American Journal of Botany

American Journal of Botany 97(8): 1391–1406. 2010.

AN EVALUATION OF CANDIDATE PLANT DNA BARCODES AND ASSIGNMENT METHODS IN DIAGNOSING 29 SPECIES IN THE GENUS *AGALINIS* (OROBANCHACEAE)[1]

JAMES B. PETTENGILL[2,4] AND MAILE C. NEEL[2,3]

[2]Behavior, Ecology, Evolution, and Systematics Graduate Program, 2174 Plant Sciences Building, University of Maryland College Park, College Park, Maryland 20742; and [3]Department of Plant Science and Landscape Architecture and Department of Entomology, 2116 Plant Sciences Building, University of Maryland College Park, College Park, Maryland 20742 USA

• *Premise of the study*: DNA barcoding has been proposed as a useful technique within many disciplines (e.g., conservation biology and forensics) for determining the taxonomic identity of a sample based on nucleotide similarity to samples of known taxonomy. Application of DNA barcoding to plants has primarily focused on evaluating the success of candidate barcodes across a broad spectrum of evolutionary divergence. Less attention has been paid to evaluating performance when distinguishing congeners or to differential success of analytical techniques despite the fact that the practical application and utility of barcoding hinges on the ability to distinguish closely related species.

• *Methods*: We tested the ability to distinguish among 92 samples representing 29 putative species in the genus *Agalinis* (Orobanchaceae) using 13 candidate barcodes and three analytical methods (i.e., threshold genetic distances, hierarchical tree-based, and diagnostic character differences). Due to questions regarding evolutionary distinctiveness of some taxa, we evaluated success under two taxonomic hypotheses.

• *Key results*: The *psbA-trnH* and *trnT-trnL* barcodes in conjunction with the "best close match" distance-based method best met the objectives of DNA barcoding. Success was also a function of the taxonomy used.

• *Conclusions*: In addition to accurately identifying query sequences, our results showed that DNA barcoding is useful for detecting taxonomic uncertainty; determining whether erroneous taxonomy or incomplete lineage sorting is the cause requires additional information provided by traditional taxonomic approaches. The magnitude of differentiation within and among the *Agalinis* species sampled suggests that our results inform how DNA barcoding will perform among closely related species in other genera.

Key words: *Agalinis*; conservation; DNA barcoding; monophyly; Orobanchaceae; taxonomy.

DNA barcoding promises to provide a fast and reliable way to determine the taxonomic identity of an individual by sequencing a small portion of its genome and comparing this nucleotide sequence with sequences in a reference database (e.g., Hebert et al., 2003). This promise along with the increasing ease with which sequence data can be obtained from a range of taxa has spawned the ambitious goal of creating a genetic catalog of the world's biodiversity (Hebert et al., 2003; Chase et al., 2005; Kress et al., 2005; Savolainen et al., 2005). Such a reference collection can be an important tool for a broad range of applications including conducting rapid biodiversity assessments and forensics tests, detecting illegal wildlife trade, identifying species during cryptic life stages, and monitoring invasive species (Armstrong and Bar, 2005; Vogler, 2006; Darling and Blum, 2007; Dawnay et al., 2007; Little and Stevenson, 2007). Barcoding has also been shown to be a useful tool in the discovery of new species (Burns et al., 2008; Murray et al., 2008; Yassin et al., 2008). However, the grand promises made by proponents of DNA barcoding have generated concern, and there is extensive debate over exactly what it can contribute to various disciplines (e.g., Will et al., 2005; Rubinoff, 2006; Trewick, 2008).

Development of a reference database of DNA sequences for plant species has been hindered by difficulties in finding suitable DNA regions that can serve as a barcode. It is most desirable to have a single locus or a few loci that have highly conserved universal primer sequences and at the same time have sufficient nucleotide variation to diagnose species. Due to low rates of nucleotide substitution and high rates of chromosomal rearrangements within the plant mitochondrial genome (Palmer, 1985) and extensive gene duplication in the nuclear genome (Alvarez and Wendel, 2003), plastid loci have been considered the most promising candidate barcodes for plant species (e.g., Chase et al., 2005; Kress et al., 2005; Lahaye et al., 2008). Candidate barcodes from the plastid genome include both slowly evolving coding regions (e.g., *rbcL*, *rpoB*, or *matK*) and more rapidly evolving loci (e.g., *rps2*, *psbA-trnH*, and *trnT-trnF*). However, when used alone, the more conserved loci may not possess enough nucleotide variation to discriminate among closely related species and the more variable loci may be problematic because of homoplasy, lack of conserved priming regions, and difficulties in DNA sequence alignment among

[1] Manuscript received 19 June 2009; revision accepted 14 June 2010.

The authors thank J. M. Canne-Hilliker, G. Dieringer, J. Koontz, D. Lewis, J. L. Neff, and P. Polloni for collecting specimens; C. Frye, B. Gulotta, J. Hays, B. Horwith, M. Jordan, N. Murray, J. Nelson, M. Pelikan, P. Polloni, C. Raithel, S. Ruhren, P. Somers, B. Sorrie, W. Tyndall, and B. Zaremba for assistance in locating populations for collection and for facilitating permits; and J. M, Canne-Hilliker and J, Hays for helpful discussions regarding the taxonomy of *Agalinis*. This research was supported in part by the U.S. Fish and Wildlife Service, the University of Maryland Agricultural Experiment Station, and the University of Maryland Graduate Program in Behavior, Ecology, Evolution, and Systematics.

[4] Author for correspondence (e-mail: jpetteng@umd.edu)

doi:10.3732/ajb.0900176

distantly related species. To overcome these problems, a multilocus DNA barcoding system has been suggested (Chase et al., 2005; Fazekas et al., 2008; CBOL Plant Working Group, 2009) in which slowly evolving loci delineate individuals into families, genera, or groups within genera and the more rapidly evolving loci differentiate species within those higher groups (Newmaster et al., 2006; Kress and Erickson, 2007). The most recent suggestion from a study by a group of 52 scientists is to use the multilocus barcode consisting of *matK* and *rbcL* (CBOL Plant Working Group, 2009).

In addition to a barcode, application of DNA barcoding requires analytical methods that accurately assign query sequences of unknown taxonomic identity to species based on the sequences contained in the reference database. The Barcode of Life Data Systems (BOLD), an online initiative to facilitate implementation of DNA barcoding, assigns identities using a pairwise genetic distance threshold of 1% for animal species; query sequences less than that specified distance from a reference sequence are assigned to that species (Ratnasingham and Hebert, 2007). Pairwise genetic distances were also used by the Consortium for the Barcode of Life (CBOL) to evaluate the efficacy of candidate barcodes and identify the plant DNA barcode of *rbc*L coupled with *matK* (CBOL Plant Working Group, 2009). The primary criticism of distance-based methods is that any particular distance threshold is unlikely to be suitable for all species. There will undoubtedly be instances where intraspecific distances within one species will exceed interspecific distances in other species (Ferguson, 2002; DeSalle et al., 2005; Meyer and Paulay, 2005; Little and Stevenson, 2007). Differences between intra- and interspecific distances vary greatly across plant genera, and in general, there is much more overlap in these values in plant species than in animal species (Fazekas et al., 2009).

Alternative analytical methods have been developed to overcome the problems posed by identifying a single distance threshold. For example, methods based on the clustering of individuals on a phylogenetic tree have been suggested because through the bootstrap a measure of statistical support can be achieved and the resulting tree provides a visual representation of the affinities of the query sequence (Armstrong and Bar, 2005). However, tree-based methods have been criticized because of the difficulties in assigning taxonomy to a query sequence based on a hierarchical tree-like structure (DeSalle et al., 2005) and because low levels of divergence at a barcode locus could be insufficient to estimate phylogenetic relationships and yet may be able to differentiate among sequences (Erickson et al., 2008). A recently developed character-based method (Sarkar et al., 2008) identifies combinations of nucleotide character states that are unique to members of a species (i.e., are diagnostic characters) and, therefore, can be used to assign taxonomy if query sequences possess that combination of nucleotide characters. However, that character-based method relies on a phylogenetic tree and, therefore, suffers to some extent from the concerns associated with tree-based methods. Because they take advantage of different sequence characteristics, the three assignment techniques may result in different correct classification rates when evaluating the same DNA barcode. As such, the efficacy of DNA barcoding may be highly dependent on the assignment method and use of different analytical techniques has potentially contributed to the lack of consensus in identifying a suitable locus to serve as the plant DNA barcode (Erickson et al., 2008). Despite their importance in success of barcoding, alternative methods have not been rigorously compared using the same loci and samples.

The search for a DNA barcode in plants has focused primarily on assessing the reliability of amplification and correct assignment across the evolutionary breadth of land plants (e.g., Chase et al., 2005; Kress et al., 2005) or within geographically or ecologically defined communities (Lahaye et al., 2008). Much less effort has been spent assessing the potential to distinguish closely related species (Fazekas et al., 2009). Regardless of the analytical methods used, DNA barcoding will be highly successful in divergent taxa that have much smaller intraspecific than interspecific differences (i.e., that possess a large "barcoding gap"; Meyer and Paulay, 2005), are reciprocally monophyletic, and have diagnostic character differences (Fazekas et al., 2009). However, when time since divergence between species is too short for sorting of shared ancestral polymorphisms (e.g., <8.7 N_e (effective population size) generations; Hudson and Coyne, 2002; Rosenberg, 2003), using short sequences of DNA to diagnose species will potentially be challenging regardless of the assignment method employed (Ross et al., 2008). Simulations show failure rates become excessive when overlap between intra- and interspecific distances is >10% and when <80% of species are monophyletic (Ross et al., 2008). Lack of differentiation can be due to inaccurate taxonomy or low levels of divergence, both of which result in nonmonophyletic relationships. Paraphyly or polyphyly may be quite extensive among plant species (e.g., Rieseberg and Brouillet, 1994; Crisp and Chandler, 1996; Fazekas et al., 2009) suggesting that overlap in intraspecific and interspecific distances may be widespread. CBOL Plant Working Group (2009) illustrates the difficulty in discriminating among closely related species in that they were able to correctly classify 100% of samples to genus but 28% of cases were assigned to the wrong congeneric species. Consequently, the utility of DNA barcoding primarily rests on the ability to discriminate among closely related species (i.e., congeners), which has only recently begun to be tested (Newmaster et al., 2008; Fazekas et al., 2009; Hollingsworth et al., 2009).

The purpose of the present research was to contribute to filling this void by evaluating the performance of DNA barcoding in distinguishing 29 putative species in the plant genus *Agalinis* (family Orobanchaceae). Although our sampling is of limited taxonomic breadth (i.e., a single genus), sampling densely within a genus and within species is necessary to clarify whether DNA barcoding can succeed in making taxonomic designations to species rather than only higher taxonomic levels (e.g., to genus or species groups) (CBOL Plant Working Group, 2009; Fazekas et al., 2009). We evaluated the utility of the 13 barcodes (eight single-locus and five 2-locus barcodes) first by considering how easy it was to obtain a reliable sequence across all individuals. We then evaluated the combined utility of loci under each of three assignment methods as a barcode's ability to correctly classify individuals to species and to determine accurately when a sequence had no conspecific in the reference database. Because it will take a substantial amount of time to populate reference databases with representative accessions of all plant species, it is critically important for analytical methods to accurately identify when no conspecific is present in the database and, thus, reduce erroneous and false-positive taxonomic assignments (e.g., Ratnasingham and Hebert, 2007). The first method we investigated employed a pairwise genetic distance threshold (Meier et al., 2006), the second was based on the relationships depicted when a hierarchical clustering algorithm was applied to a matrix of pairwise genetic distances (Ross et al., 2008), and the third used unique combinations of nucleotide

character states that differentiate taxa (i.e., diagnostic nucleotide character differences; Sarkar et al., 2008).

The genus *Agalinis* provides an excellent test case for barcoding. It is well studied taxonomically (Pennell, 1913a, b, 1928, 1929, 1935; Canne, 1979, 1980, 1981, 1983, 1984; Canne-Hilliker, 1987; Kampny and Canne-Hilliker, 1987; Canne-Hilliker and Kampny, 1991; Canne-Hilliker and Dubrule, 1993), and the array of evolutionary divergence is relatively well understood and ranges from taxa that are highly differentiated to those that may not be evolutionarily distinct (Neel and Cummings, 2004; Pettengill and Neel, 2008). As such, it reflects the range of evolutionary distinctiveness one expects in plant genera and illuminates the issues that are likely to be present when barcoding other congeneric individuals. Lack of distinction among several taxa indicates that some taxonomic alignments may not be warranted (Neel and Cummings, 2004; Pettengill and Neel, 2008), which allows us to investigate the consequences of taxonomic uncertainty for barcoding. By focusing on the practical application of our results to conservation of rare *Agalinis* species, we are also able to evaluate the utility of DNA barcoding to conservation biology. Twenty-one of the sampled species are considered imperiled (S2) or critically imperiled (S1) in at least one state in which they occur; seven of these species are also globally vulnerable (G3 or G3-G4), and four are critically imperiled (G1) (Table 1; NatureServe, 2007). Because many of the species are particularly difficult to distinguish in the field, we are interested in knowing whether DNA barcoding can serve as a tool for field botanists from state and federal agencies to identify the taxonomy of individuals that represent species that are challenging to classify based on morphology. Greater certainty in identification can increase understanding of the true distribution and abundance of putatively rare species and facilitate appropriate management and priority setting.

TABLE 1. The number of conspecifics, general locations of sampled individuals, and conservation status for the 29 *Agalinis* species investigated.

Taxon	*N*	State	Status [a]
Section Erectae			
A. acuta	10	CT, RI, MA, NY, MD	G1/S1
A. aphylla	2	FL, AL	G3-G4/S2
A. decemloba	5	VA, NC	NR
A. gattingeri	3	MO, LA	G4/S1
A. obtusifolia	9	AL, SC, FL	G4-G5-Q/S1
A. oligophylla	4	AL, TX	G4/S1
A. skinneriana	3	MD, MO	G3/S1
A. tenella	8	GA, SC	NR
A. viridis	2	LA	G4/S1
Section Heterophyllae			
A. auriculata	2	IA, IL	G3-G4/S1
A. heterophylla	3	TX	G4-G5/S1
A. calycina	1	TX	G1/S1
Section Linifoliae			
A. linifolia	2	FL, GA	G4?/S1
Section Purpureae			
A. edwardsiana	1	TX	G4/S4
A. homalantha	2	TX	G5/S1
A. pulchella	2	GA, FL	G4-G5/S3?
A. strictifolia	2	TX	G4/SNR
A. navasotensis	2	TX	G1/S1
A. fasciculata	3	TX, GA, LA	G5/S1
A. harperi	2	FL, NC	G4?/SNR
A. maritima	2	TX, MA	G5/S2
A. paupercula	2	MA, NY	G5/S1
A. purpurea	4	VA, AL, MD, SC	G5/S1
A. laxa	2	SC, GA	G3-G4/S3?
A. plukenettii	2	FL, GA	G3-G5/S1
A. setacea	4	VA, MD	
Section Tenuifoliae			
A. filicaulis	2	FL, AL	G3-G4/S1
A. divaricata	2	FL	G3?/S1
A. tenuifolia	4	LA, VA, IA, MD	G5/S1

[a] Conservation status: global ranking (G1 = critically imperiled; G2 = imperiled; G3 = vulnerable to extinction or extirpation; G4 = apparently secure; G5 = demonstrably secure or widespread)/highest state ranking for each species (S1–S5 are equivalent to the global scale but applied to within a single state) (USA). When a range or question mark (?) is given, the precise conservation status is uncertain: NR and SNR= not ranked.

MATERIALS AND METHODS

Sampling—The database used to assess performance of the different loci and methods consisted of sequences from 92 individuals representing 1–10 accessions for 29 of the ~40 putative North American *Agalinis* species (Table 1). No specimens were available for the ~30 additional *Agalinis* species that exist in Mexico, Central America, and South America. The sampled species represented all North American sections and subsections of the genus, and 27 species were represented by more than one accession. When multiple individuals of a species were included, they were from different populations that often represented disjunct portions the species range, and as a result, we can assess how well DNA barcoding performs when there is the potential for substantial intraspecific differentiation due to geographic isolation. The inclusion of a representative of a population was dictated by the samples that could be obtained from J. M. Canne-Hilliker or that we could locate based on historical records and information from local and state botanists. The two species represented by a single individual allow us to assess the performance of methods when there is no conspecific reference sequence in the database, which has been identified as problematic for methods that assign taxonomy based on the closest match in the reference database (Ratnasingham and Hebert, 2007). We did not include representatives of other genera in our database because there is strong support for the monophyly of the genus (Neel and Cummings, 2004; Pettengill and Neel, 2008).

Previous study of the phylogenetic relationships among the sampled individuals showed that *Agalinis decemloba*, *A. tenella*, and *A. acuta* were polyphyletic as were *A. paupercula* and *A. purpurea* (Pettengill and Neel, 2008) and synonymization may be warranted. Due to the potentially long evolutionary time required for the property of reciprocal monophyly to arise (Hudson and Coyne, 2002; Rosenberg, 2003), taxonomic statements cannot be based solely on genealogical exclusivity. We are currently investigating whether in fact synonymization is warranted among *A. decemloba*, *A. tenella*, and *A. acuta* based on phylogenetic, population genetic, and morphological analyses. One objective of the present study is to evaluate the efficacy of DNA barcoding in light of taxonomic uncertainty by considering the performance of candidate barcodes using two taxonomic hypotheses. The first taxonomic alignment, which we refer to as the "historical taxonomy," treated each putative species individually and in the second *A. decemloba*, *A. tenella*, and *A. acuta* were synonymized under *A. decemloba*. A second pair of species, *A. paupercula* and *A. purpurea*, have also been considered by some authors to actually represent a single species and were synonymized under *A. purpurea*. Synonymizing yielded a data set of 26 species with between 1 and 23 individuals per species.

DNA extraction, PCR, and sequencing—Total genomic DNA was isolated from fresh or frozen (−80°C) leaves and flower buds by grinding 50–100 mg of tissue to powder in liquid nitrogen with a mortar and pestle, and then using GenElute Plant Genomic DNA Kits (Sigma, St. Louis, Missouri, USA) or Qiagen DNEasy Kits (Qiagen, Valencia, California, USA) following manufacturer's instructions.

We attempted to analyze sequences for all 92 individuals from six chloroplast gene regions (four coding regions: *matK* [Cuénoud et al., 2002] *rbcL*, [Kress and Erickson, 2007], *rps2* [dePamphilis et al., 1997], and *rpoB* (http://www.kew.org/barcoding/update.html), and the noncoding *psbA-trnH* [Sang et al., 1997] and *trnT-trnF* spacer regions [Taberlet et al., 1991]) that have been recommended for barcoding in plants. Two PCR reactions using primer pairs *trnT-a/trnL-d* and *trnL-c/trnF-f* were required to obtain the full length of the

trnT-trnF locus. Given that it is best to use a single primer pair to acquire a barcode, we divided the *trnT-trnF* locus into three separate regions that represent candidate barcodes evaluated by other studies. These three regions are defined by the primers used to amplify them (Taberlet et al., 1991): (1) *trnT*-a/*trnL*-d that we refer to as the *trnT-trnL* barcode (evaluated by Edwards et al., 2008), (2) *trnL*-c/*trnL*-d that we refer to as *trnL* intron (suggested by Taberlet et al., 2007), and (3) *trnL*-c/*trnF*-f, which we refer to as the *trnL-trnF* barcode (evaluated by Chase et al., 2005). This treatment of *trnT-trnF* as three separate regions brings our total assessment to eight single-locus barcodes.

All polymerase chain reactions (PCR) were done with Eppendorf MasterTaq PCR kits (Brinkman, Westbury, New York, USA) on MJ Research (Waltham, Massachusetts) PTC-200 Thermal Cyclers. In general, the PCR temperature profile was 30 cycles of 94°C for 60 s, annealing temperature set approximately 5°C below the lower of the two primer melting temperatures for 90 s, 72°C for 150 s, and a final 15 min elongation period at 72°C. Amplified DNA fragments were visualized by agarose gel electrophoresis and purified using the QIAQuick PCR Purification Kit according to manufacturer's instructions (Qiagen, Valencia, California, USA).

We employed a conservative sequencing strategy in which four replicates of both forward and reverse sequence reads were obtained per sample, resulting in 8-fold coverage across the majority of all loci. Although this strategy is not concordant with the rapid and inexpensive acquisition of sequences promoted as an advantage of DNA barcoding, we feel it is a necessary step to ensure accuracy of nucleotide sequences when considering closely related taxa. Sequencing reactions were conducted with BigDye Terminator v3.1 Cycle Sequencing chemistry (Applied Biosystems, Foster City, California, USA) and products were sequenced using an Applied Biosystems 3730xl DNA Analyzer. Total reaction volume was 7 μL (1–3 μL PCR template, 1.5 μL 5× Sequencing Buffer, 1 μL primer [2.5 μM], 0.5 μL BigDye Terminator, and 1–3 μL ddH_2O).

The program Sequencher v4.6 (Gene Codes Corp., Ann Arbor, Michigan, USA) was used for base calling, quality assignments, and assembling consensus sequences for each sample from the replicate bidirectional sequence reads. Sequences were aligned using the default settings within MUSCLE (Edgar, 2004). We manually inspected alignments (available in the TreeBase database, http://treebase.org; study ID no. 10343), to ensure that there were no errors and, in general, avoided manually editing sequences due to the subjectivity and nonreproducibility of such actions (Morrison, 2009). Furthermore, need for extensive manual editing of a nucleotide sequence alignment is not a desirable property of a DNA barcoding locus (CBOL Plant Working Group, 2009).

Barcoding analyses—In addition to evaluating the performance of the eight single-locus barcodes, we tested five, two-barcode combinations that have either been suggested (*matK* and *psbA-trnH*, Chase et al., 2005; *rbcL* and *psbA-trnH*, Kress and Erickson, 2007; *matK* and *rbcL*, CBOL Plant Working Group, 2009), have potential given the objectives of a multilocus barcode (i.e., the relatively slowly evolving *rpoB* and fast evolving *psbA-trnH*), or represent a combination of the best performing loci in this study (i.e., *trnT-trn*L and *psbA-trnH*). The multilocus barcodes were constructed by concatenating sequences for only those samples that had sequences for both loci, which sometimes yielded different sample sizes between each single locus barcode and the two-locus combinations. To avoid confounding issues due to differences in DNA sequence alignments, we constructed multilocus barcodes using the same alignment associated with the corresponding single-locus barcodes.

We calculated variability of each of the 13 barcodes as the number and percentage of variable characters using the program PAUP* v4b10 (Swofford, 2003). We assessed the presence of a gap between interspecific and intraspecific distances by constructing a histogram based on Kimura 2-parameter distances (K2P; Kimura, 1980) obtained from the program TAXONDNA (Meier et al., 2006). We further evaluated the presence of a gap across species and barcodes by subtracting the maximum intraspecific from the minimum interspecific K2P distance obtained from PAUP* v4b1.0 for each species following Fazekas et al. (2009). We then qualitatively compared the magnitude of the differences between within-species and among-species distances in *Agalinis* to those known from other genera. Due to differential amplification success, barcodes differed in the number of species that had at least two conspecifics and thus for which a barcoding gap could be calculated. We present evidence for barcoding gaps for three candidate barcodes: the best and worst performing barcodes and that advocated by CBOL Plant Working Group (2009) to illustrate the range of divergence among the *Agalinis* species.

The software package TAXONDNA (Meier et al., 2006) was used to assess performance of each barcode based on the "best close match" option using K2P distances. This method reduces the potential for errors of commission by employing a user-specified pairwise distance threshold above which a query sequence is classified as "no match". We classified query sequences as "ambiguous" if they could be assigned to both the correct and an incorrect species; as "correct" when pairwise genetic distances between query and reference sequences were below the threshold only for conspecifics; and as "incorrect" when only heterospecific sequences had pairwise distances to the query sequence that were below the threshold. Performance of each barcode is presented as the percentage of the total number of samples for that barcode that fell into each of these categories.

We evaluated the success of each barcode at three different genetic distance thresholds (2%, 5%, and a threshold calculated from the observed levels of intra- and interspecific divergence for each barcode). We chose the first two values to test the performance of standard thresholds that have been proposed as suitable for assigning sequences to species for animals (e.g., Hebert et al., 2003; Blaxter, 2004). The third threshold was calculated separately for each barcode to represent the pairwise genetic distance at which 95% of all conspecific individuals were correctly classified. Using information contained in the reference sequence data to establish a threshold is appealing because it accounts for the differences in mutation rate among the various loci and divergence among taxa. However, application of a calculated threshold requires multiple accessions of most species be present in the reference database.

We evaluated the success of both the "liberal" and "strict" tree-based methods of assigning sequences to species as described by Ross et al. (2008). With the liberal method, a sequence is assigned to a species if it is either sister to or embedded within a group. Thus, when A is the query sequence and there are two representatives of species Z in the reference database, A is assigned to species Z if it is either ((Z, A), Z) or ((Z, Z), A). The "strict" tree-based method, which is capable of accounting for sequences that have no conspecific sequence in the database, only assigns a sequence to a species if it is embedded within a group [e.g., only when ((Z, A), Z))]. Trees were constructed for each barcode using PAUP* v4b10 based on the neighbor-joining (NJ) algorithm (Saitou and Nei, 1987) applied to K2P distances. Success was determined by how each sequence would be classified if it were the query sequence and all other sequences were present in the reference database. Results are presented as the percentage of total sequences that were assigned to the categories correct and incorrect for the liberal method and as correct, incorrect, or ambiguous for the strict method.

Diagnostic nucleotide polymorphisms provide a character-based approach that potentially captures important differences that are not identified by distance- or tree-based barcoding methods (DeSalle et al., 2005). We used the program SITES (Hey and Wakeley, 1997) to calculate the number of nucleotide sites that were fixed between species pairs. However, to determine how well samples of unknown membership could be correctly assigned using diagnostic characters we used the character attribute organization system (CAOS) (Sarkar et al., 2002, b,, 2008). The CAOS method involves two steps that are accomplished by the programs P-Gnome and P-Elf. A reduced data set that included one sample for each species (i.e., 29 samples representing 29 species) was used to infer a phylogenetic tree for each barcode using two different inference methods: NJ based on K2P distances accomplished using PAUP*v4b10 and maximum-likelihood as implemented in the program GARLI v0.951 (Genetic Algorithm for Rapid Likelihood Inference; Zwickl, 2006) using the default parameter settings. The data matrix and associated tree file were then imported into the program MacClade where, based on previous phylogenetic work (Pettengill and Neel, 2008), *A. calycina* was placed as sister to the other members of the genus. The resulting NEXUS file was used in P-Gnome to generate rules for diagnosing the species where the tree associated with the data matrix served as a guide tree against which the presence and absence of nucleotides at a given node were evaluated to determine whether they were diagnostic (i.e., characteristic attributes) of the taxa subtending that node. Samples for the guide tree were arbitrarily chosen based on the first sequence for each species when sorted numerically by sample code. The same samples were used across all barcodes to prevent confounding our interpretations of barcode performance due to inclusion of different samples among guide trees. The program P-Elf was then used to assign sequences not included in the original file to species based on the rules (i.e., diagnostic character sets) generated by P-Gnome.

RESULTS

Ease of amplification—The first step in evaluating the utility of candidate barcodes was to consider amplification ease and reliability across all 92 individuals. In this regard, *rbcL* followed by *matK* were the worst-performing barcodes with bidirectional sequence reads being obtained for 85 and 86 individuals,

respectively. All but two other barcodes (*rps2*, $N = 90$; and *rpoB*, $N = 91$) amplified across all individuals (Table 2).

Sequences were obtained from at least one representative of all 29 species for all loci, and sequences for each locus were obtained from at least 92% of the samples (Table 2 and Appendix 1). Across all loci, 720 of a total possible 736 sequences were obtained (98%), and most samples required only a single attempt at PCR and sequencing to obtain a high quality sequence. In addition to the complete failures mentioned, the locus *rps2* required the alternative forward primer *rps2*-47 instead of *rps2*-18F for 10 individuals (dePamphilis et al., 1997). Both *matK* and *rbcL* often required multiple attempts at PCR and sequencing to obtain a high quality sequence from samples.

Barcode variation—The percentage of variable characters ranged from 7.71% (*rbcL*) to 31.52% (*psbA-trnH*), (Table 2). The number of variable sites for individual loci ranged from 55 (*rpoB*) to 290 (*trnT-trnL*) (Table 2). Loci also varied in the range of intraspecific and interspecific distances within and among species. The most conserved loci were *rbcL* and the *trnL*-intron; the percentage differentiation based on pairwise K2P distances among *Agalinis* species ranged from 0.0–3.9% for both loci and averaged 1.02% for *rbcL* and 1.79 for the *trnL* intron (Fig. 1 shows the distribution of distances for *rbcL*). Within-species percentage differences for these same loci ranged from 0–1.6% (average 0.2%) for *rbcL* and from 0–0.8% (average 0. 04%) for the *trnL* intron. The *psbA-trnH* locus had the largest range of among-species K2P distances (0.0–15.35%) and averaged 6.95% (Fig. 1). The average and range of among species K2P distances for the combined *rbcL/matK* barcode were 1.92% and 0.0–5.1%, respectively (Fig. 1). After synonymization, the average pairwise distances within most species increased with changes ranging from 0.5–1.04% depending on the locus examined and average distances among species increased for all loci by 3.3–4.5%. Large pairwise genetic distances among *Agalinis* species were also observed in Neel and Cummings (2004) (i.e., maximum-likelihood estimates of pairwise distances were on the order of 0.12). The maximum estimate of pairwise K2P distances calculated from the samples of *Orobanche* species investigated with *rbcL* (Manen et al., 2004) is 11.5%, which is also similar to that which we observed.

The proportion of species exhibiting a barcoding gap varied by locus, with *rbcL* exhibiting the fewest number of gaps of all barcodes (Table 2; Appendices S1 and S2, see Supplemental Data at http://www.amjbot.org/cgi/content/full/ajb.0900176/DC1); for this locus, only 12 of the 27 species for which multiple samples were included had a larger minimum interspecific genetic distance than maximum intraspecific distance (Table 2, Fig. 2). Combining *rbcL* with *matK* increased the number of species that exhibited a barcoding gap from 12 to 21, relative to when *rbcL* was evaluated independently (Fig. 2). In contrast, for *psbA-trnH*, 22 of 27 species possessed a barcoding gap (Table 2, Fig. 2).

Performance of candidate barcodes—Success of each candidate barcode using the "best close match" distance-based method differed depending on the threshold value (Fig. 3). In all cases, the calculated threshold was substantially less than 2%, varying from 0.25% (*matK*) to 1.01% (*rbcL*) under the historical taxonomy and from 0.15% (*matK*) to 1.01% (*rbcL*) when species were synonymized (Table 3). Regardless of the taxonomic hypothesis considered, as the threshold increased from the calculated distance to 2% both correct and incorrect classifications increased by a few percent with a corresponding decrease in "no match" classifications (Fig. 3). For *matK*, *trnT-trnL*, and the multilocus barcodes, we saw this pattern repeated when comparing performance between the 2% and 5% thresholds. Ambiguous classifications remained fairly constant across the three thresholds (Fig. 3). Below we discuss the performance of the candidate barcodes considering only the calculated distance threshold because it generally had the best results of all the thresholds.

Of the single-locus barcodes, *trnL-trnF* had the highest correct classification rate and tied for the lowest incorrect classification rate with *trnT-trnL* and the *trnL* intron under the historical taxonomy; the *rbcL* region had the lowest correct classification rate and the highest percentage of incorrect classifications (Table 3). Although the percentage incorrect classification rates were relatively low for all single-locus barcodes (1.08–7.14%), there were between 26.37% (*psbA-trnH*) and 40.47% (*rbcL*) ambiguous assignments using this method (Table 3).

Classification success based on the multilocus barcodes was only marginally different (e.g., generally less than a few percentage points or one or two individuals) than when the loci were considered individually (Table 3). For example, the *matK/psbA-trnH* barcode had a higher percentage of correct classification (62.79%) than when *matK* was used by itself

TABLE 2. Summary characteristics of the 13 DNA barcodes evaluated in this study. Presence of a barcoding gap is based on the proportion of species for which there were multiple samples (N) that had a difference between the minimum interspecific and maximum intraspecific genetic distance greater than zero.

Locus	N individuals (N species)	Aligned length (bp)	Variable characters (%)	Presence of barcoding gap	
				Nonsynonymized	Synonymized
matK	86 (29)	899	168 (18.69%)	0.76 (25)	0.95 (22)
rbcL	85 (29)	1012	78 (7.71%)	0.44 (27)	0.54 (24)
rpoB	91 (29)	368	55 (14.95%)	0.62 (26)	0.78 (23)
rps2	90 (29)	660	146 (22.12%)	0.70 (27)	0.88 (24)
trnT-trnL	92 (29)	1397	290 (20.76%)	0.81 (27)	0.88 (24)
trnL-intron	92 (29)	548	84 (15.32%)	0.70 (27)	0.88 (24)
trnL-trnF	92 (29)	965	182 (18.86%)	0.67 (27)	0.83 (24)
psbA- trnH	92 (29)	809	255 (31.52%)	0.81 (27)	0.96 (24)
rpoB/psbA-trnH	91 (29)	1177	310 (26.33%)	0.77 (26)	0.96 (23)
rbcL/psbA-trnH	85 (29)	1821	333 (18.28%)	0.78 (27)	0.96 (24)
matK/psbA-trnH	86 (29)	1708	422 (24.71%)	0.76 (25)	0.95 (22)
matK/rbcL	82 (29)	1911	246 (12.88%)	0.76 (25)	0.95 (22)
trnT-trnL/psbA-trnH	92 (29)	2206	545 (24.71%)	0.74 (27)	0.92 (24)

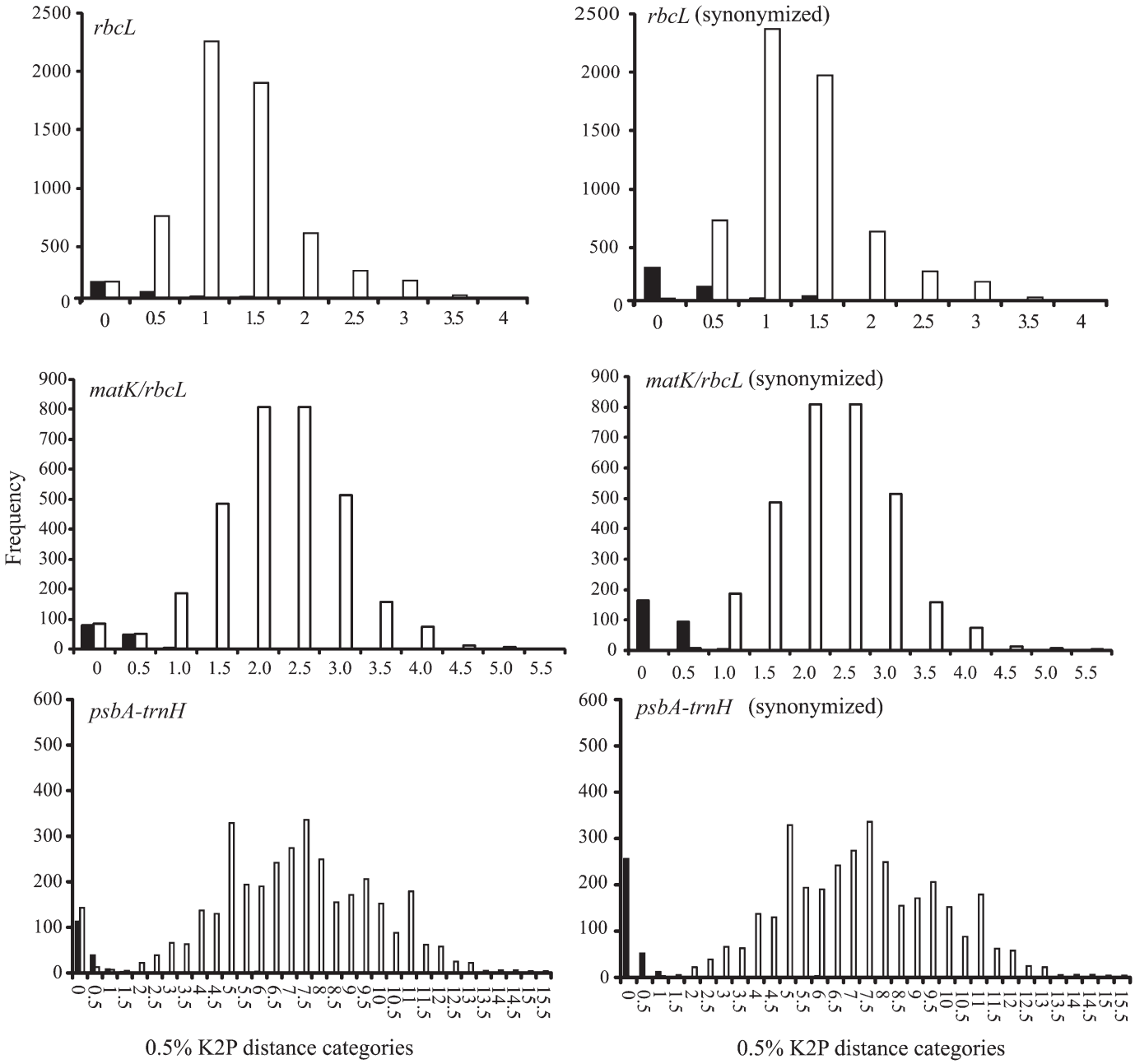

Fig. 1. Histograms of the number of intraspecific and interspecific comparisons that fell into 0.5% K2P distance categories for three candidate barcodes under two taxonomic hypotheses. Black and white bars represent intraspecific and interspecific pairwise comparisons, respectively.

(61.62%), but performance was worse than when *psbA-trnH* was used alone (64.83%); *trnT-trnL/psbA-trnH* performed better (63.04% correct classification) than when just *trnT-trnL* (60.86%) was employed but worse when only *psbA-trnH* was used (64.83%) (Table 3). Discrepancies among classification success of multilocus barcodes and the individual loci that comprise them are due to a combination of factors including differences in samples sizes, presence or absence of a conspecific (e.g., a species with two representatives may result in a correct match, but when there is no conspecific, the result may be a no match), and samples being classified differently under the different barcodes (e.g., distances to conspecifics of *A. skinneriana* 90MO are below the threshold for *psbA-trnH* but above the threshold for both *matK* and *matK/psbA-trnH*).

Synonymizing species increased the average correct classification across all candidate barcodes from 62.04 to 89.75% and correct classification was >90% for four of the single-locus barcodes (Table 3). *psbA-trnH* had the highest correct classification (94.5%) and no misclassification; *rbcL* had the lowest correct classification (79.76%). For *trnL-trnF,* there was an increase in the number of query sequences determined to have no match relative to the historical taxonomic alignment (Table 3). The most likely reason for this increase in the number of no matches is that, upon synonymizing samples with nearly identical

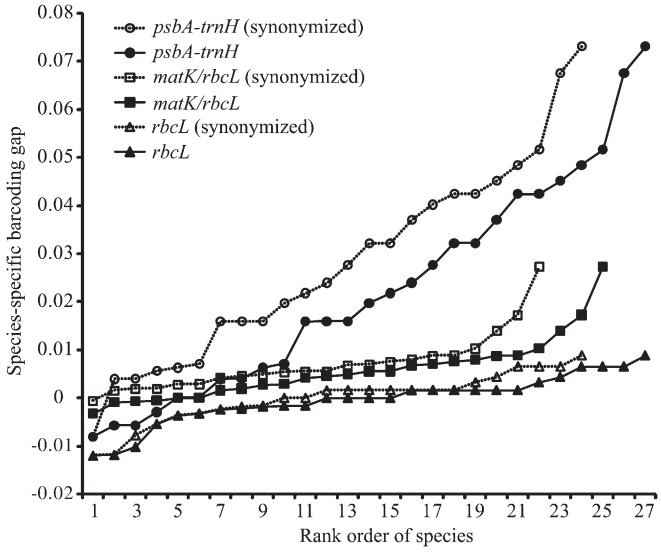

Fig. 2. Species-specific barcoding gaps calculated as the difference between the minimum interspecific and intraspecific genetic distances (K2P) for three candidate barcodes under the historical taxonomy and after synonymization based on the results of Pettengill and Neel (2008).

sequences (e.g., *A. acuta*, *A. decemloba*, and *A. tenella*), the threshold value associated with 95% correct classification was decreased such that more pairwise conspecific comparisons exceeded that threshold. The performance of the five multilocus barcodes varied after synonymizing species; only *rpoB*/*psbA-trnH* and *mat*K/*psbA-trnH* had no individuals misclassified (Table 3).

The two tree-based methods differed substantially in the degree to which they accurately handled query sequences. Under the strict method, every sequence that represented a species for which we had only sampled two individuals was classified as ambiguous, which was a result of treating one sample as a reference sequence and the other as a query sequence such that there is no way for a query sequence to be embedded among conspecifics (Table 4). The conservative nature of the strict method yielded an average 16.99% correct classification rate that increased to 42.95% when species were synonymized (Table 4). However, there were no incorrect classifications using the strict method because query sequences were never incorrectly embedded in a monospecific clade consisting of heterospecifics. Under the historical taxonomy, when samples of *A. acuta*, *A. decemloba*, *A. tenella*, *A. paupercula*, or *A. purpurea* were treated as query sequences, they were embedded within clades that contained both heterospecifics and conspecifics, and as a result, all assignments were considered ambiguous rather than incorrect. Even when species were synonymized, across all barcodes, two of the samples for species with at least three accessions were sister to rather than embedded within the synonymized group and were classified as ambiguous.

The liberal tree-based method had an average correct classification of 61.17% under the historical taxonomic framework, and 92.26% after synonymization (Table 4). Despite the inability of this method to identify sequences that do not have a conspecific in the database, it had a relatively low incorrect classification rate, which based on the synonymized taxonomic hypothesis ranged from 4.35% for *psbA-trnH* and *trnT-trnL*/*psbA-trnH* to 28.24% for *rbcL* (Table 4).

Under both tree-based methods, incorporating a second locus increased the correct classification rate relative to the performance when the more slowly evolving locus was used alone. For example, using the synonymized data set and the strict tree-based method, the correct classification rates for *rbcL* and *rpoB* were 38.82% and 34.07%, respectively; when they were combined with *psbA-trnH* correct classification was 41.18% and 45.05%, respectively (Table 4). The combination of *matK* and *psbA-trnH* did not substantially change the correct classification rate; the marginal decrease in correct classification rate of *matK*/*psbA-trnH* relative to *psbA-trnH* and *matK*/*rbcL* relative to *matK* are due to differences in which samples are present for a barcode as a result of amplification failure (Tables 2, 4).

We found diagnostic characters in all barcodes for the majority of putative *Agalinis* species pairs (online Appendix S3). The average number of fixed differences among species varied from 4.85 in *rbcL* to 30.74 in *trnT-trnL* for the single locus barcodes. The *matK*, *rps2*, *psbA-trnH*, and five multilocus barcodes had fixed nucleotide differences for 82.75% of the species; they all failed to distinguish five putative species (*A. acuta/A. decemloba/A. tenella* and *A. purpurea/A. paupercula*). The *rbcL* and *rpoB* barcodes discriminated the fewest species based on the presence of fixed nucleotides (58.62% and 72.41%, respectively). When species were synonymized, three single-locus and all five multilocus barcodes had fixed nucleotide differences that distinguished all 26 taxa, although the average number of fixed differences was generally lower than under the historical taxonomy.

The number of query sequences used to evaluate the CAOS method ranged from 56 (*rbcL*) to 63 (*psbA-trnH*, *trnT-trnL*, *trnL*-intron, and *trnL-trnF*) and always increased by three for the analyses in the synonymized data set (Table 5). Although fixed differences existed among the majority of species, percentage correct classification using the CAOS method was often substantially lower than the other methods evaluated (Tables 3–5), ranging from 17.74% for *rpoB* to 63.49% for *trnL-trnF* and *psbA-trnH* (Table 5). Misclassifications were due to both incorrect and ambiguous calls. The multilocus barcodes all performed worse than single-locus barcodes, primarily as a result of a decrease in ambiguous classifications and an increase in incorrect classifications relative to when the loci were used alone (Table 5). This result is surprising given that chloroplast loci do not assort independently of one another and, thus, should not be in conflict. Additionally, when two chloroplast loci are concatenated, they possess the combined diagnostic characters and should have more information than each single locus. In contrast to the marginal differences observed between single locus and multilocus barcodes under the other two assignment techniques, the discrepancies under the CAOS method cannot be explained solely by differences in sample inclusion, presence or absence of a conspecific, and samples being classified differently under the different barcodes; additional explanations are presented in the discussion.

Using the synonymized data set greatly increased the number of query sequences that could be correctly classified using the CAOS method with two barcodes (*psbA-trnH* and *trnL-trnF*) having correct classification rates >87% (Table 5). The multilocus barcodes still had classification success rates below those of the better performing single locus barcode. Because of differences between the two data sets in the number of samples used to create the guide tree (i.e., 26 and 29 corresponding to the two taxonomic alignments we investigated), the topologies and combinations of characters that were diagnostic also differed.

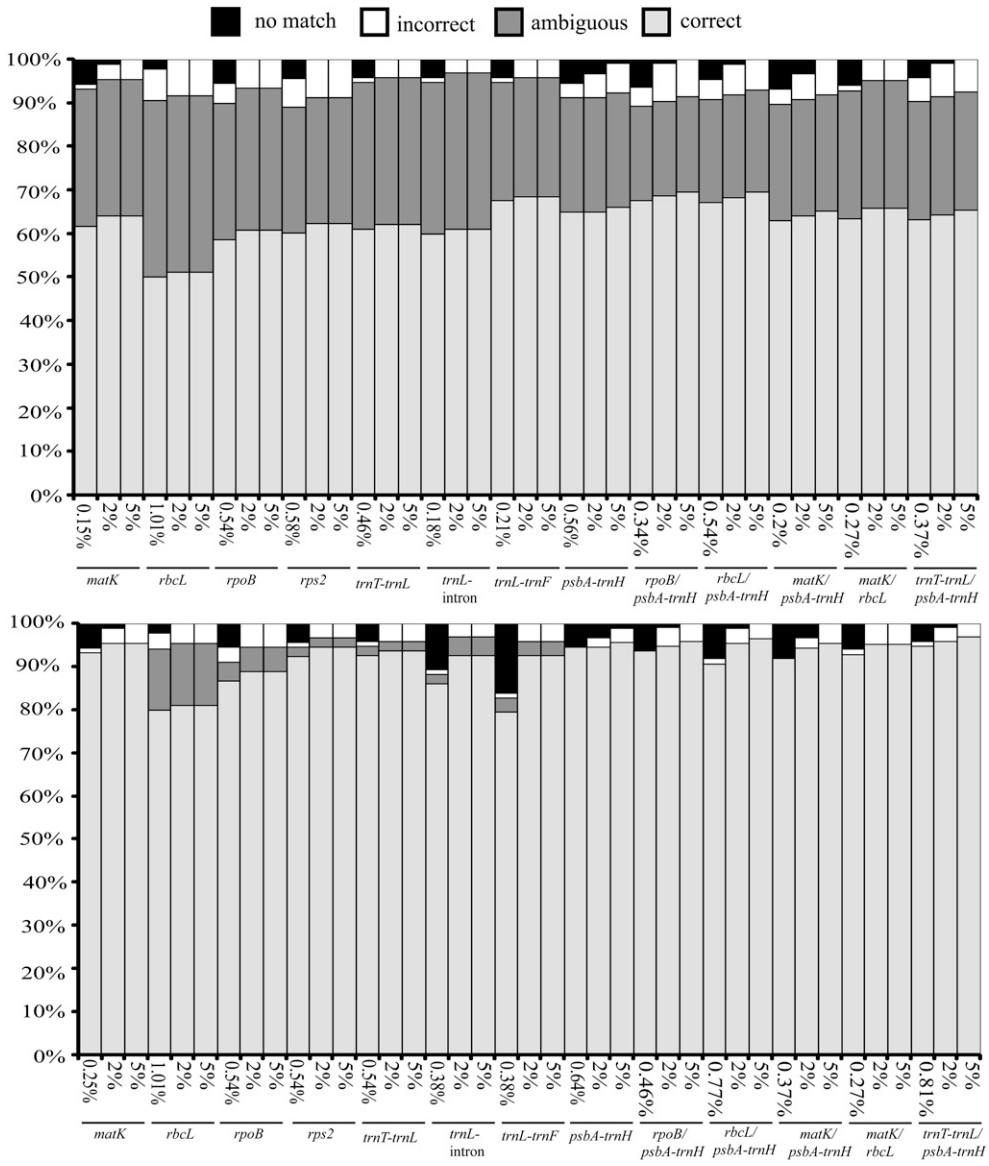

Fig. 3. Performance of 13 candidate barcodes across three different threshold values for the "best close match" distance-based method (Meier et al., 2006); the first threshold for each barcode represents the calculated value based on the samples below which 95% of sequences were correctly classified. Top panel represents classifications under the historical taxonomic alignment and the bottom panel represents the results after synonymization based on the results of Pettengill and Neel (2008).

This likely explains in part why the rank order performance of barcodes differs between the synonymized and nonsynonymized data set.

In addition to being generally low, barcoding success from CAOS depended on the method of inferring the phylogeny that served as the guide tree (NJ or maximum-likelihood) (Table 5); topologies based on the same barcode were often different between the two inference methods (online Appendix S4). Neither method was consistently better in that the percentage correct classification was higher for a tree inferred using the NJ algorithm for *rbcL*, *rpoB*, *trnL-trnF*, and the maximum-likelihood tree yielded better classification for *psbA-trnH*; the two methods did not differ for *rps*2, *trnT-trnL*, and *trnL*-intron. In general, the maximum likelihood tree had fewer ambiguous and more incorrect calls than the neighbor-joining tree.

Species-specific performances—No DNA barcode locus or analytical technique correctly classified all samples of *A. acuta*, *A. decemloba, A. tenella*, *A. paupercula, A. purpurea*, or *A. harperi*. When *A. acuta*, *A. decemloba,* and *A. tenella* were synonymized as *A. decemloba* and *A. paupercula* and *A. purpurea* were synonymized as *A. purpurea*, the distance- and liberal tree-based methods always classified all relevant sequences into the correct synonymized group. However, the CAOS method failed to correctly classify all of the samples for seven barcodes (*matK*/*psbA-trnH*, *matK*/*rbcL*, *rbcL*, *rbcL*/*psbA-trnH*, *trnT-trnL*, *trnT-trnL*/*psbA-trnH*, and *rpoB*). The two *A. harperi* samples were never correctly classified as conspecifics.

For the distance-based methods, although details varied depending on the DNA barcode, 10 samples were typically classified as "no match" because the distances among the conspecific

TABLE 3. Performance for the candidate barcodes when "best close match" distance method (Meier et al., 2006) is used to identify sample sequences (*N*). Threshold distances were calculated from the observed sequences for each locus. Numbers in parentheses indicate the performance of barcodes when species are synonymized based on the results of Pettengill and Neel (2008).

		Percentage of samples assigned to class				
Barcode	*N*	Correct	Ambiguous	Incorrect	No match	Threshold
matK	86	61.62 (93.02)	31.39 (0.0)	1.16 (1.16)	5.81 (5.81)	0.25 (0.15)
rbcL	85	50.0 (79.76)	40.47 (14.28)	7.14 (3.57)	2.38 (2.38)	1.01 (1.01)
rpoB	91	58.42 (86.51)	31.46 (4.49)	4.49 (3.37)	5.61 (5.61)	0.54 (0.54)
rps2	90	60.0 (92.22)	28.88 (2.22)	6.66 (1.11)	4.44 (4.44)	0.58 (0.58)
psbA-trnH	92	64.83 (94.5)	26.37 (0.0)	3.29 (0.0)	5.49 (5.49)	0.64 (0.56)
trnT-trnL	92	60.86 (92.39)	33.69 (2.17)	1.08 (1.08)	4.34 (4.34)	0.54 (0.46)
trnL-intron	92	59.78 (85.86)	34.78 (2.17)	1.08 (1.08)	4.34 (10.86)	0.38 (0.18)
trnL-trnF	92	67.39 (79.34)	27.17 (3.26)	1.08 (1.08)	4.34 (16.3)	0.38 (0.21)
rpoB/psbA-trnH	91	67.39 (93.47)	21.73 (0.0)	4.34 (0.0)	6.52 (6.52)	0.46 (0.34)
rbcL/psbA-trnH	85	67.05 (90.58)	23.52 (0.0)	4.7 (1.17)	4.7 (8.23)	0.77 (0.54)
matK/psbA-trnH	86	62.79 (91.86)	26.74 (0.0)	3.48 (0.0)	6.97 (8.13)	0.37 (0.2)
matK/rbcL	82	63.41 (92.68)	29.26 (0.0)	1.21 (1.21)	6.09 (6.09)	0.27 (0.27)
trnT-trnL/psbA-trnH	92	63.04 (94.5)	27.17 (0.0)	5.43 (1.08)	4.34 (4.34)	0.81 (0.37)
Average	88.92	62.04 (89.75)	29.43 (2.20)	3.47 (1.22)	5.03 (6.81)	0.54 (0.42)

individuals exceeded the designated threshold (e.g., *A. skinneriana* 90MO, the two *A. linifolia* samples, *A. oligophylla* 12AL and 5AL, *A. heterophylla* TX and 5TX, *A. pulchella* 3GA, *A. fasciculata* 1LA, and *A. setacea* 76VA). The two species for which we only had a single representative (i.e., *A. calycina* and *A. edwardsiana*) were correctly treated as having no conspecific match in the database when using *trnL-trnF*, *psbA-trnH*, *rpoB*/*psbA-trnH*, and *matK*/*psbA-trnH*. However, for the other barcodes (i.e., *rps2*, *matK*, *rbcL*, *rpoB*, *trnL* intron, *trnT-trnL*, and *rbcL*/*psbA-trnH*) classifications of *A. calycina* and *A. edwardsiana* were incorrect or ambiguous due to insufficient differentiation from heterospecific samples in the database.

Species-specific misclassification was high for the strict tree-based method where the samples from all but one species (i.e., *A. harperi*) that were represented by two individuals were classified as ambiguous. For the liberal tree-based method, the only incorrect classifications using the best performing locus (i.e., *psbA-trnH*) involved query sequences that did not have conspecifics in the database (i.e., *A. calycina* and *A. edwardsiana*) and the two polyphyletic *A. harperi* samples. The higher incorrect classification rates associated with other loci (e.g., *rbcL* and *rpoB*) were due to lack of sufficient nucleotide variation to depict accurately the relationships among the samples.

Under the diagnostic character-based method of CAOS, species-specific classifications depended on the inference method used to create the guide tree. For example, using the *psbA-trnH* barcode and the NJ algorithm, seven samples were misclassified (*A. fasciculata* 2GA, *A. fasciculata* 4LA, *A. harperi* 12NC, *A. setacea* 7MD, and *A. tenuifolia* 5IA were classified as *A. skinneriana*; *A. setacea* 3VA was classified as *A. tenuifolia*; and *A. tenuifolia* 2VA was classified as *A. aphylla*), and for the same barcode, but using a guide tree inferred using a maximum likelihood-based method, there were only four incorrect classifications (*A. setacea* 7MD, *A. setacea* 3VA, and *A. tenuifolia* 5IA were classified as *A. skinneriana* and *A. tenuifolia* 2VA was classified as *A. aphylla*).

DISCUSSION

The need to catalog the earth's rapidly declining biodiversity makes the promise of DNA barcoding highly appealing (Savolainen et al., 2005). In addition to facilitating basic inventory

TABLE 4. Performance, in terms of percentage of query sequences (*N*), of the candidate barcodes based on the strict and liberal tree-based methods of Ross et al. (2008). Numbers in parentheses indicate the performance of barcodes when species are synonymized based on the results of Pettengill and Neel (2008).

		Strict method		Liberal method	
Barcode	*N*	Correct	Ambiguous	Correct	Incorrect
matK	86	17.44 (44.19)	82.56 (55.81)	63.95 (95.35)	36.05 (4.65)
rbcL	85	12.94 (38.82)	87.06 (61.18)	52.94 (71.76)	47.06 (28.24)
rpoB	91	18.68 (34.07)	81.32 (65.93)	57.14 (89.01)	42.86 (10.99)
rps2	90	16.67 (45.56)	83.33 (54.44)	55.56 (95.56)	44.44 (4.44)
psbA-trnH	92	18.48 (45.65)	81.52 (54.35)	64.13 (95.65)	35.87 (4.35)
trnT-trnL	92	17.39 (44.57)	82.61 (55.43)	60.87 (92.39)	39.13 (7.61)
trnL-intron	92	17.39 (44.57)	82.61 (55.43)	60.87 (92.39)	39.13 (7.61)
trnL-trnF	92	17.39 (44.57)	82.61 (55.43)	60.87 (92.39)	39.13 (7.61)
rpoB/psbA-trnH	91	18.68 (45.05)	81.32 (54.95)	62.64 (93.41)	37.36 (6.59)
rbcL/psbA-trnH	85	15.29 (41.18)	84.71 (58.82)	64.71 (95.29)	35.29 (4.71)
matK/psbA-trnH	86	17.44 (44.19)	82.56 (55.81)	63.95 (95.35)	36.05 (4.65)
matK/rbcL	82	14.63 (40.24)	85.37 (59.76)	63.41 (95.12)	36.59 (4.88)
trnT-trnL/psbA-trnH	92	18.48 (45.65)	81.52 (54.35)	64.13 (95.65)	35.87 (4.35)
Average	88.92	16.99 (42.95)	83.01 (57.05)	61.17 (92.26)	38.83 (7.74)

TABLE 5. Performance of the candidate barcodes in terms of percentage of query sequences (*N*) classified into each category based on the diagnostic character-based method CAOS (Sarkar et al., 2008) when the guide tree is inferred with the neighbor-joining algorithm or a maximum-likelihood based method. Numbers in parentheses indicate the number of query sequences and performance of barcodes when species are synonymized based on the results of Pettengill and Neel (2008).

		Neighbor-joining			Maximum-likelihood		
Barcode	*N*	Correct	Ambiguous	Incorrect	Correct	Ambiguous	Incorrect
matK	57 (60)	29.82 (68.33)	43.86 (6.67)	26.32 (25.00)	36.84 (68.33)	10.53 (5.00)	52.63 (26.67)
rbcL	56 (59)	30.36 (27.12)	25.00 (30.51)	44.64 (42.37)	25.00 (28.81)	28.57 (30.51)	46.43 (40.68)
rpoB	62 (65)	22.58 (53.85)	59.68 (23.08)	17.74 (23.08)	17.74 (53.85)	58.06 (24.62)	24.19 (21.54)
rps2	61 (64)	44.26 (84.38)	45.90 (12.50)	9.84 (3.13)	54.10 (82.81)	22.95 (10.94)	22.95 (6.25)
psbA-trnH	63 (66)	58.73 (87.88)	30.16 (1.52)	11.11 (10.39)	63.49 (87.88)	30.16 (1.52)	6.35 (10.39)
trnT-trnL	63 (66)	55.56 (83.33)	12.70 (0.0)	31.75 (16.67)	55.56 (84.85)	14.29 (0.00)	30.16 (15.159)
trnL-intron	63 (66)	46.03 (83.33)	50.79 (9.09)	3.17 (7.58)	46.03 (83.33)	50.79 (12.12)	3.17 (4.55)
trnL-trnF	63 (66)	63.49 (87.88)	7.94 (6.06)	28.57 (6.06)	61.90 (84.85)	7.94 (13.64)	30.16 (1.51)
rpoB/psbA-trnH	62 (65)	45.16 (70.77)	19.35 (1.54)	35.48 (27.69)	41.94 (75.38)	9.68 (1.54)	48.39 (23.08)
rbcL/psbA-trnH	56 (59)	53.57 (76.27)	8.93 (0.0)	37.50 (23.73)	58.93 (54.24)	8.93 (0.0)	32.14 (45.76)
matK/psbA-trnH	57 (60)	42.11 (68.33)	22.81 (0.0)	35.09 (31.67)	49.12 (68.33)	19.30 (0.0)	31.58 (31.67)
matK/rbcL	53 (56)	35.84 (67.86)	7.54 (3.57)	56.60 (28.57)	35.84 (71.43)	5.66 (3.57)	58.49 (25.00)
trnT-trnL/psbA-trnH	63 (66)	30.16 (54.55)	11.11 (0.0)	58.73 (45.45)	42.86 (56.07)	15.87 (0.0)	41.27 (43.93)
Average	59.92 (62.92)	42.90 (70.30)	26.60 (7.27)	30.50 (22.43)	45.33 (69.24)	21.75 (7.96)	32.92 (22.80)

of poorly studied areas, barcoding has been proposed as an effective means of delineating and monitoring species distributions (DeSalle et al., 2005). Sequence databases created as a result of DNA barcoding have also been advocated as being useful for understanding the evolutionary diversity within rare species (Faith and Baker, 2006). Whether DNA barcoding can be used for these purposes rests on the ability to distinguish among closely related species including sister taxa. This study is among the few that have been conducted to evaluate the efficacy of DNA barcoding among congeneric species (e.g., Newmaster and Ragupathy, 2009; Starr et al., 2009). Although limited in taxonomic breadth, the range of evolutionary distances within and among the 92 samples representing 29 *Agalinis* species reflect patterns in many angiosperm genera and in certain instances even represent levels of differentiation found among members from different genera (e.g., Kress et al., 2005; Lahaye et al., 2008; Fazekas et al., 2009). The large distances among *Agalinis* species we observed may be characteristic of members of the Orobanchaceae in general within which accelerated rates of chloroplast evolution (e.g., *rps2* and *rbcL*) have been documented (dePamphilis et al., 1997). That species within the Orobanchaceae are to some degree parasitic and, therefore, not as dependent on photosynthetic proteins has been hypothesized to result in higher rates of evolution within the chloroplast (Wolfe et al., 1992). Despite the large distances we observed among congeneric species, the mean barcoding gap for different barcodes spanned the range of magnitudes observed in other genera (Fig. 4; Fazekas et al., 2009). Consequently, the results from our analysis of 29 species using three analytical techniques that clearly show barcoding has potential for distinguishing among congeners using single barcodes inform how DNA barcoding may perform among closely related species in other genera.

The ability to distinguish closely related species means that barcoding can be useful in conservation biology to confirm identities in taxonomically challenging groups that are difficult to identify in the field. Such information can serve to improve understanding of species distributions. A specific example from our study is the case of the state rare *A. skinneriana*, whose geographic range was thought to not include Mid-Atlantic coastal states. When samples from Maryland were first collected they were tentatively attributed to *A. skinneriana*, but local botanists had concerns because no other populations of this species occur within 400–500 km; barcoding unambiguously confirmed the taxonomic identity of those samples. We were also able to confirm identifications from collection locations from which individuals had originally been identified as *A. decemloba* but later annotated as *A. setacea*. However, as discussed below, successful application of DNA barcoding depends on the barcode locus, the analytical method, and the underlying taxonomy used.

Candidate barcodes—Of the loci we tested, *psbA-trnH* and *trnT-trnL* most closely met the requirements of a suitable DNA barcode identified by the Consortium for the Barcode of Life in that those loci were routinely retrievable with a single primer pair, easy to obtain bidirectional sequence reads, required little to no manual editing of sequence traces, and provided maximal discrimination among species (CBOL Plant Working Group, 2009) (Tables 2–5, Figs. 1–3). When evaluating the data set in which polyphyletic taxa had been synonymized, these two loci had a greater than 83% correct classification rate across all methods except with the strict tree-based method. Under the historical taxonomic framework, accuracy was closer to 50–70%, mostly due to ambiguous classifications among the species that were synonymized in the other data set. Although they

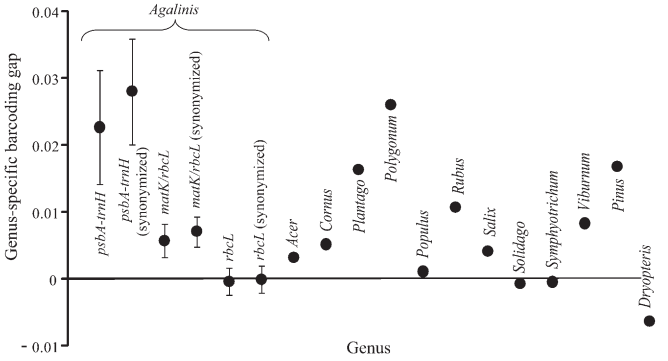

Fig. 4. Mean (±95% CI) for three candidate barcodes under two taxonomic alignments of *Agalinis* and mean values for 12 other plant genera from Fazekas et al. (2009), which are based on the concatenation of 7 plastid loci.

worked well for *Agalinis* and have been advocated by others (Kress and Erickson, 2007), these loci are known to have alignment issues (*psbA-trnH*; Lahaye et al., 2008) or suffer from a lack of significant interspecific differentiation (*trn*T-*trnL*; Edwards et al., 2008) in some taxa.

We found no substantial benefit to using multilocus barcodes; correct classification sometimes decreased or only increased by a few percentage points and an additional locus rarely discriminated among samples that were not already correctly classified using the better performing of the two loci alone (Tables 3–5). Therefore, the additional work and cost necessary to acquire a second locus does not seem warranted. Although one could argue that the closely related species we investigated do not represent the situation under which a multilocus barcode may be useful, the benefits have also been shown to be negligible with more evolutionarily divergent samples (e.g., Lahaye et al., 2008). However, multilocus barcodes have yielded improved performance in some studies (e.g., Fazekas et al., 2008; CBOL Plant Working Group, 2009; Hollingsworth et al., 2009).

We found that the plant barcode advocated by CBOL (i.e., the combination of *rbcL*/*matK*) correctly classified >92% *Agalinis* species under the synonymized taxonomy; Tables 3, 4); that barcode correctly classified only 72% of samples in the CBOL Plant Working Group (2009) study. The level of success of *rbcL*/*matK* observed by CBOL is similar to the expected 60–70% level of unambiguous correct classification that was suggested by Erickson et al. (2008) and also similar to the 71% success described in Fazekas et al. (2008). Despite the relatively poor performance of *rbcL*/*matK* in assigning samples to the correct species (i.e., 72%), that barcode did have a 100% correct classification to either the correct species or species group (i.e., the 28% of samples that were misclassified in the CBOL study were matched to groups of congeners) (CBOL Plant Working Group, 2009). Our results provide encouraging evidence that even among closely related species (i.e., congeners of *Agalinis*), there is high probability of correct classification using the *rbcL*/*matK* barcode.

However, a major drawback of the two-locus *rbcL*/*matK* system is that both sequences could only be obtained from ~89% (82 of 92) samples (Table 2). In practice, our results suggest that ~11% of samples will not be able to be identified using that barcode due to amplification failure. Others have also noted the difficulty in isolating *rbcL* and/or *matK* (e.g., Sass et al., 2007; Fazekas et al., 2008; Lahaye et al., 2008), indicating that our inability to isolate those two loci is not unique to *Agalinis* but rather will likely affect DNA barcoding in plant genera more broadly. Even CBOL could only acquire *matK* from <85% of gymnosperm and <20% of cryptogam species (CBOL Plant Working Group, 2009), which we consider unacceptably low amplification success for a universal barcode.

Differential success of assignment methods—In real barcoding applications, samples of unknown identity that are compared with reference sequences will either be unambiguously assigned to one species, assigned to multiple species, or depending on the methods used, assigned as having "no match". Some percentage of unambiguous, apparently correct assignments will in fact be incorrect (i.e., will be false positives). We argue that understanding the potential for false positives resulting from unambiguous assignment to the wrong species is more critical for evaluating the utility of barcoding than knowing the potential for correct identifications. A means of estimating this potential is to assume that it will be similar to the percentage of incorrect matches among samples of known taxonomic identity (i.e., those in the reference database). Using distance methods, our percentage of false positives was 1.25% when species were synonymized and 3.5% under the historical taxonomy (Table 3), which is comparable to levels found by Newmaster et al. (2008) in differentiating among congeners of *Compsonuera* (Myristicaceae). The low levels of correct classification under the historical taxonomy were due to ambiguous classifications because sequences were assigned to multiple species; a few samples were also classified as "no match" because they were sufficiently distant from all species. Methods that provide ambiguous and "no match" classifications are essential because they disclose patterns that would be masked if query sequences were forced into only species that exist in the database and thus highlight the need for further research to understand the cause of the ambiguities.

Although some loci were generally better than others, performance of a given barcode was highly dependent on the analytical method being used. The "best close match" distance-based method as implemented in TAXONDNA (Meier et al., 2006) yielded the best overall results across most loci. In addition to high correct classification rates, most of the samples that were not classified correctly were "ambiguous" rather than incorrect (Table 3). This method also can reduce errors of commission by identifying sequences as having no match rather than forcing incorrect classification as the closest species. Although there were cases in which samples were erroneously identified as having no conspecific match in the database, these misclassifications represent an exception to the behavior of the majority of conspecifics we examined. Such cases yield opportunities for further examination to determine if they resulted from misidentification, contamination, or something biologically interesting such as phylogeographic structure or presence of cryptic species.

The primary criticism of distance-based methods is that no single distance threshold delineates all species (Ferguson, 2002; DeSalle et al., 2005; Little and Stevenson, 2007). Although a value calculated from the data can help determine the most appropriate threshold (e.g., Meier et al., 2006), it still represents a single threshold that could be problematic when taxa with vastly different rates of chloroplast nucleotide substitution rates are present in the reference database (Fazekas et al., 2009). For example, when using a single calculated threshold, samples from taxa with relatively rapidly evolving chloroplast genomes (e.g., grasses) may be classified as having no match in the database and samples from taxa with relatively slowly evolving chloroplast genomes (e.g., palms) (Wilson et al., 1990) will be classified as ambiguous. Our calculated thresholds for *psbA-trnH* and *matK* in the synonymized data set were 0.56% and 0.15%, respectively, but within Newmaster et al. (2008) thresholds of 2.52% and 0.26%, respectively, were reported for these same loci. Although these differences suggest that a single calculated threshold may not be appropriate, there was little practical consequence of the three different threshold values we evaluated in that as the threshold value increased correct and incorrect classification rates increased by only a few percentage points, with a concomitant reduction in ambiguous and "no match" classifications. Perhaps across multiple studies a pattern will emerge indicating a suitable single threshold, like that used in BOLD to discriminate among animals (Ratnasingham and Hebert, 2007), exists for plants. Existence of such a threshold will depend on the magnitude of genetic differentiation among conspecifics relative to among heterospecifics, which has been documented for only a few genera (Fig. 4 and Fazekas et al., 2009).

Alternatively, one could avoid the problem of using a single calculated threshold for a large database of evolutionary divergent samples by using a species-specific threshold (i.e., for each species determining whether intraspecific distances exceed interspecific distances) as used by CBOL Plant Working Group (2009). This approach accounts for taxon specific differences in chloroplast nucleotide substitution rates but requires that the database includes multiple conspecific accessions of all taxa. Although the CBOL study did not use species-specific thresholds to assess the accuracy of barcodes in assigning taxonomy to a query sequence, in practice, such an approach could assign taxonomy by determining with which species the query sequence maximizes putative interspecific distances while minimizing intraspecific distances.

Although the liberal tree-based clustering method yielded correct classification rates as high as 71.76–95.65% in the synonymized data set, the inability of this method to accurately handle sequences that do not have a conspecific in the reference database render it inferior to other methods based on distances and diagnostic character differences. This conclusion is in contrast to Ross et al. (2008) who found the liberal tree-based method superior to distance methods in this regard. The strict method can identify sequences that have no conspecific in the database; however, because a query sequence needs to be embedded in a clade in order for a taxonomic identity to be assigned to it (Ross et al., 2008), success is highly dependent on the number of reference sequences in the database. Given that one sequence was treated as the reference and the other as a query sequence, the 14 species for which we sampled two individuals represent the extreme situation in which any query sequence will at best be sister to a conspecific and thus will be identified as ambiguous under the strict method. Given the dependence of the success of this method on the number of conspecifics sampled, it has been suggested that at least five conspecifics should be present in the reference database to ensure accurate identifications (Ross et al., 2008). The authors note that this level of sampling may not be achievable for the majority of species.

The character-based CAOS method (Sarkar et al., 2008) that uses fixed nucleotide differences across all taxa is conceptually appealing because it does not require multiple conspecifics to be in the reference database and it can deal appropriately with query sequences for which there is no conspecific in the database. Unfortunately, despite the number of segregating sites that were fixed among a priori groups (online Appendix S3), the CAOS method yielded a wide range in the percentage correct classification rate and some barcodes performed dismally compared to their performance using other methods (e.g., *matK*, *rbcL*, *rpoB*, and the multilocus barcodes; Table 5). A number of the samples that failed to be correctly classified under the CAOS method were accurately identified by the distance- and tree-based methods and formed strongly supported monophyletic groups in Pettengill and Neel (2008) (e.g., the phylogenetically distinct samples *A. setacea* 7MD and *A. tenuifolia* 5IA were classified as *A. skinneriana* using the *psbA-trnH* barcode). These misclassifications might have resulted because ancestral relationships among these species lack statistical support even when phylogenetic relationships are inferred based on many chloroplast loci (Pettengill and Neel, 2008), such that few diagnostic characters exist within individual barcodes to differentiate such relationships. The CAOS method does not incorporate uncertainty in phylogenetic relationships in that a single fully resolved topology is used, which may result in the spurious identification of diagnostic characters. This problem increases when additional sequence data are used suggesting that the method is perhaps statistically inconsistent in that there is a higher incidence of incorrect classifications at the expense of ambiguous classification when a multilocus barcode is used (Table 5). Another potential problem with the CAOS method is whether enough diagnostic character differences will be present for any one barcode when many species (e.g., thousands) are considered in the P-Gnome process. The issue is similar to the problem observed in phylogenetic studies where the degree of homoplasy increases with increased taxon sampling (Sanderson and Donoghue, 1989). In barcoding, diagnostic characters that distinguish certain species may no longer be diagnostic when additional species are considered.

Importance of taxonomy—Beyond barcode characteristics and analytical methods, classification success was conditional upon the taxonomic hypothesis used. Results under the historical taxonomy of 29 species and the "best close match" distance method yielded correct identification of ~60% of *Agalinis* individuals using the loci we sequenced (Table 3). Such low correct classification rates can result from polyphyletic or paraphyletic relationships among species that are in fact reproductively isolated and distinguishable by experts based on morphological features or from incorrect taxonomy. There is no way to distinguish these two possible causes of misclassification without additional information. In the case of *Agalinis*, previous phylogenetic studies (Neel and Cummings, 2004; Pettengill and Neel, 2008) suggested that separate species status may be unwarranted for two groups of species, first *A. acuta*, *A. decemloba*, and *A. tenella* and second *A. paupercula* and *A. purpurea*. When species in these two groups were synonymized with one another, correct classification exceeds 80% for all loci and was above 90% for the best performing loci (Table 3). Given that all three of the barcoding methods we evaluated can be misleading when species are not reciprocally monophyletic (Ross et al., 2008) and the potentially high frequency of such relationships (Rieseberg and Brouillet, 1994; Crisp and Chandler, 1996; Funk and Omland, 2003), there is potential that DNA barcoding may be challenging in many taxonomic groups (Hollingsworth et al., 2009).

Although DNA barcoding alone cannot resolve taxonomic uncertainty, it may be a useful tool for detecting it. The ability to detect cryptic species with barcoding is well known (e.g., Elias-Gutierrez and Valdez-Moreno, 2008; Yassin et al., 2008; Ragupathy et al., 2009). The technique can also be a useful tool for detecting when two putative entities may not actually be different (Fazekas et al., 2009). The "best close match" method effectively identified reference samples of *A. acuta*, *A. decemloba*, and *A. tenella* as ambiguous, and therefore, sequences of these species were assignable to more than one reference species. Had we had no other information prior to this study, the ambiguous classifications could have triggered further investigation. Thus, rather than replacing traditional taxonomic approaches, DNA barcoding can be seen as part of an iterative process in which query sequences that do not match sequences in the database or that are assigned to multiple taxa prompt examination of taxonomic hypotheses used to assign membership. The results from these additional studies may either support the current taxonomic hypothesis or suggest taxonomic revisions. There is an undeniably disconcerting circularity to this logic: if samples that cannot be delineated to a single species using barcoding are synonymized, then those samples become distinguishable

with barcodes. However, not all instances of "no match" or ambiguous classifications will result in taxonomic revisions when additional data are examined. The ability to detect incorrect taxonomy is highly dependent on the density of taxonomic coverage in the reference database (e.g., Ross et al., 2008), and the frequency at which this iterative process will be repeated will depend on the degree to which current taxonomic hypotheses are incorrect and putative species are not reciprocally monophyletic.

Conclusions—Our results show that the success of DNA barcoding varies depending on three factors: (1) the actual barcode being employed, (2) the analytical method being used to determine the taxonomic identity of a query sequence, and (3) the related factors of accuracy of the taxonomy associated with the sequences in the reference database and the degree to which species are not monophyletic. Within limits imposed by the factors above, we found two loci (*psbA-trnH* and *trnT-trnL*) to be most effective at distinguishing among congeners in the genus *Agalinis*. The "best close match" distance method generally outperformed other methods due to its high correct classification rates, potential for ambiguous classification, and identification of query sequences with no match in the database. Patterns of intra- vs. interspecies differentiation indicate that these results are applicable beyond *Agalinis*.

This success indicates that DNA barcoding can be useful in a conservation context by determining the identity of morphologically confusing species or populations that appear to represent an extension of a species' range. Although it has been argued that application of DNA barcoding to conservation is dependent on having taxonomically extensive representation in the reference database (e.g., Rubinoff, 2006), our results suggest this is not the case because methods exist that can accurately identify samples that have no conspecifics present in the database. Assuming that taxonomic hypotheses are accurate and given an appropriate method, DNA barcoding will provide a means of identifying understudied and putatively rare species that warrant additional studies to evaluate their evolutionary distinctiveness and phylogenetic affinity.

LITERATURE CITED

ALVAREZ, I., AND J. F. WENDEL. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29: 417–434.

ARMSTRONG, K. F., AND S. L. BAR. 2005. DNA barcodes for biosecurity: Invasive species identification. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences* 360: 1813–1823.

BLAXTER, M. L. 2004. The promise of a DNA taxonomy. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences* 359: 669–679.

BURNS, J. M., D. H. JANZEN, M. HAJIBABAEI, W. HALLWACHS, AND P. D. N. HEBERT. 2008. DNA barcodes and cryptic species of skipper butterflies in the genus *Perichares* in Area de Conservación Guanacaste, Costa Rica. *Proceedings of the National Academy of Sciences, USA* 105: 6350–6355.

CANNE, J. M. 1979. A light and scanning electron microscope study of seed morphology in *Agalinis* (Scrophulariaceae) and its taxonomic significance. *Systematic Botany* 4: 281–296.

CANNE, J. M. 1980. Seed surface features in *Aureolaria*, *Brachystigma Tomanthera*, and certain South American *Agalinis* (Scrophulariaceae). *Systematic Botany* 5: 241–252.

CANNE, J. M. 1981. Chromosome counts in *Agalinis* and related taxa (Scrophulariaceae). *Canadian Journal of Botany* 59: 1111–1116.

CANNE, J. M. 1983. The taxonomic significance of seedling morphology in *Agalinis* (Scrophulariaceae). *Canadian Journal of Botany* 62: 454–456.

CANNE, J. M. 1984. Chromosome numbers and the taxonomy of North American *Agalinis* (Scrophulariaceae). *Canadian Journal of Botany* 62: 454–456.

CANNE-HILLIKER, J. M. 1987. Patterns of floral development in *Agalinis* and allies (Scrophulariaceae). II. Floral development of *Agalinis densiflora. American Journal of Botany* 74: 1419–1430.

CANNE-HILLIKER, J. M., AND M. DUBRULE. 1993. A new species of *Agalinis* (Scrophulariaceae) from Grimes County, Texas. *Sida* 15: 425–440.

CANNE-HILLIKER, J. M., AND C. M. KAMPNY. 1991. Taxonomic significance of leaf and stem anatomy of *Agalinis* (Scrophulariaceae) from the U.S.A. and Canada. *Canadian Journal of Botany* 69: 1935–1950.

CBOL PLANT WORKING GROUP. 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences, USA* 106: 12794–12797.

CHASE, M. W., N. SALAMIN, M. WILKINSON, J. M. DUNWELL, R. P. KESANAKURTHI, N. HAIDAR, AND V. SAVOLAINEN. 2005. Land plants and DNA barcodes: Short-term and long-term goals. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences* 360: 1889–1895.

CRISP, M. D., AND G. T. CHANDLER. 1996. Paraphyletic species. *Telopea* 6: 813–844.

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, atp*B, and *matK* DNA sequences. *American Journal of Botany* 89: 132–144 .

DARLING, J. A., AND M. J. BLUM. 2007. DNA-based methods for monitoring invasive species: A review and prospectus. *Biological Invasions* 9: 751–765.

DAWNAY, N., R. OGDEN, R. MCEWING, G. R. CARVALHO, AND R. S. THORPE. 2007. Validation of the barcoding gene COI for use in forensic genetic species identification. *Forensic Science International* 173: 1–6.

DEPAMPHILIS, C. W., N. D. YOUNG, AND A. D. WOLFE. 1997. Evolution of plastid gene *rps2* in a lineage of hemiparasitic and holoparasitic plants: Many losses of photosynthesis and complex patterns of rate variation. *Proceedings of the National Academy of Sciences, USA* 94: 7367–7372.

DESALLE, R., M. G. EGAN, AND M. SIDDALL. 2005. The unholy trinity: Taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences* 360: 1905–1916.

EDGAR, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.

EDWARDS, D., A. HORN, D. TAYLOR, V. SAVOLAIN, AND J. A. HAWKINS. 2008. DNA barcoding of a large genus, *Aspalathus* L. (Fabaceae). *Taxon* 57: 1317–1327.

ELIAS-GUTIERREZ, M., AND M. VALDEZ-MORENO. 2008. A now cryptic species of *Leberis* Smirnov, 1989 (Crustacea, Cladocera, Chydoridae) from the Mexican semi-desert region, highlighted by DNA barcoding. *Hidrobiologica* 18: 63–74.

ERICKSON, D. L., J. SPOUGE, A. RESCH, L. A. WEIGT, AND W. J. KRESS. 2008. DNA barcoding in land plants: Developing standards to quantify and maximize success. *Taxon* 57: 1304–1316.

FAITH, D. P., AND A. M. BAKER. 2006. Phylogenetic diversity (PD) and biodiversity conservation: Some bioinformatics challenges. *Evolutionary Bioinformatics Online* 2: 70–77.

FAZEKAS, A. J., K. S. BURGESS, P. R. KESANAKURTI, S. W. GRAHAM, S. G. NEWMASTER, B. C. HUSBAND, D. M. PERCY, ET AL. 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS ONE* 3: e2802.

FAZEKAS, A. J., P. R. KESANAKURTI, K. S. BURGESS, D. M. PERCY, S. W. GRAHAM, S. C. H. BARRETT, S. G. NEWMASTER, ET AL. 2009. Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Molecular Ecology Resources* 9: 130–139.

FERGUSON, J. W. H. 2002. On the use of genetic divergence for identifying species. *Biological Journal of the Linnean Society* 75: 509–516.

FUNK, D. J., AND K. E. OMLAND. 2003. Species-level paraphyly and polyphyly: Frequency, causes and consequences with insights from animal mitochondrial DNA. *Annual Review of Ecology Evolution and Systematics* 34: 397–423.

HEBERT, P. D. N., A. CYWINSKA, S. L. BALL, AND J. R. DEWAARD. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London, B, Biological Sciences* 270: 313–321.

HEY, J., AND J. WAKELEY. 1997. A coalescent estimator of the population recombination rate. *Genetics* 145: 833–846.

HOLLINGSWORTH, M. L., A. A. CLARK, L. L. FORREST, J. RICHARDSON, R. T. PENNINGTON, D. G. LONG, R. COWAN, ET AL. 2009. Selecting barcoding loci for plants: Evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Molecular Ecology Resources* 9: 439–457.

HUDSON, R. R., AND J. A. COYNE. 2002. Mathematical consequences of the genealogical species concept. *Evolution* 56: 1557–1565.

KAMPNY, C. M., AND J. M. CANNE-HILLIKER. 1987. Patterns of floral development in *Agalinis* and allies (Scrophulariaceae). I. Floral development of *Agalinis fasciculata* and *A. tenuifolia*). *Canadian Journal of Botany* 65: 2255–2262.

KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide-sequences. *Journal of Molecular Evolution* 16: 111–120.

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.

KRESS, W. J., K. J. WURDACK, E. A. ZIMMER, L. A. WEIGT, AND D. H. JANZEN. 2005. Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences, USA* 102: 8369–8374.

LAHAYE, R., M. VAN DER BANK, D. BOGARIN, J. WARNER, F. PUPULIN, G. GIGOT, O. MAURIN, ET AL. 2008. DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences, USA* 105: 2923–2928.

LITTLE, D. P., AND D. W. STEVENSON. 2007. A comparison of algorithms for the identification of specimens using DNA barcodes: Examples from gymnosperms. *Cladistics* 23: 1–21.

MANEN, J., C. HABASHI, D. JEANMONOD, J. PARK, AND G. M. SCHNEEWEISS. 2004. Phylogeny and intraspecific variability of holoparasitic *Orobanche* (Orobanchaceae) inferred from plastid *rbcL* sequences. *Molecular Phylogenetics and Evolution* 33: 482–500.

MEIER, R., K. SHIYANG, G. VAIDYA, AND P. K. L. NG. 2006. DNA barcoding and taxonomy in diptera: A tale of high intraspecific variability and low identification success. *Systematic Biology* 55: 715–728.

MEYER, C. P., AND G. PAULAY. 2005. DNA barcoding: Error rates based on comprehensive sampling. *PLoS Biology* 3: 2229–2238.

MORRISON, D. A. 2009. Why would phylogeneticists ignore computerized sequence alignment? *Systematic Biology* 58: 150–158.

MURRAY, T. E., U. FITZPATRICK, M. J. F. BROWN, AND R. J. PAXTON. 2008. Cryptic species diversity in a widespread bumble bee complex revealed using mitochondrial DNA RFLPs. *Conservation Genetics* 9: 653–666.

NATURESERVE. 2007. NatureServe Explorer: An online encyclopedia of life [web application], version 6.3. NatureServe, Arlington, Virginia, USA. Website http://www.natureserve.org/explorer.

NEEL, M. C., AND M. P. CUMMINGS. 2004. Section-level relationships of North American *Agalinis* (Orobanchaceae) based on DNA sequence analysis of three chloroplast gene regions. *BMC Evolutionary Biology* 4: 15.

NEWMASTER, S. G., A. J. FAZEKAS, AND S. RAGUPATHY. 2006. DNA barcoding in land plants: Evaluation of *rbcL* in a multigene tiered approach. *Canadian Journal of Botany* 84: 335–341.

NEWMASTER, S. G., A. J. FAZEKAS, R. A. D. STEEVES, AND J. JANOVEC. 2008. Testing candidate plant barcode regions in the Myristicaceae. *Molecular Ecology Resources* 8: 480–490.

NEWMASTER, S. G., AND S. RAGUPATHY. 2009. Testing plant barcoding in a sister species complex of pantropical *Acacia* (Mimosoideae, Fabaceae). *Molecular Ecology Resources* 9: 172–180.

PALMER, J. D. 1985. Evolution of chloroplast and mitochondrial DNA in plants and algae, Plenum, New York, New York, USA.

PENNELL, F. W. 1913a. Studies in the Agalinanae, a subtribe of the Rhinanthaceae. I. Nomenclature of the Neartic genera. *Bulletin of the Torrey Botanical Club* 40: 119–130.

PENNELL, F. W. 1913b. Studies in the Agalinanae, a subtribe of the Rhinanthaceae. II. Species of the Atlantic Coastal Plain. *Bulletin of the Torrey Botanical Club* 40: 401–439.

PENNELL, F. W. 1928. *Agalinis* and allies in North America I. *Proceedings of the Academy of Natural Sciences of Philadelphia* 80: 339–449.

PENNELL, F. W. 1929. *Agalinis* and allies in North America II. *Proceedings of the Academy of Natural Sciences of Philadelphia* 81: 111–249.

PENNELL, F. W. 1935. The Scrophulariaceae of eastern temperate North America (41 *Gerardia*). *Academy of Natural Sciences of Philadelphia Monographs* 1: 419–476.

PETTENGILL, J. B., AND M. C. NEEL. 2008. Phylogenetic patterns and conservation among North American members of the genus *Agalinis* (Orobanchaceae). *BMC Evolutionary Biology* 8: 264.

RAGUPATHY, S., S. G. NEWMASTER, M. MURUGESAN, AND V. BALASUBRAMANIAM. 2009. DNA barcoding discriminates a new cryptic grass species revealed in an ethnobotany study by the hill tribes of the Western Ghats in southern India. *Molecular Ecology Resources* 9: 164–171.

RATNASINGHAM, S., AND P. D. N. HEBERT. 2007. BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes* 7: 355–364.

RIESEBERG, L. H., AND L. BROUILLET. 1994. Are many plant species paraphyletic? *Taxon* 43: 21–32.

ROSENBERG, N. A. 2003. The shapes of neutral gene genealogies in two species: Probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. *Evolution; International Journal of Organic Evolution* 57: 1465–1477.

ROSS, H. A., S. MURUGAN, AND W. L. S. LI. 2008. Testing the reliability of genetic methods of species identification via simulation. *Systematic Biology* 57: 216–230.

RUBINOFF, D. 2006. Utility of mitochondrial DNA barcodes in species conservation. *Conservation Biology* 20: 1026–1033.

SAITOU, N., AND M. NEI. 1987. The neighbor-joining method—A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406–425.

SANDERSON, M. J., AND M. J. DONOGHUE. 1989. Patterns of variation in levels of homoplasy. *Evolution* 43: 1781–1795.

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 .

SARKAR, I. N., P. J. PLANET, T. E. BAEL, S. E. STANLEY, M. SIDDALL, R. DESALLE, AND D. H. FIGURSKI. 2002b. Characteristic attributes in cancer microarrays. *Journal of Biomedical Informatics* 35: 111–122.

SARKAR, I. N., P. J. PLANET, AND R. DESALLE. 2008. CAOS software for use in character-based DNA barcoding. *Molecular Ecology Resources* 8: 1256–1259.

SARKAR, I. N., J. W. THORNTON, P. J. PLANET, D. H. FIGURSKI, B. SCHIERWATER, AND R. DESALLE. 2002a. An automated phylogenetic key for classifying homeoboxes. *Molecular Phylogenetics and Evolution* 24: 388–399.

SASS, C., D. P. LITTLE, D. W. STEVENSON, AND C. D. SPECHT. 2007. DNA barcoding in the Cycadales: Testing the potential of proposed barcoding markers for species identification of cycads. *PLoS ONE* 2: e1154.

SAVOLAINEN, V., R. S. COWAN, A. P. VOGLER, G. K. RODERICK, AND R. LANE. 2005. Towards writing the encyclopedia of life: An introduction to DNA barcoding. *Philosophical transactions of the Royal Society of London, B, Biological Sciences* 360: 1805–1811.

STARR, J. R., R. F. C. NACZI, AND B. N. CHOUINARD. 2009. Plant DNA barcodes and species resolution in sedges (*Carex*, Cyperaceae). *Molecular Ecology Resources* 9: 151–163.

SWOFFORD, D. L. 2003. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0. Sinauer, Sunderland, Massachusetts, USA.

TABERLET, P., E. COISSAC, F. POMPANON, L. GIELLY, C. MIQUEL, A. VALENTINI, T. VERMAT, ET AL. 2007. Power and limitations of the chloroplast *trn*L (UAA) intron for plant DNA barcoding. *Nucleic Acids Research* 35: e14.

TABERLET, P., L. GIELLY, G. PAUTOU, AND J. BOUVET. 1991. Universal primers for amplification of 3 noncoding regions of chloroplast DNA. *Plant Molecular Biology* 17: 1105–1109.

TREWICK, S. A. 2008. DNA barcoding is not enough: Mismatch of taxonomy and genealogy in New Zealand grasshoppers (Orthoptera: Acrididae). *Cladistics* 24: 240–254.

VOGLER, A. 2006. Will DNA barcoding advance efforts to conserve biodiversity more efficiently than traditional taxonomic methods? *Frontiers in Ecology and the Environment* 5: 272–273.

WILL, K. W., B. D. MISHLER, AND Q. D. WHEELER. 2005. The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology* 54: 844–851.

WILSON, M. A., B. GAUT, AND M. T. CLEGG. 1990. Chloroplast DNA evolves slowly in the palm family (Arecaceae). *Molecular Biology and Evolution* 7: 303–314.

WOLFE, K. H., C. W. MORDEN, S. C. EMS, AND J. D. PALMER. 1992. Rapid evolution of the plastid translational apparatus in a nonphotosynthetic plant: loss or accelerated sequence evolution of tRNA and ribosomal protein genes. *Journal of Molecular Evolution* 35: 304–317.

YASSIN, A., P. CAPY, L. MADI-RAVAZZI, D. OGEREAU, AND J. R. DAVID. 2008. DNA barcode discovers two cryptic species and two geographical radiations in the invasive drosophilid *Zaprionus indianus*. *Molecular Ecology Resources* 8: 491–501.

ZWICKL, D. J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence data sets under the maximum likelihood criterion. Ph.D. dissertation, University of Texas, Austin, Texas, USA.

APPENDIX 1. North American *Agalinis* samples examined in this study, general sampling location, known voucher information, and GenBank accession numbers. Voucher specimens are deposited in the following herbaria: University of Texas at Austin = GLV, University of Maryland College Park = NBH, University of Guelph = OAC. Voucher specimens were not collected from the endangered species, *A. acuta*, because these populations are well documented by state Natural Heritage Programs and the U.S. Fish and Wildlife Service. (N/A = not available.)

SECTION, SUBSECTION, *Species*, Sample locality, Voucher (herbarium), GenBank accessions: *matK*, *rbcL*, *rps2*, *trnT*-a-*trnF*-f, *trnH-psbA*, *rpoB*.

SECTION ERECTAE. *A. acuta* 125CT, Windham Co., CT, N/A, GU943656, GU943571, EU827965, EU828128.2, EU827882, EU828046. *A. acuta* 139RI, Washington Co., RI, N/A, GU943657, GU943572, EU827966, EU828129, EU827883, EU828047. *A. acuta* 13PCMA, Barnstable Co., MA, N/A, GU943658, GU943573, EU827967, EU828130, EU827884, EU828048. *A. acuta* 1BVMA, Barnstable Co., MA, N/A, GU943660, GU943575, EU827968, EU828131, EU827885, EU828049. *A. acuta* 211HPNY, Nassau Co., NY, N/A, GU943661, GU943576, EU827969, EU828132, EU827886, EU828050. *A. acuta* 229MDNY, Suffolk Co., NY, N/A, GU943662, GU943577, EU827970, EU828133, EU827887, EU828051. *A. acuta* 265SMNY, Suffolk Co., NY, N/A, GU943663, GU943578, EU827971, EU828134, EU827888, EU828052. *A. acuta* 33SNMA, Dukes Co., MA, N/A, GU943664, GU943579, EU827972, EU828135, EU827889, EU828053. *A. acuta* 51MD, Baltimore Co., MD, N/A, GU943665, GU943580, EU827973, EU828136, EU827890, EU828054. *A. acuta* 161RI, Washington Co., RI, N/A, GU943659, GU943574, GU943545, GU943559, GU943532, GU943519. *A. aphylla* 3FL, Liberty Co., FL, JCH 3545 (OAC), GU943666, GU943581, EU827974, EU828137, EU827891, EU828055. *A. aphylla* 4AL, Geneva Co., AL, JCH 3579 (OAC), GU943667, GU943582, EU827975, EU828138, EU827892, EU828056. *A. decemloba* 1VA, Lunenberg Co., VA, N/A, GU943671, N/A, GU943548, GU943558, GU943534, GU943521. *A. decemloba* 6VA, Lunenberg Co., VA, JCH 3502 (OAC), GU943673, GU943588, EU827982, EU828145.2, EU827899, EU828063. *A. decemloba* 45VA, Lunenberg Co., VA, N/A, GU943672, GU943587, GU943549, GU943560, GU943535, GU943522. *A. decemloba* 19NC, Wake Co., NC, JBP 072328 (NBH), N/A, GU943586, GU943547, GU943561, GU943533, GU943520. *A. decemloba* 9NC, Randolph Co., NC, JCH 3506 (OAC), GU943674, GU943589, EU827983, EU828146, EU827900, EU828064. *A. gattingeri* 45MO, Crawford Co., MO, JCH 3604 (OAC), GU943684, GU943598, EU827993, EU828156, EU827910, EU828074. *A. gattingeri* 8MO, Hickory Co., MO, JCH 3606 (OAC), GU943685, GU943599, EU827994, EU828157, EU827911, EU828075. *A. gattingeri* 1LA, Webster Parish, LA, JCH 3614 (OAC), GU943683, GU943597, EU827992, EU828155, EU827909, EU828073. *A. obtusifolia* 177SC, Dorchester Co., SC, JBP 072326 (NBH), GU943704, GU943618, GU943553, GU943564, GU943538, GU943525. *A. obtusifolia* 169SC, Lee Co., SC, JBP 072330 (NBH), GU943703, GU943617, GU943552, GU943565, GU943537, GU943524. *A. obtusifolia* 18FL, FL, JCH 3595 (OAC), N/A, N/A, GU943554, GU943562, GU943539, GU943526. *A. obtusifolia* 10AL, Baldwin Co., AL, JCH 3577 (OAC), N/A, N/A, GU943551, GU943563, GU943536, GU943523. *A. obtusifolia* 13AL, Geneva Co., AL, JCH 3580 (OAC), GU943701, GU943615, EU828008, EU828171, EU827925, EU828089. *A. obtusifolia* 14AL, Mobile Co., AL, N/A, GU943702, GU943616, N/A, EU828172, EU827926, EU828089. *A. obtusifolia* 20FL, Liberty Co., FL, JCH 3598 (OAC), GU943706, GU943619, EU828010, EU828174, EU827928, EU828092. *A. obtusifolia* 6AL, Mobile Co., AL, JCH 3571 (OAC), GU943707, GU943620, EU828011, EU828175, EU827929, EU828093. *A. obtusifolia* 8AL, Geneva Co., AL, JCH 3576 (OAC), GU943708, GU943621, N/A, EU828176, EU827930, EU828094. *A. oligophylla* 12AL, Tyler Co., TX, JCH 3633 (OAC), GU943709, GU943622, EU828012, EU828177, EU827931, EU828095. *A. oligophylla* 1AL, Mobile Co., AL, JCH 3564 (OAC), GU943710, GU943623, EU828013, EU828178, EU827932, EU828096. *A. oligophylla* 5AL, Mobile Co., AL, JCH 3574 (OAC), N/A, GU943624, EU828014, EU828179, EU827933, EU828097. *A. oligophylla* 8TX, Vernon Parish, LA, JCH 3626 (OAC), GU943711, GU943625, EU828015, EU828180, EU827934, EU828098. *A. skinneriana* 106MD, Prince Georges Co., MD, N/A, GU943724, GU943639, EU828028, EU828193, EU827947, EU828110. *A. skinneriana* 78MD, Dorchester Co., MD, N/A, GU943725, GU943640, EU828029, EU828194, EU827948, EU828111. *A. skinneriana* 90MO, Vernon Co., MO, JCH 3607 (OAC), GU943726, GU943641, EU828030, EU828195, EU827949, EU828112. *A. tenella* 1GA, Ware Co., GA, JCH 3537 (OAC), GU943731, GU943646, EU828009, EU828173, EU827927, EU828091. *A. tenella* 79GA, Brooks Co., GA, NBH 072331 (NBH), N/A, N/A, GU943555, GU943569, GU943542, GU943529. *A. tenella* 11GA, Lowndes Co., GA, JCH 3540 (OAC), GU943729, GU943644, EU828032, EU828197, EU827951, EU828114. *A. tenella* 13GA, Grady Co., GA, JCH 3543 (OAC), GU943730, GU943645, EU828033, EU828198, EU827952, EU828115. *A. tenella* 91GA, Grady Co., GA, JBP 072329 (NBH), GU943734, N/A, GU943556, GU943570, GU943543, GU943530. *A. tenella* 3SC, Colleton Co., SC, JCH 3520 (OAC), GU943732, GU943647, EU828034, EU828199, EU827953, EU828116. *A. tenella* 4GA, Ware Co., GA, JCH 3537 (OAC), GU943733, GU943648, EU828035, EU828200, EU827954, EU828117. *A. tenella* 9GA, Lanier Co., GA, JCH 3539 (OAC), GU943735, GU943649, EU828036, EU828201, EU827955, EU828118. *A. viridis* 2LA, Natchitoches Parish, LA, JCH 3620 (OAC), GU943740, GU943654, EU828040, EU828205, EU827959, EU828122. *A. viridis* 9IL, DeSoto Parish, LA, JCH 3617 (OAC), GU943741, GU943655, EU828041, EU828206, EU827960, EU828123.

SECTION HETEROPHYLLAE. *A. auriculata* 1IA, Story Co., IA, DH 1 (N/A), GU943668, GU943583, EU827976, EU828139, EU827893, EU828057. *A. auriculata* 7IL, Will Co., IL, J. Koontz 5 (N/A), GU943669, GU943584, EU827977, EU828140, EU827894, EU828058. *A. calycina*, Pecos Co., TX, N/A, GU943670, GU943585, EU827978, EU828141, EU827895, EU828059. *A. heterophylla* 5TX, Cameron Co., TX,

Cabrera and Dieringer 1057 (N/A), GU943687, GU943602, EU827997, EU828160, EU827914, EU828078. *A. heterophylla* 8TX, Stephens Co., TX, JCH 3636 (OAC), GU943688, GU943603, EU827998, EU828161, EU827915, EU828079. *A.* cf. *heterophylla* TX, Grimes Co., TX, JLN 01-10-07-02 (GLV), GU943689, GU943604, EU827979, EU828142, EU827896, EU828060.

SECTION LINIFOLIAE. *A. linifolia* 2FL, Liberty Co., FL, JCH 3521 (OAC), GU943694, GU943609, EU828003, EU828166, EU827920, EU828084. *A. linifolia* 4GA, Cinch Co., GA, JCH 3538 (OAC), GU943695, GU943610, EU828004, EU828167, EU827921, EU828085.

SECTION PURPUREAE, **SUBSECTION PEDUNCULARES**. *A. edwardsiana* 1TX, Stephens Co., TX, JLN 01-10-07-01 (GLV), GU943677, GU943592, EU827986, EU828149, EU827903, EU828067. *A. homalantha* 1TX, Tyler Co., TX, JCH 3630 (OAC), GU943690, GU943605, EU827999, EU828162, EU827916, EU828080. *A. homalantha* 2TX, Jasper Co., TX, JCH 3642 (OAC), GU943691, GU943606, EU828000, EU828163, EU827917, EU828081. *A. pulchella* 3GA, Grady Co., GA, JCH 3544 (OAC), GU943716, GU943630, EU828020, EU828185, EU827939, EU828102. *A. pulchella* 4FL, Florida, JCH 3596 (OAC), N/A, GU943631, EU828021, EU828186, EU827940, EU828103. *A. strictifolia* 4, Stephens Co., TX, JLN 01-10-07-03 (GLV), GU943727, GU943642, EU828031, EU828196, EU827950, EU828113. *A.* cf. *strictifolia* TX, Cameron Co., TX, Cabrera & Dieringer 1056 (N/A), GU943728, GU943643, EU827981, EU828144, EU827898, EU828062. *A. navasotensis* 1TX, Tyler Co., TX, JCH 3634 (OAC), GU943698, GU943613, EU828006, EU828169, EU827923, EU828087. *A. navasotensis* 5TX, Grimes Co., TX, JCH 3637 (OAC), GU943699, GU943614, EU828007, EU828170, EU827924, EU828088.

SUBSECTION PURPUREAE. *A. fasciculata* 1LA, Grimes Co., TX, N/A, GU943678, GU943593, EU827987, EU828150, EU827904, EU828068. *A. fasciculata* 2GA, Long Co., GA, JCH 3529 (OAC), GU943679, GU943594, EU827988, EU828151, EU827905, EU828069. *A. fasciculata* 4LA, Caddo Parish, LA, JCH 3615 (OAC), GU943680, N/A, EU827989, EU828152, EU827906, EU828070. *A. harperi* 13FL, Liberty Co., FL, JCH 3597 (OAC), N/A, GU943600, EU827995, EU828158, EU827912, EU828076. *A. harperi* 14NC, Brunswick Co., NC, JCH 3516 (OAC), GU943686, GU943601, EU827996, EU828159, EU827913, EU828077. *A. maritima* TX, Cameron Co., TX, Cabrera & Dieringer 1058 (N/A), GU943697, GU943612, EU827980, EU828143, EU827897, EU828061. *A. maritima* 2MA, Barnstable Co., MA, PP 3, GU943696, GU943611, EU828005, EU828168, EU827922, EU828086. *A. paupercula* 4MA, Barnstable Co., MA, N/A, GU943712, GU943626, EU828016, EU828181, EU827935, EU828099. *A. paupercula* 7NY, Suffolk Co., NY, N/A, GU943713, GU943627, EU828017, EU828182, EU827936, EU828100. *A. purpurea* 101VA, Fauquier Co., VA, N/A, GU943717, GU943632, EU828022, EU828187, EU827941, EU828104. *A. purpurea* 1AL, Mobile Co., AL, JCH 3573 (OAC), GU943718, GU943633, EU828023, EU828188, EU827942, EU828105. *A. purpurea* 64MD, Dorchester Co., MD, N/A, GU943719, GU943634, EU828024, EU828189, EU827943, EU828106. *A. purpurea* 6SC, Harry Co., SC, JCH 3518 (OAC), GU943720, GU943635, EU828025, EU828190, EU8279414, EU828107. *A. tenuifolia* 2VA, Prince Edward Co., VA, JCH 3498 (OAC), GU943737, GU943651, EU828038, EU828203, EU827957, EU828120. *A. tenuifolia* 5IA, Story Co., IA, N/A, GU943738, GU943652, EU828039, EU828204, EU827958, EU828121. *A. tenuifolia* 10LA, Caddo Parish, LA, JCH 3616 (OAC), GU943736, GU943650, EU828037, EU828202, EU827956, EU828119. *A. tenuifolia 46*FCVA, Fauquier Co., VA, N/A, GU943739, GU943653, GU943557, GU943568, GU943544, GU943531.

SUBSECTION SETACEAE. *A. laxa* 3SC, Colleton Co., SC, JCH 3521 (OAC), GU943692, GU943607, EU828001, EU828164, EU827918, EU828082. *A. laxa* 4GA, Long Co., GA, JCH 3526 (OAC), GU943693, GU943608, EU828002, EU828165, EU827919, EU828083. *A. plukenettii* 2FL, Washington Co., FL, JCH 3558 (OAC), GU943714, GU943628, EU828018, EU828183, EU827937, N/A. *A. plukenettii* 4GA, GA, JCH 3601 (OAC), GU943715, GU943629, EU828019, EU828184, EU827938, EU828101. *A. setacea* 3VA, Prince Edward Co., VA, JCH 3499 (OAC), GU943722, GU943637, EU828026, EU828191, EU827945, EU828108. *A. setacea* 7MD, Wicomico Co., MD, N/A, GU943723, GU943638, EU828027, EU828192, EU827946, EU828109. *A. setacea* 76VA, Lunenberg Co., VA, N/A, N/A, N/A, GU943550, GU943566, GU943541, GU943528. *A. setacea* 100VA, Fauquier Co., VA, N/A, GU943721, GU943636, GU943546, GU943567, GU943540, GU943527.

SECTION TENUIFOLIAE. *A. filicaulis* 5FL, Grady Co., GA, JCH 3544 (OAC), GU943682, GU943596, EU827991, EU828154, EU827908, EU828072. *A. filicaulis* 1AL, Mobile Co., AL, JCH 3569 (OAC), GU943681, GU943595, EU827990, EU828153, EU827907, EU828071. *A. divaricata* 3FL, Liberty Co., FL, JCH 3559 (OAC), GU943675, GU943590, EU827985, EU828147, EU827901, EU828065. *A. divaricata* 5FL, Washington Co., FL, JCH 3552 (OAC), GU943676, GU943591, EU827985, EU828148, EU827902, EU828066.