratophyllum* L., *Cymbidium* Sw., *Helianthus* L., *Magnolia* L., and *Nelumbo* Adans.) and for subspecies of *F. vesca*, *G. herbaceum*, *O. europaea*, and *O. sativa* were compared to 130 primer pairs published by Shaw et al. (2005, 2007), Scarcelli et al. (2011), Dong et al. (2012), and those designed here. Complete chloroplast genome sequences were downloaded from GenBank (accession numbers, taxonomic identity, and original publication information provided in Appendix 3) and aligned manually in Sequencher (Gene Codes Corporation, Ann Arbor, Michigan, USA). A separate file containing the primer sequences was imported and automatically assembled using the settings "dirty data" and 100% sequence similarity with a minimum overlap of 16 bp. Additional rounds of alignment were conducted with successively lower levels of sequence similarity. Primers that failed to align automatically, or that aligned incorrectly, were realigned manually whenever possible (guided by the GenBank annotations). Alignment of the two *Gossypium* sequences required inversion of a large region of one taxon (arbitrarily selected as *G. herbaceum* subsp. *africanum* (G. Watt) Vollesen) approximately corresponding to bases 115,132–135,355 in the final alignment. The *Oryza* alignment includes *O. nivara* Sharma & Shastry because it is a potential progenitor of *O. sativa* (Li et al., 2006; but see Huang et al., 2012 for an alternative view point).

As mentioned above, degenerate primers provide broader utility, but reduced amplification efficiency. If a mismatch was detected in the last five bases at the 3′ end of the primer, the mismatch was inferred to be fatal (IDT, 2009). If more than three mismatches were detected within any given primer, amplification was inferred to be unsuccessful. These criteria are arbitrary but have worked for me personally and are probably more strict than necessary.

***Sequence variability within species*—**The sequences of *F. vesca*, *G. herbaceum*, *O. europaea*, and *O. sativa* were examined manually to assess the variation of the 130 regions. Length of the inferred amplicon was noted along with the number of mismatched bases (aka inferred substitutions; excluding primer regions), the number of insertion/deletion (indel) events, and the number of inversions. These data provided an estimate of the utility of the regions for inferring phylogeny among closely related subspecies, and potential for application to phylogeographic studies. Shaw et al. (2014) specifically avoided these types of comparisons due to the very small number of parsimony informative characters. Sequence diversity was estimated using three criteria calculated as: (1) [(number of substitutions*2)+(number of indels)+(number of inversions)]/amplicon length, (2) number of substitutions+indels+inversions, and (3) sequence diversity (number of substitutions/sequence length). The first criterion (criterion 1) is a weighted rank, and includes information on the number of inferred substitutions (weighted twice as heavily as the other two components), indels, and inversions. Substitutions were weighted more heavily because chloroplast indels may be more homoplasious (Kelchner and Clark, 1997), especially among closely related taxa. Inversions are often low in homoplasy (Graham et al., 2000) and thus could be weighted more heavily, but are relatively rare so weighting was not employed. The 10 most variable regions for each species were identified, as measured under each criterion. Frequency of any specific "top 10" primer pair was summed across the four species.

RESULTS

***Primers designed here*—**The 72 primers targeted noncoding regions of the chloroplast genome with amplicon sizes of 500–1800 bp. Degenerate primers were avoided because they were assumed to decrease priming efficiency, as were mismatches within the last five bases at the 3′ end of the primer. Only two primers required degenerate bases: one primer with two degenerate bases and another primer with one degenerate base. None of these degeneracies were located within the last five bases. In

Downloaded From: https://complete.bioone.org/journals/Applications-in-Plant-Sciences on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

TABLE 1. Region, primer name, primer sequence, amplicon position, and amplicon length for plastid noncoding regions relative to the *Nicotiana tabacum* L. (GenBank Z00044.1) genome.

Region	Primer name	T_m (°C)[a]	Primer sequence	Amplicon position	Amplicon length (bp)
trnQ(UUG)–*psb*K IGS	trnQ-IGSR	62.7	ACCCGTTGCCTTACCGCTTGG	7457–8018	562
	psbK-IGSR	50.9	ATCGAAAACTTGCAGCAGCTTG		
psbK–*trnS*(GCU) IGS	psbK-IGSF	47.9	CCAATCGTAGATGTTATGCC	7937–8719	783
	trnS_GCU-IGSF	56.1	GGAGAGATGGCTGAGTGGA		
trnG(UCC)–*atpA* IGS	trnG_UCC-IGSF	56.3	CCTTCCAAGCTAACGATGCG	10,219–10,796	577
	atpA-IGSF	50.3	TGGACAGGTGAAGAAATTTC		
atpF intron	atpF-E2R	47.3	CTCTGTTTTCGATTATCTAATAAAT	12,582–13,372	791
	atpF-E1F	48.1	AGCAACAAATCCAATAAATCT		
atpF–*atpH* IGS	atpF-E1R	46.5	TAGATTTATTGGATTTGTTGC	13,352–13,927	575
	atpH-IGSF	48.5	CTTTTATGGAAGCTTTAACAATTTA		
atpH–*atpI* IGS	atpH-IGSR	56.9	CCAGCAGCAATAACGGAAGC	14,059–15,400	1341
	atpI-IGSF	48.2	GTTGTTGTTCTTGTTTCTTTAG		
rpoC1 intron	rpoC1-intR	49.9	AAGTGGGATGCTGTATTTC	23,004–23,976	973
	rpoC1-intF	49.2	ACGAAGGTATCAAATGGG		
trnS(UGA)–*psbZ* IGS	trnS_UGA-IGSR	55.0	ATCAACCACTCGGCCATC	37,209–37,620	412
	psbZ-IGS	45.6	AATAGCCAATTGAAAAGC		
psaA–*ycf3* IGS	psaA-IGSR	50.2	CGGCGAACGAATAATCAT	43,469–44,295	827
	ycf3-E3F	48.4	CCCGGTAATTATATTGAAGC		
ycf3 intron 2	ycf3-E3R	54.5	ATCTCCCTGTCGAATGGC	44,362–45,193	832
	ycf3-E2F	53.2	GGCCGTGATCTGTCATTAC		
ycf3 intron 1	ycf3-E2R	50.0	TTCCGCGTAATTTCCTTC	45,370–46,163	794
	ycf3-E1F	48.1	CATTTACCTATTACAGAGATGG		
ycf3–*trnS*(GGA) IGS	ycf3-E1R	45.5	ACAATTGAAAAGGTCTTATC	46,214–47,174	961
	trnS_GGA-IGSR	47.9	CAAAAGCCTACATAGCAG		
rpS4-*trnT*(UGU)	rpS4-IGSR1	56.2	TCCTCGGTAACGCGACAT	48,065–48,570	506 max.
	rpS4-IGSR2	45.9	GGCTTTTTATTAGTTAGTCC		
	trnT_UGU-IGSF1	53.0	AGGTTAGAGCATCGCATTTG		
	trnT_UGU-IGSF2	47.9	GAGCATCGCATTTGTAAT		
trnF(GAA)–*ndhJ* IGS	trnF-IGSF	56.4	ATCCTCGTGTCACCAGTTCAAA	50,277–51,024	747
	ndhJ-IGSF	49.3	RCCCCTAATTTYTATGAAATACA		
ndhC–*trnV*(UAC) IGS	ndhC-IGSR	52.9	ATCATATTCGTGAAGCAGAAACAT	52,644–53,776	1132
	trnV_UAC-E2F	58.3	GGTTCGAGTCCGTATAGCCCT		
trnV(UAC) intron	trnV_UAC-E2R	57.1	GGGCTATACGGACTCGAACC	53,757–54,380	624
	trnV_UAC-E1F	52.8	GTAGAGCACCTCGTTTACAC		
trnV(UAC)–*atpE* IGS	trnV_UAC-E1R	52.8	GTGTAAACGAGGTGCTCTAC	54,361–55,032	672
	atpE-IGSF	56.6	AGTGACATTGATCCRCAAGAAGC		
atpB–*rbcL* IGS	atpB-IGSR	48.4	AAGTAGTAGGATTGATTCTCAT	56,756–57,615	859
	rbcL-IGSR	53.9	AGTCTCTGTTTGTGGTGACAT		
rbcL–*accD* IGS	rbcL-IGSF	58.5	GCTGCTGCTTGTGAGGTATGG	58,960–59,865	905
	accD-IGSR	51.1	AATTGAACCCACATTTTTCCATA		
accD–*psaI* IGS	accD-IGSF	48.2	GGTAAAAGAGTAATTGAACAAAC	61,143–62,161	1018
	psaI-IGSR	49.7	ATAAAGAAGCCATTGCAATTG		
psaI–*ycf4* IGS	psaI-IGSF	51.8	CCTAGTCTTTCCGGCAAT	62,127–62,682	556
	ycf4-IGSR	49.5	CCCCGTTATAAGTTCTATCC		
ycf4–*ycf10* IGS	ycf4-IGSF	47.0	ATTAGCCTATTTCTTGCG	63,153–63,541	389
	ycf10-IGSR	51.9	GCCCAGTATTCCACCAA		
petA–*psbJ* IGS	petA-IGSF	50.8	GAAACAGTTTGAGAAGGTTCA	65,255–66,388	1133
	psbJ-IGSF	55.8	ATTCCGCATTGGGCTCATC		
petL–*psaJ* IGS	petL-IGSF	48.4	TCTATTAGCGGCTTTAACTATA	68,322–69,671	1350
	psaJ-IGSR	52.4	GCATCCGGGAATAAACGA		
psaJ–*rpL20* IGS	psaJ-IGSF	46.5	ATGCGAGATCTAAAAACATA	69,565–71,404	1840
	rpL20-IGSF	46.6	CAGAATTAAACGGGGATATA		
rpL20–*rpS12* IGS	rpL20-IGSR	51.3	CGTCTCCGAGCTATATATCC	71,372–72,319	947
	rpS12-IGSF	47.3	CAACTTATTAGAAACACAAGAC		
clpP intron 2	clpP-E3R	51.6	TTGCCTGTTCTTTGTACATAAAC	72,573–73,466	893
	clpP-E2F	50.9	GCTATTTATGACGCTATGCAA		
clpP intron 1	clpP-E2R	50.9	TTGCATAGCGTCATAAATAGC	73,446–74,451	1005
	clpP-E1F	54.9	TTGGGTTGACATATAGTGCGAC		
clpP–*psbB* IGS	clpPE1-IGSR	52.2	AGGGACTTTTGGAACACC	74,481–74,970	490
	psbB-IGSR	51.5	ATACCAAGGCAAACCCAT		
psbH–*petB* IGS	psbH-IGSF	48.5	AACTACTCCTTTGATGGG	77,214–78,377	1163
	petB-E2R	44.1	TAGTAAAAAGTCATAGCAAA		
petB–*petD* IGS	petBE2-IGSF	50.8	ATGCACTTTCCAATGATACG	78,805–79,760	956
	petD-E2R	59.8	CCCGAGGGAACCGGACAT		
rpS3–*rpS19* IGS	rpS3-IGSR	50.5	CAGTCTGAAACCAAGTGG	85,863–86,504	642
	rpS19-IGSF	45.9	TTTATATAACGGATAGTATGGT		
ccsA-*ndhD* IGS	ccsA-IGSF	45.5	ATGATATTTTCAACCTTAGA	116,344–117,614	1271
	ndhD-IGSF	43.6	CCGTAATAGGTATTGGTAT		

Downloaded From: https://complete.bioone.org/journals/Applications-in-Plant-Sciences on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Table 1. Continued.

Region	Primer name	T_m (°C)[a]	Primer sequence	Amplicon position	Amplicon length (bp)
psaC–ndhE IGS	psaC-IGSR	44.9	TCCTATACACGTATCATAAA	119,351–119,713	363
	ndhE-IGSF	42.4	TTCATCAATTTATCGTAAC		
ndhE–ndhI IGS	ndhE-IGSR	45.6	GAAAATAAATAGGCACTCAA	119,912–121,251	1340
	ndhI-IGSF	46.9	CAATGACCGAAGAATATGA		
rpS15–ycf1 IGS	rpS15-IGSR	47.7	GCAATTCTAAATGTGAAGTAAG	125,374–126,001	628
	ycf1-IGSR	45.6	ATTATCGATTAGAAGATTTAGC		

[a]Melting temperature (T_m) based on 50 mM NaCl solution.

contrast, 17 of the Scarcelli et al. (2011) primers have at least one degenerate base in the last five bases at the 3′ end of the primer, and so are assumed to fail for at least some taxa.

***Primer evaluation*—**Three of the four sets of primers examined here were equally likely to amplify target chloroplast regions (81–85% should work; see Table 2). The Dong et al. (2012) primers were least likely to work based on the 12 species examined here (65% on average) and were particularly poorly matched to the *Oryza* genome (29% amplification success predicted), and only moderately suited for *Amborella* (52%), *Cymbidium* (52%), and *Helianthus* (57%). However, the Dong et al. (2012) primer pair *trnH-psbA* was not expected to work on any of the target species, possibly due, in part, to an extra "A" near the 3′ end of the published sequence for the *trnH* primer. The primers designed here were poorly matched to three of the four monocots (*Cymbidium*, *Oryza*, and *Canna*; 61%, 64%, and 67%, respectively), despite being a good match for *Acorus* (81%). Scarcelli et al. (2011) primers were designed with monocots in mind and did an exceptional job matching the monocot genomes examined here, with amplification success ranging from 82–97%. They were almost equally good for the dicots examined here, with amplification success of 72–93%. The Shaw et al. (2005, 2007) primers were useful across the angiosperm phylogeny, with all anticipated amplification success percentages above 78%.

On average, the Shaw et al. (2005, 2007) and Scarcelli et al. (2011) primers are more degenerate, yet they were only slightly more likely to amplify the target sequences than the nondegenerate primers designed here, at least for nonmonocot taxa. With so many different primers available, most regions could be amplified in almost all target taxa provided an appropriate primer pair was selected. Indeed, many primer pairs should work in all 12 species examined here. Details of the inferred priming success are provided in Appendix S1, and species-specific notes on primer/sequence mismatches are provided in Appendix S2.

***Primer utility* × *sequence variability*—**Shaw et al. (2014) conveniently summarized sequence variability across the chloroplast genome including the identification of the 13 fastest-evolving regions for six taxonomic groups (magnoliids, monocots, eurosids I, eurosids II, euasterids I, and euasterids II). Summing across these major groups, 28 different regions were identified as the most variable. Primers to amplify those 28 regions are detailed in Table 3, along with the Shaw et al. (2014) rank for each region (in bold typeface above each primer region), for each taxon examined here. Multiple primer pairs are available for each of the 28 regions except the *trnT-trnL* (Shaw et al., 2005 only), *ycf4-ycf10* (or *cemA*; current study only), and *ndhD-psaC* (none of the publications examined). The *ndhD-psaC* region was ranked 10th fastest for eurosids I, but as there are no primers to be evaluated this region will not be discussed further. Primers are available for each of the remaining 27 regions.

Among the basal dicot grade (*Amborella* and *Magnolia*), successful primers are available for all 27 regions. Primer selection is more challenging for *Amborella* than for *Magnolia*. The top ranked region was the *rpl32-trnL* intergenic spacer (IGS). Shaw et al. (2007) primers will work for both taxa; Dong et al. (2012) primers will not. In contrast, *rps16-trnQ*, the second highest ranked region, has three sets of primers available (Shaw et al., 2007; Scarcelli et al., 2011; and Dong et al., 2012), all of which should work.

Among the monocots sampled (*Acorus*, *Cymbidium*, *Oryza*, and *Canna*), *Acorus* was the least difficult sequence to match and *Oryza* the most difficult. Structural rearrangements are the primary reason for failure to amplify across all available primers (e.g., *rbcL-accD* in *Oryza* and *petA-psbJ* in *Cymbidium*). One region cannot be amplified in *Acorus*—the *accD-psaI* IGS, despite the availability of four different primer pairs. In all, four regions cannot be amplified in *Cymbidium* with the primers studied here: *petN-psbM*, *psbM-trnD*, *atpB-rbcL*, and *petA-psbJ*. The *ndhA* region can be amplified in only some species of *Cymbidium* due to fatal substitutions in some species for all three primer pairs evaluated here. In *Oryza*, the *trnS*[GCU]-*trnG*[GCC], *trnT-psbD*, *rbcL-accD*, *accD-psaI*, and *rps15-ycf1* cannot be amplified using any primer pair. In *Canna*, *ndhF-rpl32* will not amplify with either of the available primer pairs. Unfortunately, according to Shaw et al. (2014), *ndhF-rpl32* is the most variable and *psbM-trnD* is the third most variable region for monocots.

Basal eudicots were not evaluated by Shaw et al. (2014) in detail, so direct comparisons cannot be made here. Fortunately, at least one primer pair was successful for each of the 27 fastest-evolving regions, with the exception of the *ycf4-ycf10* region. The only available primers for this region were designed here, and they will not work for *Ceratophyllum*. In general, *Ceratophyllum* was more difficult to match than was *Nelumbo*.

Shaw et al. (2014) detailed variability of higher eudicots for four major groups: eurosids I, eurosids II, euasterids I, and euasterids II. Only a single species representing each group was included here. *Fragaria* (eurosids I) could not be amplified for a single region, the *ycf4-ycf10* IGS. According to Shaw et al. (2014), the fastest region for this clade was the *ndhA* intron. Both the Shaw et al. (2007) and Scarcelli et al. (2011) primers should work, but the Dong et al. (2012) primers will not. The second fastest region was the *trnS*[GCU]-*trnG*[GCC], which should amplify with any of the primer pairs (Shaw et al., 2005; Scarcelli et al., 2011; or Dong et al., 2012).

The sole representative of eurosids II and euasterids I (*Gossypium* and *Olea*, respectively) could successfully be amplified by at least one pair of primers studied here. The fastest region

Downloaded From: https://complete.bioone.org/journals/Applications-in-Plant-Sciences on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

TABLE 2. Summary of amplification success probability for 130 pairs of chloroplast primers.

			Basal dicot grade/Magnoliids		Monocots			Basal eudicot grade		Eurosids I	Eurosids II	Euasterids I	Euasterids II
Publication[a]	No. of regions	Average % ampl.	*Amborella*	*Magnolia*	*Acorus*	*Cymbidium*	*Oryza*	*Canna*	*Ceratophyllum*	*Nelumbo*	*Fragaria*	*Gossypium*	*Olea*	*Helianthus*
Dong	21	65	11 (52%)	16 (76%)	14 (67%)	11 (52%)	6 (29%)	15 (71%)	15 (71%)	17 (81%)	16 (76%)	14 (67%)	17 (81%)	12 (57%)
Current study	36	81	31 (86%)	32 (89%)	29 (81%)	22 (61%)	23 (64%)	24 (67%)	32 (89%)	32 (89%)	28 (78%)	33 (92%)	31 (86%)	32 (89%)
Scarcelli	99	83	71 (72%)	92 (93%)	96 (97%)	92 (93%)	81 (82%)	87 (88%)	71 (72%)	88 (89%)	73 (74%)	80 (81%)	79 (80%)	75 (76%)
Shaw	33	85	27 (82%)	31 (94%)	29 (88%)	26 (79%)	26 (79%)	29 (88%)	28 (85%)	28 (85%)	27 (82%)	27 (82%)	29 (88%)	28 (85%)

[a] Dong et al., 2011; Scarcelli et al., 2011; Shaw et al., 2005, 2007.

for eurosids II was the *ndhF-rpl32* IGS. The Shaw et al. (2007) primer pair should work, but the Scarcelli et al. (2011) primer pair likely will not. The second most variable region was the *psbZ-trnG* IGS. For this region, both the Scarcelli et al. (2011) and Dong et al. (2012) primers should work, but the Shaw et al. (2005; as *trnfM-trnS*) primers will not. In euasterids I, the fastest region was the *rps16-trnQ* IGS. For *Olea*, the Shaw et al. (2007) and Scarcelli et al. (2011) primers should work, but not so the Dong et al. (2012) primers. The next-fastest region was the *rpl32-trnL* IGS. Both the Shaw et al. (2007) and Dong et al. (2012) primers should work.

Primer failure in *Helianthus* (euasterids II) was primarily due to structural rearrangements (e.g., *trnS*[GCU]-*trnG*[GCC], *rpoB-trnC*, *trnE-trnT*, *rbcL-accD*). *rpl32-trnL* IGS was the fastest region according to Shaw et al. (2014), and either the Shaw et al. (2007) or Dong et al. (2012) primers should successfully amplify this region. The adjacent *ndhF-rpl32* IGS was the second most variable region. Both the Shaw et al. (2007) or the Scarcelli et al. (2011) primers should work.

Subspecific sequence variability—Intraspecific sequence variation was evaluated in four species: *F. vesca*, *G. herbaceum*, *O. europaea*, and *O. sativa*. This represents a tiny fraction of angiosperm diversity, but is the first analysis of subspecific diversity across the entire chloroplast genome for multiple species, in the context of available primer resources. Appendix S3 identifies the fastest-evolving regions among the four species, under three different criteria. On average, only five inversions per chloroplast genome were detected here and the distribution across species was very different. *Gossypium* and *Oryza* each had 10 inversions, *Fragaria* none, and *Olea* only one. Details of subspecific comparisons for all regions are provided in Appendix S2.

No single genic region was identified as the top 10 fastest for all four species. Pooling data across all three criteria, the most frequently identified genic region was the *psbZ-trnfM* IGS with eight occurrences out of a maximum of 12 possible, followed by the *trnS*(GCU)-*trnG*(GCC) IGS, with six occurrences, *rps16-trnQ* IGS and *trnT*(GGU)-*psbD* IGS each with five, and *rps12-psbB* IGS and *rps4-trnT*(UGU) IGS each with four occurrences. Data for individual species have limited general application, but are provided below.

Oryza sativa, the only monocot in this comparison, showed highest variation, based on rank, for *clpP-psbB* (0.0195, 924 bp), *atpB-rbcL* (0.0168, 1070 bp), and *psbM-trnD*(GUC) (0.0150, 523 bp). Two of the same regions were identified as fastest under criterion 2, *atpB-rbcL* (12 characters, 1070 bp) and *clpP-psbB* (11 characters, 924 bp), plus *rbcL-accD* (13 characters, 1824 bp). Sequence divergence was highest in and around the *clpP* region including what would be the *clpP* intron 2 (1.9455%, 257 bp), *clpP* intron 1 (1.0050%, 199 bp), and *clpP-psbB* (0.7576%, 924 bp). In contrast, the three fastest regions per Shaw et al. (2014) for monocots were *ndhF-rpl32* (rank 1), *ndhC-trnV* (rank 2), and *psbM-trnD* (rank 3).

The highest variation for *Fragaria* under criterion 1 was for *trnW*(CCA)-*psaJ* (0.0101, 789 bp), *trnT*(GGU)-*psbD* (0.0098, 1527 bp), and *trnP*(UGG)-*rps18* (0.0090, 1563 bp). Under criterion 2: *trnT*(GGU)-*psbD* (eight characters; 1527 bp), *trnP*(UGG)-*rps18* (eight characters, 1563 bp), and *petN-trnD* (seven characters, 2504 bp). Under criterion 3, the top three regions were *trnT*(GGU)-*psbD* (0.4584%, 1527 bp), *psbB-psbH* (0.4451%, 674 bp), and *rps4-trnT*(UGU) (0.4435%, 451 bp). Shaw et al. (2014) eurosids I top three regions were *ndhA* intron

Downloaded From: https://complete.bioone.org/journals/Applications-in-Plant-Sciences on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

TABLE 3. Amplification success prediction for the 28 fastest Shaw et al. (2014) regions.[a]

Approx. *Nicotiana* order	Genomic region	Publication[b]	Basal dicot grade/Magnoliids		Monocots				Basal eudicot grade		Eurosids I	Eurosids II	Euasterids I	Euasterids II	Average
			Amborella	*Magnolia*	*Acorus*	*Cymbidium*	*Oryza*	*Canna*	*Ceratophyllum*	*Nelumbo*	*Fragaria*	*Gossypium*	*Olea*	*Helianthus*	
1	*trnH-psbA* IGS												**8**[c]		
	trnH-psbA IGS	Dong et al.	NO**	NO**	NO	NO**	NO	NO	NO	NO**	NO	NO	NO	NO**	0%
	trnH-psbA IGS	Scarcelli et al.	YES	YES	YES	YES	YES	NO	YES	NO	YES	YES	YES	YES	83%
	trnH-psbA IGS	Shaw et al.	YES	YES	YES	NO	YES	YES	YES	YES	YES	YES	YES	YES	92%
5	*matK* exon		**12**[c]								**6**[c]	**12**[c]			
	trnK (including *matK*)	Dong et al.	YES	YES	YES	NO	YES	YES	YES	YES	YES	YES	YES	YES	92%
	matK exon	Scarcelli et al.	YES	YES	YES	YES*	YES	YES	YES	YES	YES	YES	YES	YES	100%
7	*trnK-rps16* IGS		**13**[c]		**5**[c]						**13**[c]		**7**[c]	**12**[c]	
	trnK-rps16	Scarcelli et al.	YES	YES	YES	YES	YES	YES	YES	YES	NO	YES	YES	YES*	92%
	trnK-3′rpS16	Shaw et al.	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	100%
8	*rps16* intron				**4**[c]						**3**[c]			**5**[c]	
	rps16 intron	Scarcelli et al.	YES	YES	YES	YES	YES	YES	NO	YES	NO	YES	YES	YES	83%
	rpS16 intron	Shaw et al.	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	100%
9	*rps16-trnQ* IGS		**2**[c]								**11**[c]		**1**[c]	**13**[c]	
	rps16-trnQ	Dong et al.	YES	YES	YES	YES	NO	NO	YES	YES	NO	NO	NO	YES	58%
	rps16-trnQ	Scarcelli et al.	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	100%
	5′rpS16-trnQ	Shaw et al.	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	100%
12	*trnS-trnG* IGS				**11**[c]						**2**[c]		**12**[c]		
	trnS-trnG (and intron)	Dong et al.	NO	YES	YES	YES	NO	YES	YES	YES	YES	YES	YES	NO	75%
	trnS-trnG	Scarcelli et al.	NO	YES	YES	YES	NO	YES	YES	YES	YES	YES	YES	NO	75%
	trnS-trnG	Shaw et al.	YES	YES	YES	YES	NO	YES	YES	YES	YES	YES	YES	NO	83%
16	*atpF* intron										**5**[c]				
	atpF intron	Prince (here)	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	100%
	atpF intron/exon	Scarcelli et al.	NO	YES	YES	NO	YES	YES	NO	YES	NO	YES	YES	YES	67%
18	*atpH-atp* IGS		**9**[c]								**12**[c]		**4**[c]		
	atpH-atpI	Dong et al.	YES	YES	YES	YES	YES	YES	YES	YES	YES	NO	YES	YES	92%
	atpH-atpI	Prince (here)	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	100%
	atpH-atpI	Scarcelli et al.	YES	YES	YES	YES	YES	YES	YES	YES	NO	YES	YES	YES	92%
	atpH-atpI	Shaw et al.	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	100%
26	*rpoB-trnC* IGS				**8**[c]							**10**[c]	**11**[c]	**7**[c]	
	rpoB-trnC	Dong et al.	YES	YES	YES	NO	NO	NO	YES	YES	YES	YES	YES	NO	67%
	rpoB-trnC	Scarcelli et al.	NO	YES	YES	YES	YES	YES	NO	YES	YES	YES	YES	NO	75%
	rpoB-trnC	Shaw et al.	YES	YES	YES	YES	YES	YES	YES	YES	YES	NO	YES	NO	83%
29–31	*petN-psbM* IGS												**6**[c]	**10**[c]	
	petN-trnD	Scarcelli et al.	YES	YES	YES	NO	YES	YES	YES	YES	YES	YES	YES	NO	83%
	petN-psbM	Dong et al.	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	17%
	ycf6-psbM	Shaw et al.	YES	YES	YES	NO	YES	NO	YES	YES	YES	YES	NO	YES	75%
32	*psbM-trnD* IGS		**8**[c]		**3**[c]							**9**[c]			
	psbM-trnD	Dong et al.	YES	YES	YES	NO	NO	YES	YES	YES	YES	YES	YES	YES	83%
	psbM-trnD	Shaw et al.	NO	NO	YES	NO	YES	YES	YES	YES	YES	YES	YES	NO	67%
33	*trnE-trnT* IGS											**8**[c]		**6**[c]	
	trnD-trnT	Scarcelli et al.	YES	YES	YES	YES	NO	YES	YES	YES	YES	YES	YES	NO	83%
	trnD-trnT	Shaw et al.	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	NO	92%
34	*trnT-psbD* IGS		**4**[c]								**8**[c]	**4**[c]		**8**[c]	
	trnT-psbD	Dong et al.	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	NO	67%
	trnT-psbD	Scarcelli et al.	NO	YES	YES	YES	NO	YES	YES	YES	YES	YES	YES	YES	83%
	trnT-psbD	Shaw et al.	YES	YES	YES	YES	NO	YES	YES	YES	YES	YES	YES	YES	92%
38–41	*psbZ-trnG* IGS										**7**[c]	**2**[c]			
	trnS-trnG	Dong et al.	YES	YES	YES	YES	NO	YES	YES	YES	YES	YES	NO	YES	83%
	trnS-trnfM	Shaw et al.	YES	YES	NO	YES	YES	YES	YES	NO	YES	NO	NO	YES	67%

http://www.bioone.org/loi/apps

Downloaded From: https://complete.bioone.org/journals/Applications-in-Plant-Sciences on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

TABLE 3. Continued.

Approx. *Nicotiana* order	Genomic region	Publication[b]	Basal dicot grade/Magnoliids		Monocots				Basal eudicot grade		Eurosids I	Eurosids II	Euasterids I	Euasterids II	Average
			Amborella	*Magnolia*	*Acorus*	*Cymbidium*	*Oryza*	*Canna*	*Ceratophyllum*	*Nelumbo*	*Fragaria*	*Gossypium*	*Olea*	*Helianthus*	
	psbZ-trnfM	Scarcelli et al.	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	100%
50	*trnT-trnL* IGS		**11**[c]		**9**[c]							**3**[c]			
	trnT-trnL	Shaw et al.	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	100%
55	*ndhC-trnV* IGS		**5**[c]		**2**[c]								**3**[c]	**3**[c]	
	ndhC-trnV	Dong et al.	YES	YES	YES	YES*	YES	YES	YES	YES	YES	YES	YES	YES	100%
	ndhC-trnV	Prince (here)	YES	YES	YES	YES	YES	YES	YES	YES	NO	YES	YES	YES	92%
	ndhC-trnV	Scarcelli et al.	YES	YES	YES	YES*	YES	YES	YES	YES	YES	YES	YES	YES	100%
	ndhC-trnV	Shaw et al.	YES	YES	YES	YES	YES	YES	YES	YES	NO	YES	YES	YES	92%
60	*atpB-rbcL* IGS										**9**[c]				
	atpB-rbcL	Prince (here)	YES	YES	YES	NO	YES	YES	YES	YES	NO	NO	YES	YES	75%
	atpB-rbcL	Scarcelli et al.	NO	YES	YES	NO	YES	YES	YES	YES*	NO	YES	YES	NO	67%
62	*rbcL-accD* IGS				**12**[c]								**13**[c]		
	rbcL-accD	Dong et al.	NO	YES	YES	YES	NO	YES	YES	YES	YES	YES	YES	NO	75%
	rbcL-accD	Prince (here)	YES	YES	NO	YES	NO	NO	YES	NO	NO	YES	NO	NO	42%
	rbcL-accD	Scarcelli et al.	NO	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	8%
64	*accD-psaI* IGS		**10**[c]		**10**[c]										
	accD-psaI	Dong et al.	NO	YES	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	67%
	accD-psaI	Prince (here)	NO	YES	NO	YES	NO	YES	YES	YES	YES	YES	YES	NO	67%
	accD-psaI	Scarcelli et al.	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	YES	YES	58%
	accD-psaI	Shaw et al.	YES	YES	NO	YES	NO	NO	YES	YES	YES	YES	YES	YES	75%
67	*ycf4-cemA* (*ycf10*) IGS											**11**[c]			
	ycf4-ycf10	Prince (here)	YES	YES	YES	YES	YES	NO	NO	YES	NO	YES	YES	YES	75%
70	*petA-psbJ* IGS		**6**[c]		**6**[c]							**5**[c]	**5**[c]		
	petA-psbJ	Dong et al.	YES	YES	YES	NO	NO	YES	YES	YES	YES	YES	YES	NO	75%
	petA-psbJ	Prince (here)	YES	YES	YES	NO	YES	NO	YES	YES	YES	NO	NO	YES	67%
	petA-psbJ	Shaw et al.	YES	YES	YES	NO	YES	YES	YES	YES	NO	NO	YES	YES	75%
72	*psbE-petL* IGS		**7**[c]		**7**[c]						**4**[c]	**13**[c]		**9**[c]	
	psbE-petL	Dong et al.	NO	NO	NO	YES*	NO	YES	YES	NO	YES	NO	YES	YES	50%
	psbE-petL	Shaw et al.	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	100%
76, 77	*psaJ-rpl33* IGS				**13**[c]										
	trnP-rps18	Scarcelli et al.	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	NO	92%
	psaJ-rpL20	Prince (here)	NO	YES	NO	NO	NO	YES	YES	YES	NO	YES	YES	YES	58%
116	*ndhF-rpl32* IGS		**3**[c]		**1**[c]							**1**[c]	**9**[c]	**2**[c]	
	ndhF-rpl32	Scarcelli et al.	YES	YES	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	83%
	ndhF-rpl32	Shaw et al.	NO	YES	YES	YES	NO	NO	NO	YES	YES	YES	YES	YES	67%
118	*rpl32-trnL* IGS		**1**[c]									**6**[c]	**2**[c]	**1**[c]	
	rpL32-trnL	Dong et al.	NO	YES	YES	NO	YES	YES	YES	YES	YES	YES	YES	YES	83%
	rpL32-trnL	Shaw et al.	YES	YES	YES	YES	YES	YES	YES	YES	NO	YES	YES	YES	92%
121.5	*ndhD-psaC* IGS										**10**[c]				
127	*ndhA* intron										**1**[c]		**10**[c]	**11**[c]	
	ndhA intron	Dong et al.	NO	NO	NO	YES*	YES	NO	NO	NO	NO	YES	NO	NO	25%
	ndhA intron	Scarcelli et al.	YES	YES	YES	YES*	YES	YES	YES	YES	YES	YES	YES	YES	100%
	ndhA intron	Shaw et al.	YES	YES	YES	YES*	NO	YES	YES	YES	YES	YES	YES	YES	92%
129	*rps15-ycf1* IGS											**7**[c]		**4**[c]	
	rpS15-ycf1	Prince (here)	YES	YES	YES	NO	NO	YES	YES	YES	YES	YES	YES	YES	83%
	rps15-ycf1	Scarcelli et al.	YES	NO	YES	YES	NO	YES	YES	NO	YES	YES	NO	YES	67%

[a] YES* = will not work for at least one species in the genus; NO** = will work if psbA primer is synthesized with one fewer A at the 3′ end.

[b] Shaw et al., 2005, 2007; Scarcelli et al., 2011; Dong et al., 2012.

[c] Shaw et al. (2014) rank for the region within the specified taxonomic group.

http://www.bioone.org/loi/apps

Downloaded From: https://complete.bioone.org/journals/Applications-in-Plant-Sciences on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

(rank 1), *trnS*(GCU)-*trnG*(GCC) (rank 2), and *rps16* intron (rank 3).

In *Gossypium*, the most informative regions under criterion 1 were *psbZ-trnfM*(CAU) (0.0534, 1179 bp), *trnH*(GUG)-*psbA* (0.0444, 496 bp), and *rps4-trnT*(UGU) (0.0425, 635 bp). Criterion 2 fastest regions were *trnS*(UGA)-*trnG*(GCC) with 39 variable characters over 1673 bp, followed by *psbZ-trnfM*(CAU) with 37 characters for 1179 bp, and *trnT*(UGU)-*trnL*(UAA) with 33 characters over 1470 bp. Sequence divergence (criterion 3) was highest for *psbZ-trnfM*(CAU) (2.2053%, 1179 bp), then *trnS*(UGA)-*trnG*(GCC) (1.6736%, 1673 bp), and finally the *rps16* intron (1.6181%, 927 bp). Eurosids II top three regions for Shaw et al. (2014) were *ndhF-rpl32* (rank 1), *psbZ-trnG* (rank 2), and *trnT-trnL* (rank 3).

For *Olea*, the most informative regions under criterion 1 were *psbC-psbZ* (0.0411, 1045 bp), *trnS*(UGA)-*trnfM* (0.0333, 1203 bp), and *clpP* intron 2 (0.0313, 702 bp). The highest number of variable characters (criterion 2) were found in *rps16-trnQ* (29 characters, 2739 bp), *psbC-psbZ* (22 characters, 1045 bp), and *trnS*(UGA)-*trnfM* (21 characters, 1203 bp). Criterion 3 (percent sequence divergence) was highest in the same three regions as under criterion 1: *psbC-psbZ* (2.0096%, 1045 bp), *trnS*(UGA)-*trnfM* (1.5794%, 1203 bp), and *clpP* intron 2 (1.4245%, 702 bp). Shaw et al. (2014) euasterids I top three included *rps16-trnQ* (rank 1), *rpl32-trnL* (rank 2), and *ndhC-trnV* (rank 3).

DISCUSSION

A large number of "universal" primers have been published for amplification of various chloroplast regions. Some are more degenerate than others, presumably to be more widely applicable. Degeneracy is not required, however, and may not lead to greater success in the laboratory. On the other hand, nondegenerate primers with poor fit are likely to fail, and some primers published as "universal" are not necessarily so. The universal barcoding primers of Dong et al. (2012) were the least likely to be useful across the 12 taxa examined here, with an average success rate of 65%, and a very poor 29% success rate in *Oryza*. In contrast, the primers designed by Scarcelli et al. (2011) specifically for monocots were exceedingly well-matched to the monocots sampled (97% in *Acorus*, 93% in *Cymbidium*, 92% in *Oryza*, and 88% in *Canna*), and a good match across all angiosperms.

Unlike previous analyses, this study used published genomes and primer sequences to infer the likelihood of amplification success. Only a small number of published primers were evaluated, and additional primers will be added to future analyses. Indeed, as mentioned in the introduction, Ebert and Peakall (2009) and Dong et al. (2013) have primers that could be evaluated as well as those of Doorduin et al. (2011) designed for species of Asteraceae. The evaluation conducted here shows parallels to prior studies in that general conclusions or recommendations are difficult to distill. For each region, there may be a number of primer pair options. Which primer pair is best is highly variable and depends upon the taxon being investigated. Scarcelli et al. (2011) primers are the best option for monocots in general, but will fail in specific combinations (e.g., *trnH-psbA* for *Canna*, *atpF* intron/exon for *Cymbidium*, and *trnD-trnT* for *Oryza*). Dong et al. (2012) primers are generally less successful, but they are the only primers that will work for *psbM-trnD* in *Amborella* and *Magnolia*. In several instances, a primer will work for some, but not all species in a genus, like the Scarcelli et al. (2011) *matK* primers in *Cymbidium* or the *trnK-rps16* primers in *Helianthus*. Table 3 provides a quick summary of primer match for the top regions according to Shaw et al. (2014).

Prior studies have done an excellent job assessing variability of various noncoding regions across a diversity of angiosperms, particularly the recent work of Shaw et al. (2014). Those studies focused on infrageneric or even intergeneric comparisons. Here I compare sequence variability within species to see if the same markers are identified as the most variable, under slightly different criteria. This comparison was specifically avoided by Shaw et al. (2014) due to the small number of variable characters. The fastest regions identified here for *Oryza* were (depending upon criterion) *clpP-psbB*, *atpB-rbcL*, *psbM-trnD*, and *rbcL-accD*. In contrast, Shaw identified *ndhF-rpl32*, *ndhC-trnV*, and *psbM-trnD* as the fastest regions for monocots, with only one region of overlap between the two. For *Fragaria* (eurosids I), the list has no overlap at all. *Olea* (eurosids II) and *Gossypium* (euasterids I) each only overlap for a single region between the two studies. The lack of consensus over which region is the most variable at lower taxonomic levels has been pointed out by a number of papers including Särkinen and George (2013) for *Solanum*, and for 19 species pairs as demonstrated by Shaw et al. (2014). The comparison made here only adds to the argument that there is an acute need for additional comparative information.

Shaw et al. (2014) provided a solid foundation for which markers evolve the most quickly in major angiosperm clades, yet the fastest regions identified here for subspecies comparisons share little overlap with Shaw's regions. This finding suggests the need for a thorough exploration of markers prior to undertaking a large comparative sequencing project. The methods employed here to examine expected primer utility can easily be applied to any taxon, provided complete chloroplast genomic data are available. When complete genome data are lacking, the results presented here can provide a rough estimate of the "best primers," but this remains a work in progress.

LITERATURE CITED

Asano, T., T. Tsudzuki, S. Takahashi, H. Shimada, and K. Kadowaki. 2004. Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*) chloroplast genome: A comparative analysis of four monocot chloroplast genomes. *DNA Research* 11: 93–99.

Asmussen, C. B., and M. W. Chase. 2001. Coding and noncoding plastid DNA in palm systematics. *American Journal of Botany* 88: 1103–1117.

Barrett, C. F., C. D. Specht, J. Leebens-Mack, D. W. Stevenson, W. B. Zomlefer, and J. I. Davis. 2014. Resolving ancient radiations: Can complete plastid gene sets elucidate deep relationships among the tropical gingers (Zingiberales)? *Annals of Botany* 113: 119–133.

Besnard, G., P. Hernandez, B. Khadari, G. Dorado, and V. Savolainen. 2011. Genomic profiling of plastid DNA variation in the Mediterranean olive tree. *BMC Plant Biology* 11: 80.

Bock, D. G., N. C. Kane, D. P. Ebert, and L. H. Rieseberg. 2014. Genome skimming reveals the origin of the Jerusalem Artichoke tuber crop species: Neither from Jerusalem nor an artichoke. *New Phytologist* 201: 1021–1030.

Bortiri, E., D. Coleman-Derr, G. R. Lazo, O. D. Anderson, and Y. Q. Gu. 2008. The complete chloroplast genome sequence of *Brachypodium distachyon*: Sequence comparison and phylogenetic analysis of eight grass plastomes. *BMC Research Notes* 1: 61. 10.1186/1756-0500-1-61.

Bruneau, A., J. J. Doyle, and J. D. Palmer. 1990. A chloroplast DNA inversion as a subtribal character in the Phaseoleae (Leguminosae). *Systematic Botany* 15: 378–386.

Downloaded From: https://complete.bioone.org/journals/Applications-in-Plant-Sciences on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Calsa, T. Jr., D. M. Carraro, M. R. Benatti, A. C. Barbosa, J. P. Kitajima, and H. Carrer. 2004. Structural features and transcript-editing analysis of sugarcane (*Saccharum officinarum* L.) chloroplast genome. *Current Genetics* 46: 366–373.

Dong, W., J. Liu, J. Yu, L. Wang, and S. Zhou. 2012. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE* 7: e35071 10.1371/journal.pone.0035071.

Dong, W., C. Xu, T. Cheng, K. Lin, and S. Zhou. 2013. Sequencing angiosperm plastid genomes made easy: A complete set of universal primers and a case study on the phylogeny of Saxifragales. *Genome Biology and Evolution* 5: 989–997.

Doorduin, L., B. Gravendeel, Y. Lammers, Y. Ariyurek, T. Chin-A-Woeng, and K. Vrierling. 2011. The complete chloroplast genome of 17 individuals of pest species *Jacobaea vulgaris*: SNPs, microsatellites and barcoding markers for population phylogenetic studies. *DNA Research* 18: 93–105.

Ebert, D., and R. Peakall. 2009. A new set of universal *de novo* sequencing primers for extensive coverage of noncoding chloroplast DNA: New opportunities for phylogenetic studies and cpSSR discovery. *Molecular Ecology Resources* 9: 777–783.

Gaut, B. S., S. V. Muse, W. D. Clark, and M. T. Clegg. 1992. Relative rates of nucleotide substitution at the *rbcL* locus of monocotyledonous plants. *Journal of Molecular Evolution* 35: 292–303.

Goremykin, V. V., K. I. Hirsch-Ernst, S. Wölfl, and F. H. Hellwig. 2003. Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. *Molecular Biology and Evolution* 20: 1499–1505.

Goremykin, V. V., K. I. Hirsch-Ernst, S. Wölfl, and F. H. Hellwig. 2004. The chloroplast genome of *Nymphaea alba*: Whole-genome analyses and the problem of identifying the most basal angiosperm. *Molecular Biology and Evolution* 21: 1445–1454.

Goremykin, V. V., B. Holland, K. I. Hirsch-Ernst, and F. H. Hellwig. 2005. Analysis of *Acorus calamus* chloroplast genome and its phylogenetic implications. *Molecular Biology and Evolution* 22: 1813–1822.

Graham, S. W., P. A. Reeves, A. C. E. Burns, and R. G. Olmstead. 2000. Microstructural changes in noncoding chloroplast DNA: Interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *International Journal of Plant Sciences* 161: S83–S96.

Grivet, D., B. Heinze, G. G. Vendramin, and R. J. Petit. 2001. Genome walking with consensus primers: Application to the large single copy region of chloroplast DNA. *Molecular Ecology Notes* 1: 345–349.

Heinze, B. 2007. A database of PCR primers for the chloroplast genomes of higher plants. *Plant Methods* 3: 4.

Hiratsuka, J., H. Shimada, R. Whittier, T. Ishibashi, M. Sakamoto, M. Mori, C. Kondo, et al. 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Molecular & General Genetics* 217: 185–194.

Huang, X., N. Kurata, X. Wei, Z.-X. Wang, A. Wang, Q. Zhao, Y. Zhao, et al. 2012. A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490: 497–501.

Hupfer, H., M. Swiatek, S. Hornung, R. G. Herrmann, R. M. Maier, W. L. Chiu, and B. Sears. 2000. Complete nucleotide sequence of the *Oenothera elata* plastid chromosome, representing plastome I of the five distinguishable Euoenothera plastomes. *Molecular & General Genetics* 263: 581–585.

IDT (Integrated DNA Technologies). 2009. Degenerate sequences and non-standard bases: A quick look. Technical publication downloaded from http://www.idtdna.com/pages/support/technical-vault/reading-room/technical-reports [accessed 3 December 2014].

Jansen, R. K., and J. D. Palmer. 1987a. A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). *Proceedings of the National Academy of Sciences, USA* 84: 5818–5822.

Jansen, R. K., and J. D. Palmer. 1987b. Chloroplast DNA from lettuce and *Barnadesia* (Asteraceae): Structure, gene localization, and characterization of a large inversion. *Current Genetics* 11: 553–564.

Kato, T., T. Kaneko, S. Sato, Y. Nakamura, and A. Tabata. 2000. Complete structure of the chloroplast genome of a legume, *Lotus japonicus. DNA Research* 7: 323–330.

Kelchner, S. A., and L. G. Clark. 1997. Molecular evolution and phylogenetic utility of the chloroplast *rpl16* intron in *Chusquea* and the Bambusoideae (Poaceae). *Molecular Phylogenetics and Evolution* 8: 385–397.

Leseberg, C. H., and M. R. Duvall. 2009. The complete chloroplast genome of *Coix lacryma-jobi* and a comparative molecular evolutionary analysis of plastomes in cereals. *Journal of Molecular Evolution* 69: 311–318.

Li, C., A. Zhou, and T. Sang. 2006. Genetic analysis of rice domestication syndrome with the wild annual species, *Oryza nivara. New Phytologist* 170: 185–194.

Maier, R. M., K. Neckermann, G. L. Igloi, and H. Kossel. 1995. Complete sequence of the maize chloroplast genome: Gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *Journal of Molecular Biology* 251: 614–628.

Moore, M. J., C. D. Bell, P. S. Soltis, and D. E. Soltis. 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences, USA* 104: 19363–19368.

Njuguna, W., A. Liston, R. Cronn, T. L. Ashman, and N. Bassil. 2013. Insights into phylogeny, sex function and age of *Fragaria* based on whole chloroplast genome sequencing. *Molecular Phylogenetics and Evolution* 66: 17–29.

Ogihara, Y., K. Isono, T. Kojima, A. Endo, M. Hanaoka, T. Shiina, T. Terachi, et al. 2002. Structural features of a wheat plastome as revealed by complete sequencing of chloroplast DNA. *Molecular Genetics and Genomics* 266: 740–746.

Rambaut, A. 1996. Se-Al (v2.0a11) Sequence Alignment Editor. Available at http://evolve.zoo.ox.ac.uk/. University of Oxford, Oxford, United Kingdom.

Rychlik, W. 2002. Oligo Primer Analysis Software v. 6. Molecular Biology Insights, Cascade, Colorado, USA.

Särkinen, T., and M. George. 2013. Predicting plastid marker variation: Can complete plastid genomes from closely related species help? *PLoS ONE* 8: e82266. 10.1371/journal.pone.0082266.

Saski, C., S.-B. Lee, H. Daniell, T. C. Wood, J. Tomkins, H.-G. Kim, and R. K. Jansen. 2005. Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes. *Plant Molecular Biology* 59: 309–322.

Saski, C., S.-B. Lee, S. Fjellheim, C. Guda, R. K. Jansen, H. Juo, J. Tomkins, et al. 2007. Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes. *Theoretical and Applied Genetics* 115: 571–590.

Sato, S., Y. Nakamura, T. Kaneko, E. Asamizu, and S. Tabata. 1999. Complete structure of the chloroplast genome of *Arabidopsis thaliana. DNA Research* 6: 283–290.

Scarcelli, N., A. Barnaud, W. Eiserhardt, U. A. Treier, M. Seveno, A. d'Anfray, Y. Vigouroux, and J.-C. Pintaud. 2011. A set of 100 chloroplast DNA primer pairs to study population genetics and phylogeny in monocotyledons. *PLoS ONE* 6: e19954 10.1371/journal.pone.0019954.

Schmitz-Linneweber, C., R. M. Maier, J. P. Alcaraz, A. Cottet, R. G. Herrmann, and R. Mache. 2001. The plastid chromosome of spinach (*Spinacia oleracea*): Complete nucleotide sequence and gene organization. *Plant Molecular Biology* 45: 307–315.

Schmitz-Linneweber, C., R. Regel, T. G. Du, H. Hupfer, R. G. Herrmann, and R. M. Maier. 2002. The plastid chromosome of *Atropa belladonna* and its comparison with that of *Nicotiana tabacum*: The role of RNA editing in generating divergence in the process of plant speciation. *Molecular Biology and Evolution* 19: 1602–1612.

Shahid Masood, M., T. Nishikawa, S. Fukuoka, P. K. Njenga, T. Tsudzuki, and K. Kadowaki. 2004. The complete nucleotide sequence of wild rice (*Oryza nivara*) chloroplast genome: First genome wide comparative sequence analysis of wild and cultivated rice. *Gene* 340: 133–139.

Downloaded From: https://complete.bioone.org/journals/Applications-in-Plant-Sciences on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Shaw, J., E. B. Lickey, J. B. Beck, S. B. Farmer, W. Liu, J. Miller, K. C. Siripun, et al. 2005. The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* 92: 142–166.

Shaw, J., E. B. Lickey, E. E. Schilling, and R. L. Small. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. *American Journal of Botany* 94: 275–288.

Shaw, J., H. L. Shafer, O. R. Leonard, M. J. Kovach, M. Schorr, and A. B. Morris. 2014. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: The tortoise and the hare IV. *American Journal of Botany* 101: 1987–2004.

Shinozaki, K., M. Ohme, M. Tanaka, T. Wakasugi, N. Hayashida, T. Matsubayashi, N. Zaita, et al. 1986. The complete nucleotide sequence of tobacco chloroplast genome: Its gene organization and expression. *EMBO Journal* 5: 2043–2049.

Shulaev, V., D. J. Sargent, R. N. Crowhurst, T. C. Mockler, O. Folkerts, A. L. Delcher, P. Jaiswal, et al. 2011. The genome of woodland strawberry (*Fragaria vesca*). *Nature Genetics* 43: 109–116.

Tang, J., H. Xia, M. Cao, X. Zhang, W. Zeng, S. Hu, W. Tong, et al. 2004. A comparison of rice chloroplast genomes. *Plant Physiology* 135: 412–420.

Timme, R. E., J. V. Kuehl, J. L. Boore, and R. K. Jansen. 2007. A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: Identification of divergent regions and categorization of shared repeats. *American Journal of Botany* 94: 302–312.

Xu, Q., G. Xiong, P. Li, F. He, Y. Huang, K. Wang, Z. Li, and J. Hua. 2012. Analysis of complete nucleotide sequences of 12 *Gossypium* chloroplast genomes: Origin and evolution of allotetraploids. *PLoS ONE* 7: E37128. 10.1371/journal.pone.0037128.

Yang, J. B., M. Tang, H. T. Li, Z. R. Zhang, and D. Z. Li. 2013. Complete chloroplast genome of the genus *Cymbidium*: Lights into the species identification, phylogenetic implications and population genetic analyses. *BMC Evolutionary Biology* 13: 84.

Appendix 1. Complete chloroplast genome sequences used to design universal flowering plant primers for 36 plastid noncoding regions. Format: Organism; GenBank number and version; publication.

Basal Dicot Grade:

1. *Amborella trichopoda* Baill.; NC_005086.1; Goremykin et al., 2003.

Monocots:

2. *Oryza nivara* Sharma & Shastry; NC_005973.1; Shahid Masood et al., 2004.
3. *Oryza sativa* L.; NC_001320.1; Hiratsuka et al., 1989.
4. *Saccharum* hybrid; NC_005878.2; Calsa et al., 2004.
5. *Saccharum officinarum* L.; NC_006084.1; Asano et al., 2004.
6. *Triticum aestivum* L.; NC_002762.1; Ogihara et al., 2002.
7. *Zea mays* L.; NC_001666.2; Maier et al., 1995.

Eudicots:

8. *Arabidopsis thaliana* (L.) Heynh.; NC_000932.1; Sato et al., 1999.
9. *Atropa belladonna* L.; NC_004561.1; Schmitz-Linneweber et al., 2002.
10. *Calycanthus floridus* L. var. *glaucus* (Willd.) Torr. & A. Gray; NC_004993.1; Goremykin et al., unpublished (Goremykin, V., K. Hirsch-Ernst, S. Wolfl, and F. Hellwig. Complete structure of the chloroplast genome of *Calycanthus fertilis*. Direct GenBank submission 9 July 2003).
11. *Lotus japonicus* (Regel) K. Larsen; AP002983.1; Kato et al., 2000.
12. *Medicago truncatula* Gaertn.; NC_003119.6; Lin et al., unpublished (Lin, S., H. Wu, H. Jia, P. Zhang, R. Dixon, G. May, R. Gonzales, and B. A. Roe. *Medicago truncatula* variety Jema Long A-17 chloroplast, complete sequence. Direct GenBank submission 31 August 2001).
13. *Nicotiana tabacum* L.; Z00044.1; Shinozaki et al., 1986. Note: this sequence has been updated since this article was published.
14. *Nymphaea alba* L.; NC_006050.1; Goremykin et al., 2004.
15. *Oenothera elata* Kunth subsp. *hookeri* (Torr. & A. Gray) W. Dietr. & W. L. Wagner; NC_002693.1; Hupfer et al., 2000. Note: this sequence has been updated since this article was published.
16. *Spinacia oleracea* L.; NC_002202.1; Schmitz-Linneweber et al., 2001.

Downloaded From: https://complete.bioone.org/journals/Applications-in-Plant-Sciences on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Appendix 2. Comparison of chloroplast regions with published primer pairs.

Approx. *Nicotiana* order[a]	Primary type	Location[b]	Genomic region	Shaw et al., 2005, 2007	Ebert and Peakall, 2009	Scarcelli et al., 2011	Dong et al., 2012	Dong et al., 2013	Current study
1	IGS	LSC	*trnH*(GUG)-*psbA*	✓		✓	✓	✓	
2	Exon	LSC	*psbA* exon			✓		✓	
3	IGS	LSC	*psbA-trnK*(UUU)	✓		✓		✓	
4	IGS	LSC	3′*trnK*(UUU)-*matK*	✓	✓				
5	Exon	LSC	*matK* exon			✓	*	✓	
6	IGS	LSC	*matK-trnK5′*	✓	✓			✓	
7	IGS	LSC	*trnK*(UUU)-*rps16*	✓	✓	✓		✓	
8	Intron	LSC	*rps16* intron	✓	✓	✓		✓	
9	IGS	LSC	*rps16-trnQ*(UUG)	✓	✓	✓	✓	✓	
10	IGS	LSC	*trnQ*(UUG)-*psbK*		✓	✓		*	✓
11	IGS	LSC	*psbK-trnS*(GCU)		✓	✓		*	✓
12	IGS	LSC	*trnS*(GCU)-*trnG*(UCC) and intron	✓	✓	✓	✓	*	
13	Intron	LSC	*trnG*(UCC) intron	✓	✓	✓			
14	IGS	LSC	*trnG*(UCC)-*atpA*		*	✓		✓	✓
15	Exon	LSC	*atpA* exon			✓		✓	
16	IGS	LSC	*atpA-atpF*		✓			✓	
17	Intron	LSC	*atpF* intron		✓	✓		✓	✓
18	IGS	LSC	*atpF-atpH*		✓	✓		✓	✓
19	IGS	LSC	*atpH-atpI*	✓	✓	✓	✓	✓	
20	Exon	LSC	*atpI* exon			✓		✓	
21	IGS	LSC	*atpI-rps2*		✓	✓		✓	
22	Exon	LSC	*rps2* exon			✓		*	
23	IGS	LSC	*rps2-rpoC2*		✓	✓			
24	IGS	LSC	*rpoC2-rpoC1*			✓		*	
25	Intron	LSC	*rpoC1* intron/exon 1		✓	✓		✓	✓
26	Exon	LSC	*rpoC1* exon 2			✓		✓	
27	Exon	LSC	*rpoB2* exon					✓	
28	IGS	LSC	*rpoB-trnC*(GCU)	✓	✓	✓	✓	✓	
29	IGS	LSC	*trnC*(GCU)-*ycf6*	✓					
30	IGS	LSC	*trnC*(GCU)-*petN*		✓	✓		✓	
31	IGS	LSC	*petN-trnD*			✓			
32	IGS	LSC	*petN-psbM*		✓		✓	✓	
33	IGS	LSC	*ycf6-psbM*	✓					
34	IGS	LSC	*psbM-trnD*(GUC)	✓	✓		✓	✓	
35	IGS	LSC	*trnD*(GUC)-*trnT*(GGU)		✓	✓		✓	
36	IGS	LSC	*trnT*(GGU)-*psbD*	✓	✓	✓	✓	✓	
37	Exon	LSC	*psbD* exon			✓		✓	
38	Exon	LSC	*psbC* exon			✓		✓	
39	IGS	LSC	*psbC-psbZ*		✓	✓		*	
40	IGS	LSC	*trnS*(UGA)-*trnG*(GCC)				✓		
41	IGS	LSC	*trnG*(GCC)-*rpS14*		✓				
42	IGS	LSC	*trnS*(UGA)-*trnfM*	✓					
43	IGS	LSC	*trnS*(UGA)-*psbZ*						✓
44	IGS	LSC	*psbZ-trnfM*(CAU)			✓			
45	IGS	LSC	*trnfM*(CAU)-*psaB*			✓			
46	Exon	LSC	*psaB* exon					✓	
47	Exon	LSC	*psaA* exon					✓	
48	IGS	LSC	*psaA-ycf3*		✓	✓		✓	✓
49	Intron	LSC	*ycf3* intron 2		✓	✓		✓	✓
50	Intron	LSC	*ycf3* intron 1		✓	✓		✓	✓
51	IGS	LSC	*ycf3-trnS*(GGA)		✓				✓
52	IGS	LSC	*ycf3-rps4*			✓		✓	
53	IGS	LSC	*trnS*(GGA)-*rpS4-trnT*(UGU)	✓					
54	IGS	LSC	*rpS4-trnT*(UGU)					*	✓
55	IGS	LSC	*trnT*(UGU)-*trnL*(UAA)	✓	✓			*	
56	Intron	LSC	*trnL*(UAA) intron	✓		✓		*	
57	IGS	LSC	*trnL*(UAA)-*trnF*(GAA)	✓				*	
58	IGS	LSC	*trnL*(UAA)-*ndhJ*	✓		✓		✓	
59	IGS	LSC	*trnF*(GAA)-*ndhJ*		✓				✓
60	IGS	LSC	*ndhJ-ndhC*					✓	
61	IGS	LSC	*ndhC-trnV*(UAC)	✓	✓	✓	✓	✓	✓
62	Intron	LSC	*trnV*(UAC) intron		✓	✓		✓	✓
63	IGS	LSC	*trnV*(UAC)-*atpE*					✓	✓
64	IGS	LSC	*trnV*(UAC)-*atpB*			✓		✓	
65	Exon	LSC	*atpB* exon			✓		✓	
66	IGS	LSC	*atpB-rbcL*		✓	✓		✓	✓
67	Exon	LSC	*rbcL* exon			✓		✓	
68	IGS	LSC	*rbcL-accD*		✓	✓		✓	✓

Downloaded From: https://complete.bioone.org/journals/Applications-in-Plant-Sciences on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Appendix 2. Continued.

Approx. *Nicotiana* order[a]	Primary type	Location[b]	Genomic region	Shaw et al., 2005, 2007	Ebert and Peakall, 2009	Scarcelli et al., 2011	Dong et al., 2012	Dong et al., 2013	Current study
69	Exon	LSC	*accD* exon			✓		✓	
70	IGS	LSC	*accD-psaI*	✓	✓	✓	✓	*	✓
71	IGS	LSC	*psaI-ycf4*		✓	✓		*	✓
72	Exon	LSC	*ycf4* exon			✓		✓	
73	IGS	LSC	*ycf4-ycf10*(*cemA*)		*			✓	✓
74	Exon	LSC	*cemA*					✓	
75	IGS	LSC	*ycf4-petA*			✓		*	
76	Exon	LSC	*petA* exon			✓		✓	
77	IGS	LSC	*petA-psbJ*	✓	✓		✓	✓	✓
78	IGS	LSC	*psbJ-psbE*					✓	
79	IGS	LSC	*petA-psbL*			✓			
80	IGS	LSC	*psbE-petL*	✓	✓		✓	✓	
81	IGS	LSC	*petL-psaJ*						✓
82	IGS	LSC	*petL-trnP*(UGG)			✓		✓	
83	IGS	LSC	*trnW*(CCA)-*psaJ*				✓	✓	
84	IGS	LSC	*trnP*(UGG)-*rps18*		*	✓			
85	IGS	LSC	*psaJ-rpl20*		*			*	✓
86	IGS	LSC	*rps18-rps12*			✓		*	
87	IGS	LSC	*rpl20-rps12*	✓				*	✓
88	IGS	LSC	*rps12-psbB*			✓			
89	IGS	LSC	*rps12-clpP*		✓	✓		*	
90	Intron	LSC	*clpP* intron 2		✓	✓	✓	✓	✓
91	Intron	LSC	*clpP* intron 1		✓	✓	✓	✓	✓
92	IGS	LSC	*clpP-psbB*		✓	✓		✓	✓
93	Exon	LSC	*psbB* exon			✓		✓	
94	IGS	LSC	*psbB-psbH*	✓				✓	
95	IGS	LSC	*psbH-petBE2*		✓			✓	✓
96	Intron	LSC	*petB* intron/exon 2			✓		✓	
97	IGS	LSC	*petBE2-petDE2*		✓	✓	✓	✓	✓
98	Intron	LSC	*petD* intron/exon 2			✓		✓	
99	IGS	LSC	*petD-rpoA*			✓		✓	
100	Exon	LSC	*rpoA* exon					✓	
101	IGS	LSC	*rpoA-rps11*					✓	
102	IGS	LSC	*rps11-rps8*		✓	✓		✓	
103	Exon	LSC	*rps8* exon					✓	
104	IGS	LSC	*rpl36-rpl14*	✓					
105	IGS	LSC	*rps8-rpl16*		✓	✓		✓	
106	Intron	LSC	*rpl16* intron	✓		✓			
107	IGS	LSC	*rpl16-rps3*		✓	✓		✓	
108	Exon	LSC	*rps3* exon			✓		✓	
109	IGS	LSC	*rps3-rps19*		✓			*	✓
110	IGS	LSC	*rpl22-rpl2*			✓		*	
111	Intron	IRb	*rpl2* intron/exon 1-2			✓		✓	
112	IGS	IRb	*rpl23-ycf2*			✓		*	
113	Exon	IRb	*ycf2* exon					✓	
114	IGS	IRb	*ycf2-ndhB*			✓		✓	
115	Exon	IRb	*ndhB* exon 2			✓		✓	
116	Intron	IRb	*ndhB* intron/exon 1			✓		✓	
117	IGS	IRb	*ndhB-rps7*			✓		✓	
118	IGS	IRb	*rps7-rps12*					✓	
119	Intron	IRb	*rps12* intron/exon			✓			
120	IGS	IRb	*rps12-trnV*(GAC)			✓		✓	
121	IGS	IRb	*trnV*(GAC)-*rrn16*			✓		✓	
122	Exon	IRb	*rrn16* exon			✓		✓	
123	IGS	IRb	*rrn16-trnI*(GAU)			✓		✓	
124	Intron	IRb	*trnI*(GAU) intron			✓		*	
125	Intron	IRb	*trnA*(UGC) intron			✓		*	
126	IGS	IRb	*trnA*(UGC)-*rrn23*			✓		*	
127	Exon	IRb	*rrn23* exon					✓	
128	IGS	IRb	*rrn4,5-trnN*(GUU)			✓		✓	
129	IGS	IRb	*trnN*(GUU)-*ycf1*					✓	
130	IGS	IRb/SSC	*ycf1-ndhF*					✓	
131	Exon	SSC	*ndhF* exon				✓	✓	
132	IGS	SSC	*ndhF-rpl32*	✓		✓		✓	
133	IGS	SSC	*rpl32-ccsA*			✓		✓	
134	IGS	SSC	*rpl32-trnL*(UAG)	✓			✓		
135	Exon	SSC	*ccsA* exon			✓		✓	
136	IGS	SSC	*ccsA-ndhD*			✓		✓	✓

Downloaded From: https://complete.bioone.org/journals/Applications-in-Plant-Sciences on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Appendix 2. Continued.

Approx. *Nicotiana* order[a]	Primary type	Location[b]	Genomic region	Shaw et al., 2005, 2007	Ebert and Peakall, 2009	Scarcelli et al., 2011	Dong et al., 2012	Dong et al., 2013	Current study
137	Exon	SSC	*ndhD* exon			✓		✓	
138	IGS	SSC	*ndhD-ndhE*					✓	
139	IGS	SSC	*psaC-ndhE*						✓
140	IGS	SSC	*psaC-ndhG*			✓			
141	IGS	SSC	*ndhE-ndhI*					✓	✓
142	Exon	SSC	*ndhG* exon			✓		*	
143	IGS	SSC	*ndhG-ndhI*			✓		*	
144	Intron	SSC	*ndhA* intron	✓		✓	✓	✓	
145	IGS	SSC	*ndhA-ndhH*					✓	
146	Exon	SSC	*ndhH* exon			✓		✓	
147	IGS	SSC	*ndhH-rps15*					✓	
148	IGS	SSC/IRa	*rps15-ycf1*			✓			✓
149	IGS	IRa	*ycf1-rrn5*			✓			
Bonus	IGS	LSC	*rbcL-psaI*						✓
Bonus	IGS	LSC	*trnS-psbD*				✓		

[a] Several regions overlap.
[b] IR = inverted repeat; LSC = large single-copy region; SSC = small single-copy region.
* Slightly different region from that listed.

Downloaded From: https://complete.bioone.org/journals/Applications-in-Plant-Sciences on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Appendix 3. Complete chloroplast genome sequences used to assess primer utility. Format: Organism; GenBank number and version; publication.

Basal Dicot Grade:

1. *Amborella trichopoda* Baill.; NC_005086.1; Goremykin et al., 2003.
2. *Magnolia grandiflora* L.; NC_020318.1; Li et al., unpublished (direct GenBank submission dated 22 February 2013).

Monocots:

3. *Acorus calamus* L.; AJ879453.1; Goremykin et al., 2005.
4. *Cymbidium aloifolium* (L.) Sw.; NC_021429.1; Yang et al., 2013.
5. *Cymbidium mannii* Rchb. f.; NC_021433.1; Yang et al., 2013.
6. *Cymbidium sinense* (Jacks. ex Andrews) Willd.; NC_021430.1; Yang et al., 2013.
7. *Cymbidium tortisepalum* Fukuy.; NC_021431.1; Yang et al., 2013.
8. *Cymbidium tracyanum* Rolfe; NC_021432.1; Yang et al., 2013.
9. *Oryza nivara* Sharma & Shastry; NC_005973.1; Shahid Masood et al., 2004.
10. *Oryza sativa* L. Indica group; NC_008155.1; Tang et al., 2004.
11. *Oryza sativa* L. Japonica group; NC_001320.1; Hiratsuka et al., 1989.
12. *Canna indica* L.; KF601570.1; Barrett et al., 2014.

Basal Eudicot Grade:

13. *Ceratophyllum demersum* L.; NC_009962.1; Moore et al., 2007.
14. *Nelumbo lutea* Willd.; NC_015605.1; Quan and Ding, unpublished (direct GenBank submission dated 16 February 2009).
15. *Nelumbo nucifera* Gaertn.; NC_015610; Quan and Ding, unpublished (direct GenBank submission dated 16 February 2009).

Eurosids I:

16. *Fragaria vesca* L. subsp. *bracteata* (A. Heller) Staudt; NC_018766.1; Njuguna et al., 2013.
17. *Fragaria vesca* L. subsp. *vesca*; NC_015206.1; Shulaev et al., 2011.

Eurosids II:

18. *Gossypium herbaceum* L.; NC_023215.1; Shang et al., unpublished (Shang, M., K. Wang, J. Hua, F. Liu, C. Wang, X. Zhang, Y. Wang, and S. Li. *Gossypium herbaceum* chloroplast, complete genome. Direct GenBank submission 11 February 2011).
19. *Gossypium herbaceum* L. subsp. *africanum* (G. Watt) Vollesen; NC_016692.1; Xu et al., 2012.

Euasterids I:

20. *Olea europaea* L.; NC_013707.2; Messina, unpublished (Messina, R. *Olea europaea* chloroplast, complete genome. Direct GenBank submission 3 March 2007).
21. *Olea europaea* L. subsp. *cuspidata* (Wall. ex G. Don) Cif.; NC_015604.1; Besnard et al., 2011.
22. *Olea europaea* L. subsp. *europaea*; NC_015401.1; Besnard et al., 2011.
23. *Olea europaea* L. subsp. *maroccana* (Greuter & Burdet) P. Vargas; NC_015623.1; Besnard et al., 2011.

Euasterids II:

24. *Helianthus annuus* L.; NC_007977.1; Timme et al., 2007.
25. *Helianthus decapetalus* L.; NC_023110.1; Bock et al., 2014.
26. *Helianthus divaricatus* L.; NC_023109.1; Bock et al., 2014.
27. *Helianthus giganteus* L.; NC_023107.1; Bock et al., 2014.
28. *Helianthus grosseserratus* M. Martens; NC_023108.1; Bock et al., 2014.
29. *Helianthus hirsutus* Raf.; NC_023111.1; Bock et al., 2014.
30. *Helianthus maximiliani* Schrad.; NC_023114.1; Bock et al., 2014.
31. *Helianthus strumosus* L.; NC_023113.1; Bock et al., 2014.
32. *Helianthus tuberosus* L.; NC_023112.1; Bock et al., 2014.

Downloaded From: https://complete.bioone.org/journals/Applications-in-Plant-Sciences on 28 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use