American Journal of Botany

American Journal of Botany 98(5): 896–908. 2011.

Molecular systematics of the parasitic genus *Conopholis* (Orobanchaceae) inferred from plastid and nuclear sequences[1]

Anuar G. Rodrigues[2,4], Alison E.L. Colwell[3], and Saša Stefanović[2]

[2]Department of Biology, University of Toronto Mississauga, Mississauga, Ontario, L5L 1C6, Canada; and [3]Division of Biological Resources, Yosemite National Park, El Portal, California 95318 USA

- *Premise of the study*: Little is known of the evolutionary relationships within *Conopholis*, a small holoparasitic genus belonging to the broomrape family. Presently, *Conopholis* is described as having two species, *C. americana* and *C. alpina*. This classification is based on a combination of presence/absence of morphological characters along with a number of quantitative traits. We assessed the relationships among populations and species of this genus to determine whether the present taxonomic hypothesis is reflected in molecular phylogenies.
- *Methods*: We conducted the first phylogenetic study of *Conopholis* using plastid (*trnfM-E* intergenic spacer and *clpP* gene/introns) and nuclear (*PHYA* intron 1) sequences from a wide taxonomic sampling covering its entire geographical range in North America. Analyses were carried out using a variety of phylogenetic inference approaches.
- *Key results*: Reciprocal monophyly between the two traditionally accepted species has not yet been achieved. Instead, three distinct genetic clusters were recovered. *Conopholis alpina* is clearly paraphyletic and shows evidence of belonging to at least two distinct lineages. Specimens found in Costa Rica and Panama form a distinct group from those located in northern Mexico and the southwestern United States. The monophyly of *C. americana* was also not recovered; however, the possibility of it being monophyletic could not be rejected with confidence.
- *Conclusions*: These analyses recovered three distinct lineages indicating that there could be a minimum of three species within the genus. A reevaluation of morphological features within *Conopholis* may reveal shared features that could further corroborate our molecular findings.

Key words: *Conopholis*; molecular phylogeny; nuclear *PHYA*; Orobanchaceae; parasitic plants; plastid *trnfM-E*; plastid *clpP*.

Orobanchaceae, as redefined by Young et al. (1999), Olmstead et al. (2001), and the Angiosperm Phylogeny Group III (APG III, 2009), is a morphologically diverse family comprised of herbaceous parasitic plants containing approximately one-half of all known parasitic angiosperms (ca. 1800 species), circumscribed in some 90 genera (Nickrent, 2010). With the notable exception of species belonging to several small genera (e.g., *Lindenbergia*, *Rehmannia*, and *Triaenophora*; Jensen et al., 2008; Albach et al., 2009; Xia et al., 2009), all other members of this family are facultative or obligate root parasites. They may either retain the various degrees of capability to photosynthesize (hemiparasitic species) or be completely dependent on their host for nutrients and water (holoparasitic species). The evolution of advanced holoparasitism is accompanied generally by an extreme reduction or modification, entailing both physiological and morphological changes. Those include the loss or reduction of chlorophyll production, photosynthesis, and vegetative structures such as leaves, roots, and branches along with the gain of haustoria, the organs that enable these plants to connect to their hosts vascular systems. As a result of this overall reduction in morphological features, also known as the "parasitic reduction syndrome" (Colwell, 1994), holoparasites remain difficult to study from a taxonomic and systematic point of view.

Conopholis is one such morphologically distinct group, one of ca. 20 holoparasitic genera in Orobanchaceae. Members of this genus are obligate perennial achlorophyllous parasites (Kuijt, 1969; Haynes, 1971). The mature plant body consists of several erect flowering stalks arising from a swollen subterranean haustorium (consisting of both parasite and host tissue) that connects the plant to the vascular system of its host (oaks; Baird and Riopel, 1986a). Leaves are reduced to scales and roots are absent. Following 3–4 yr of subterranean tubercle growth, these plants reach reproductive maturity and floral meristems may erupt above ground and produce inflorescences. The mature plant will continue to grow for ca. 10 yr, after which it dies, presumably due to a disruption of the haustorial connection (Baird and Riopel, 1986b).

Populations of *Conopholis* are best described as locally abundant but rare and isolated, at times separated by many kilometers of forest or other habitat in which no individuals occur. On occasion, bumblebees have been seen visiting populations of *Conopholis* (Haynes, 1971; Gomez, 1980; Baird and Riopel, 1986b), but the actual pollination by bees has not been confirmed.

[1]Manuscript received 24 September 2010; revisions accepted 16 February 2011.

The authors warmly thank C. Campbell, W. Flynn, L. Goertzen, S. Hill, M. Kuzmina, J. Meyers, S. O'Kane, G. Porta, N. Prentiss, J. Robertson, A. Salina Tovar, G. Schatz, K. Shaw, E. Tsai, and the curators/directors of ARIZ, INBIO, MO, NY, and US for supplying plant material and G. Yatskievych and T. Collins and two anonymous reviewers for critical comments on the manuscript. Financial support from the Natural Sciences and Engineering Research Council of Canada Discovery grant (326439-06) and University of Toronto Connaught New Staff Matching grant to S.S. are gratefully acknowledged.

[4] Author for correspondence (e-mail: anuar.rodrigues@utoronto.ca)

doi:10.3732/ajb.1000375

These plants do not have floral nectaries and are not known to produce scents that may attract bees. Bagging experiments performed to investigate the roles of wind or insects in *Conopholis* pollination (Baird and Riopel, 1986b) showed only a slight reduction in viable seed set (85%) compared to unbagged controls (87%). In addition, studies of flowers postanthesis have found that the anthers are in physical contact with the stigma. In aggregate, these observations suggest a predominant selfing mode of pollination for *Conopholis*. Dispersal of seeds occurs either from drying and decomposition of the capsule and the washing away of seeds following periods of rain or following the consumption of the inflorescence by mammals, especially deer (Baird and Riopel, 1986b). The role of ants has not been investigated in detail, but anecdotal observation suggests that their role in both pollination and dispersal of *Conopholis* seems minimal.

Little is known of the evolutionary relationships among populations and species of *Conopholis* or of their postglacial history. In the most recent taxonomic classification of this genus, taking into account geographic distribution, morphology, reproductive isolation, and host specificity, *Conopholis* is described as having two species: *C. americana* and *C. alpina* (Haynes, 1971). *Conopholis americana* parasitizes red oaks (*Quercus* section *Lobatae*; Manos et al., 2001) in moist, deciduous or mixed forests and is found across eastern North America, from Florida north to Nova Scotia west to Wisconsin and south to Alabama (Fig. 1). *Conopholis alpina* parasitizes various oak species (predominantly white oak; *Quercus* section *Quercus*) in oak woodlands and mixed montane forests found in southwestern North America. This species is divided into two varieties. *Conopholis alpina* var. *mexicana* is found from the Trans-Pecos area in Texas through northern New Mexico and central Arizona south down to the Trans-Mexican volcanic belt (TMVB), a large mountain range running from east to west in the central portion of Mexico, located approximately along the 19°N parallel and well-known as a gene-flow barrier for other plant species (Haynes, 1971; Nixon, 1993). The type variety (*C. alpina* var. *alpina*) is distributed from this same central area of Mexico south to Costa Rica and Panama (Haynes, 1971; Fig. 1).

Morphologically, the classification proposed by Haynes (1971) is based on a combination of presence/absence of characters along with a number of quantitative traits such as the size and relative proportions of bracts as well as the shape of the calyx. However, Haynes (1971) states that the calyx is the most variable part of the plant, and as a result, differences pertaining to it cannot alone be used as criteria for taxonomic placement of specimens. In his view, the two species of *Conopholis* are morphologically distinct, yet "No single character can be relied upon to determine all specimens encountered…" (p. 252). Haynes (1971) acknowledged this to be a conundrum, suggesting further that one needs to consider several characters at the same time to identify a specimen to the correct taxon. Nevertheless, he recognized them as two separate species because of their clear present day geographic isolation (>1400 km; Fig. 1) and apparent differential host specificity. This bicentric geographic pattern of *Conopholis* in North America with an east–west disjunction is also found in other plant groups (e.g., *Chamaecyparis*, Mylecraine et al., 2004; and *Platanus*, Feng et al., 2005).

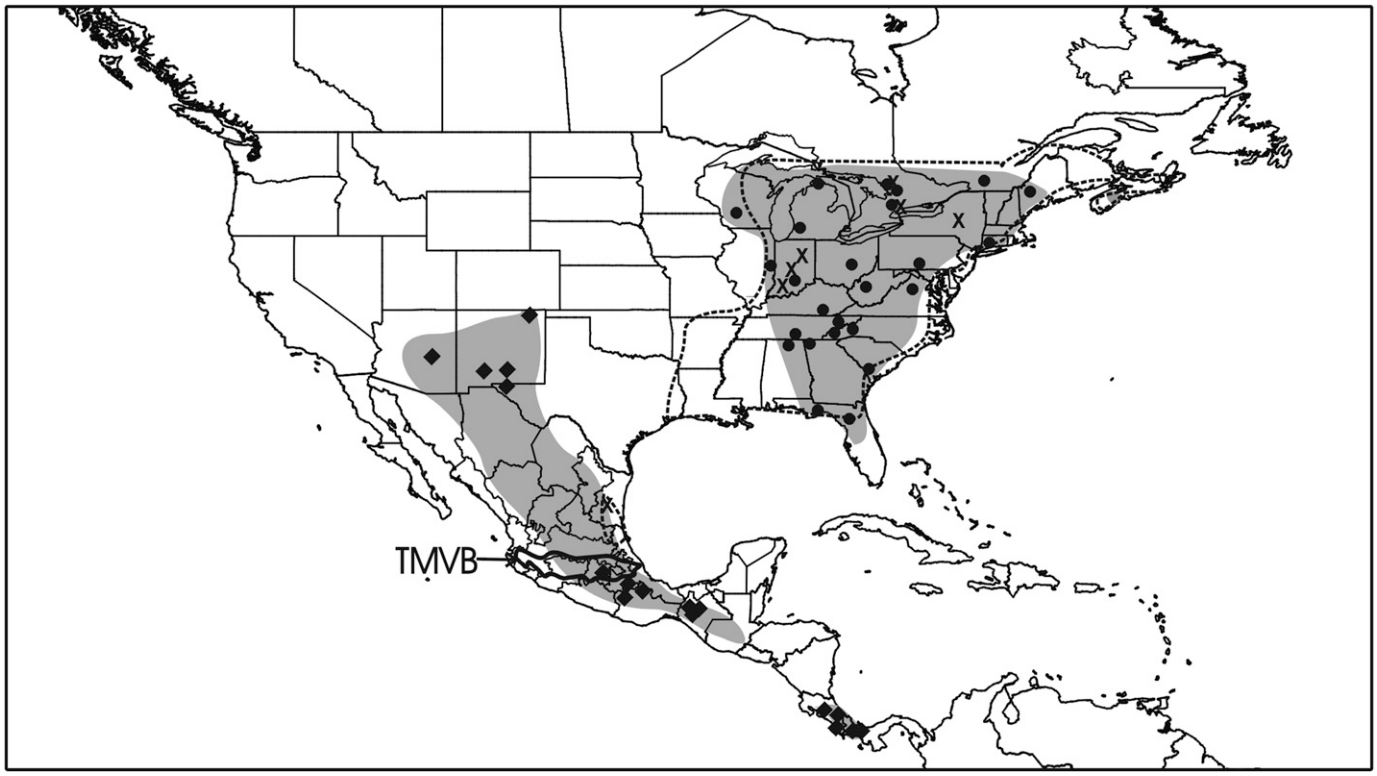

Fig. 1. Distribution of *Conopholis* (shaded) and *Epifagus* (dashed outlines) across their geographic ranges in eastern and western North America (modified from Haynes [1971] and Thieret [1969], respectively). The approximate position of sampling sites used in this study are indicated (for details, see Appendix 1). Circles represent sampling sites for populations of *C. americana*, diamonds represent those of *C. alpina*, while X symbols stand for *E. virginiana*. TMVB = Trans-Mexican volcanic belt.

Our overall research on *Conopholis* was undertaken with several major goals in mind: (1) to test the current taxonomic hypothesis by Haynes and assess the status of the proposed species; (2) to investigate the relationships between species and populations of this genus; (3) to conduct phylogeographic and population level analyses that may shed light on the postglaciation migration pattern(s) of the eastern North American species; (4) to investigate morphological character evolution within the genus and conduct morphometric analyses; and (5) to develop, in conjunction with the re-evaluation of the taxonomic characters used, a comprehensive phylogeny-based classification.

This present study is concerned with the first two of the aforementioned goals, i.e., develop a well-supported molecular phylogenetic hypothesis for *Conopholis* and assess the monophyly of the proposed species. To address these aims, we developed a multilocus molecular data set consisting of both plastid and nuclear DNA sequences. This is the first comprehensive molecular phylogenetic study of *Conopholis*, accounting for the entire geographic and morphological range of the species in North America.

MATERIALS AND METHODS

Taxon sampling—A total of 42 specimens representing the two presently recognized species of *Conopholis* were sampled in this study. A complete list of species, voucher information, DNA extraction numbers, and approximate locality of sampled populations is provided in Appendix 1. These accessions represent individuals spanning the entire geographic range of the genus (Fig. 1) and of all three traditionally described morphological types (*C. americana*, *C. alpina* var. *alpina*, and *C. alpina* var. *mexicana*). The names applied to these accessions follow the species delimitations by Haynes (1971), which emphasizes geographical distinctions between the species. However, given the morphological variation shown by *Conopholis*, we also use the phylogenetic species concept (PSC) approach. Unlike morphological or various mechanistic species concepts, the PSC is historically based (Baum and Donoghue, 1995) and uses the criteria of monophyly and exclusivity to define species (de Queiroz and Donoghue, 1990; Baum, 1992; Baum and Shaw, 1995).

In addition, *Epifagus virginiana*, represented here by seven accessions (Appendix 1), was chosen as the outgroup, based on the well-supported sister-group relationship of *Conopholis* and *Epifagus* resulting from previous broad molecular studies of Orobanchaceae (Nelson et al., 1999; Wolfe et al., 2005; Bennett and Mathews, 2006).

DNA extraction, amplification, and sequencing—Total genomic DNA was extracted from fresh, silica dried, or herbarium material using a modified hexadecyltrimethylammonium bromide (CTAB) technique from Doyle and Doyle (1987) and purified using Wizard minicolumns (Promega, Madison, Wisconsin, USA). The polymerase chain reaction (PCR) was used to obtain the double-stranded DNA fragments of interest. The plastid genome region containing the spacer between the *trnfM* (CAU) and *trnE* (UUC) exons (hereafter called *trnfM-E*) was amplified using the *trnfM*-r and *trnE* primers described by Doyle et al. (1992). These two genes are known to be more than 5 kb apart in *Nicotiana* (Wakasugi et al., 1998). However, in *Epifagus*, they are closer to each other due to significant intervening deletions in the plastid genome (dePamphilis and Palmer, 1990; Wolfe et al., 1992). Amplicons of the plastid *clpP* gene and its introns were generated via PCR following Stefanović et al. (2004). Primers used to amplify nuclear-encoded phytochrome A (*PHYA*) sequences in Orobanchaceae (Bennett and Mathews, 2006) were used to generate the first round of *Conopholis PHYA* sequences. Nested within these sequences, we subsequently designed a new set of primers, specific to *Conopholis*, targeting the *PHYA* intron 1 (PHYA F-a678f—GAGATGGTCCGTTTGATTGAG and PHYA R-a787—CGATGAAACATACTCCCACC). For all three of the regions, PCR reactions were carried out in 50-μL volumes with annealing temperature ranging between 50° and 55°C. Amplified products were cleaned by polyethylene glycol/NaCl precipitations or by Wizard minicolumns (Promega, Madison, Wisconsin, USA). To ensure accuracy, we sequenced both stands of cleaned PCR products (for *trnfM-E* and *PHYA,* using external primers only; for *clpP*, two internal primers were used in addition). Cleaned fragments were sequenced using the DYEnamic ET dye terminator sequencing kit (GE Healthcare, Baie-d'Urfe, Quebec, Canada) on an Applied Biosystems model 377 automated DNA sequencer (PE Biosystems, Foster City, California, USA). To screen nuclear sequences for polymorphisms, we cloned the cleaned *PHYA* PCR products into the pSTBlue-1 Acceptor vector (EMD Biosciences, San Diego, California, USA) and sequenced multiple clones. Sequence chromatograms were proofed, edited, and contigs assembled using the program Sequencher version 4.8. (Gene Codes Corp., Ann Arbor, Michigan, USA). All sequences generated in this study are deposited in GenBank (accessions HQ895589–HQ895712; Appendix 1).

Phylogenetic analyses—Sequences were aligned manually with the program Se-Al version 2.0a11 (Rambaut, 2002). The sequences were readily alignable among all ingroup accessions in both the plastid and nuclear matrices. Gaps in the alignments were treated as missing data. Indels were coded with the program Seqstate version 1.4.2 (Muller, 2005) using the procedure of Simmons and Ochoterna (2000) and appended to the respective sequence matrices. Phylogenetic analyses were conducted under a variety of distance- and character-based methods.

Network analyses—To investigate relationships among and within the species and populations of *Conopholis*, we initially constructed phylogenetic networks for each individual data set. The networks were constructed using a neighbor-net (NN) algorithm (Bryant and Moulton, 2004), as implemented in the program SplitsTree version 4.11.3 (Huson and Bryant, 2006). Prior to network analyses, sequences were corrected by imposing corresponding models of DNA evolution. The program ModelTest version 3.7 (Posada and Crandall, 1998) was used to determine the model of sequence evolution that fits best for each of the three data sets. The Akaike information criterion (AIC) method selected the F81 + I, HKY85 + I, and F81 models of DNA substitution for *trnfM-E*, *clpP*, and *PHYA Conopholis* matrices, respectively. The TVM + G model was chosen for the *clpP* matrix containing both ingroup and outgroup taxa (network not shown).

Parsimony analyses—Each data matrix was analyzed separately as well as in a single combined matrix using the program PAUP* version 4.0b10 (Swofford, 2002). In all of those analyses, heuristic searches for the most parsimonious (MP) trees were conducted using 1000 replicates with stepwise random sequence addition and tree-bisection-reconnection (TBR) branch swapping. All trees were saved during the search (MULTREES on). Support for relationships was inferred from nonparametric bootstrapping (Felsenstein, 1985) implemented in PAUP* by using 500 pseudoreplicates, each with 20 random sequence addition cycles, TBR branch swapping, and MULTREES option off (DeBry and Olmstead, 2000). Conflict between data sets was evaluated by visual inspection, looking for strongly supported yet conflicting tree topologies resulting from individual data matrices.

Bayesian analyses—Searches under the Bayesian criterion were done using the program MrBayes version 3.1.2 (Ronquist and Huelsenbeck, 2003) on the combined data set only (both with and without outgroup taxa). This combined plastid and nuclear data set was split into four partitions, three containing the *trnfM-trnE*, *clpP*, and *PHYA* sequences, respectively, and fourth with the combined indel characters. The models of sequence evolution as determined before were imposed for each sequence partition. The coded gaps (for *trnfM-trnE*, *clpP*, and *PHYA*) were included and analyzed separately from the sequence data. These characters were set to follow the Mk model (Lewis, 2001) with the possibility that some indels may be changing at different rates (Mk + G). Two runs starting from random trees were carried out. The Metropolis-coupled Markov chain Monte Carlo algorithm was used with four simultaneous chains set initially to one million generations and sampled every 100 generations. The likelihoods of the independent runs were considered indistinguishable when the average standard deviation of split frequencies was <0.01%, as suggested by Ronquist and Huelsenbeck (2003). To determine the burn-in cut-off point, we plotted the −ln likelihood values against generation time and discarded pre-asymptotic samples. The remaining data were analyzed in PAUP* where the 50% majority-rule consensus tree was constructed. With no significant difference between the two runs observed, we only report topologies and posterior probabilities based on pooled trees from the independent Bayesian analyses.

Evaluation of the rooting—In most phylogenies, the root node of a tree is usually determined extrinsically, by imposing outgroup(s). Alternatively, the enforcement of the molecular clock will result in ultrametric trees, rooted

intrinsically by the tree-building algorithm itself (Felsenstein, 2004). It is generally assumed that among closely related species, which tend to have similar metabolic rates, life histories, and generation times, the rates of evolution for a particular gene are likely to be comparable, resulting in a "local" molecular clock (Li, 1993; Sanderson, 2002). To identify the position of the root within *Conopholis* but without the use of outgroups and to compare the divergence times (rates) of *Conopholis* with those of *Epifagus*, we evaluated the molecular clock hypothesis using the likelihood ratio tests (LRT; Felsenstein, 1981; Goldman 1993). These tests were conducted on two data sets: the *Conopholis*-only matrix, including all available sequence data but without outgroups, as well as the *clpP*-only matrix, with *Epifagus* included. To assess whether the molecular clock could be applied to these data, we conducted maximum likelihood searches utilizing the models of DNA sequence evolution and parameter estimates as identified by Modeltest (Posada and Crandall, 1998; see above). For each data set, likelihood searches, with and without the molecular clock enforced, were performed using a two-stage strategy with PAUP*. First, the analyses involved 20 replicates with stepwise random taxon addition, TBR branch swapping saving no more than 10 trees per replicate, and MULTREES option off. The second round of analyses was performed on all trees in memory with the same settings except with the MULTREES option on. Both stages were conducted to completion or until 100 000 trees were found. The resulting likelihood estimates, with the clock imposed (H_{null}) and no enforcement of clock (H_{alt}), were then compared using the LRT with $N - 2$ degrees of freedom (where N is the number of operational units).

Testing of alternative topologies—Two alternative topologies, designed to investigate the monophyly of species as circumscribed traditionally, were constructed and their cost in parsimony assessed using PAUP* (Swofford, 2002). Constraining the monophyly of *C. americana* and that of *C. alpina* was done by using the combined data set (comprises all three sequence matrices) and including both ingroup and outgroup taxa (*Epifagus*). To statistically test and compare these alternatively enforced phylogenetic hypotheses with the optimal trees, we conducted two statistical tests. First, we conducted one-tailed Shimodaira–Hasegawa (SH) tests (Shimodaira and Hasegawa, 1999; Goldman et al., 2000) in PAUP* using 1000 replicates and full parameter optimization of the model. Second, we also carried out the less conservative, approximately unbiased tests (AU tests; Shimodaira, 2002). The *P*-values for the AU test were calculated in the program CONSEL version 0.1j (Shimodaira and Hasegawa, 2001), using 10 repetitions of multiscale bootstrapping, each consisting of 10 sets with 10 000 bootstrap replicates.

RESULTS

DNA regions and alignments—The characteristics of the three sequenced regions as well as statistics of MP trees derived from separate and combined analyses are described in Table 1.

Sequences for the *trnfM-E* region were obtained from all accessions of *Conopholis* and *Epifagus* used in this study. These sequences were relatively easy to align within *Conopholis* as well as within *Epifagus*. However, we could not achieve unambiguous alignment between the ingroup and outgroup taxa. When indels were coded for the *Conopholis trnfM-E* aligned data matrix, 22 additional binary characters were obtained. However, the indel characters arising from complex gaps in the alignment produced from single base repeats (>8) were excluded from the analyses, leaving 14 coded gaps used in subsequent analyses.

Most specimens of *Conopholis* (40 of 42) were readily amplifiable for *clpP*, as were all seven individuals of *E. virginiana*. Attempts to amplify the remaining two accessions of *C. alpina*, including partial fragment amplification with internal primers, were unsuccessful. This was probably due to the poor quality of DNA extracted from herbarium material. Unlike *trnfM-E*, *clpP* gene/introns sequences were easily aligned across both ingroup and outgroup samples. Gap characters were scored using the modified complex indel-coding method (Simmons and Ochoterna, 2000; Müller 2006) and resulted in 17 additional characters (coded for *Conopholis* only).

TABLE 1. Summary descriptions for sequences included in, and maximum parsimony trees derived from, individual and combined data sets of *Conopholis* and its close outgroup *Epifagus*.

Description	Plastid *clpP* [a]	Plastid *trnfM-trnE* [b]	Nuclear *PHYA* [b]	All combined [a]
Number of OTUs included	47	42	35	49
Sequence characteristics				
Analyzed length [c]	1667	475	479	2621
Number of coded gaps	24	14	3	41
Variable sites [d]	232	24	8	264
Parsimony informative sites [d]	209	16	5	230
Mean AT content	0.70	0.73	0.66	0.70
Tree characteristics				
Number of trees	711	120	2	930
Length	287	28	9	326
CI/RI	0.930/0.986	0.857/0.953	1/1	0.920/0.983

Note: CI, consistency index; RI, retention index; OTU, operational taxonomic unit

[a] Including outgroup taxa (*Epifagus*)

[b] Excluding outgroup taxa that could not be aligned with the ingroup accessions

[c] Excluding portions of the alignment spanning primer regions and ambiguously aligned regions

[d] Including coded gaps

Intron 1 sequences of the *PHYA* gene were readily obtained for all 35 individuals of *Conopholis* acquired from fresh or silica dried material. A direct sequencing approach yielded results without polymorphisms being observed in sequence trace chromatograms. A number of amplicons were cloned and up to 10 clones sequenced from each. In all cases, clones from the same individual produced identical sequences. Several attempts to amplify this single-copy nuclear region from the remaining seven *Conopholis* accessions, all of which were from herbarium sources, met with failure. In addition, amplification of outgroup samples for this nuclear region proved to be difficult, and the few *PHYA* sequences for *Epifagus* that did amplify could not be aligned with those of *Conopholis*. Indel coding for available sequences provided three additional binary characters.

Individual data set analyses—To explore the data, we initially conducted a number of distinct phylogenetic analyses on individual matrices using distance, parsimony, and Bayesian approaches.

The phylogenetic networks of individual data sets each revealed three distinct clusters, labeled informally A–C (Fig. 2). Group A consists of all accessions of *C. alpina* var. *mexicana* from Arizona, New Mexico, Texas, and north of the Trans-Mexican volcanic belt (TMVB). Group B contains samples of *C. alpina* var. *alpina* exclusively from Costa Rica and Panama. Group C includes populations from southern Mexico (*C. alpina* var. *alpina* from the states of Chiapas, Oaxaca, and Puebla), found interspersed within this group that otherwise contains all samples of *C. americana* from eastern North America.

Tree characteristics for MP searches are shown in Table 1. Topological agreement was found among the three separate analyses (trees not shown). Parsimony analyses of the individual data sets produced clades identical to their respective phylogenetic network already described (Fig. 2). Taking the results from all three separate analyses into account, both under network and tree approaches, we deemed these three matrices to show no significant topological incongruence and thus combined them into one data set.

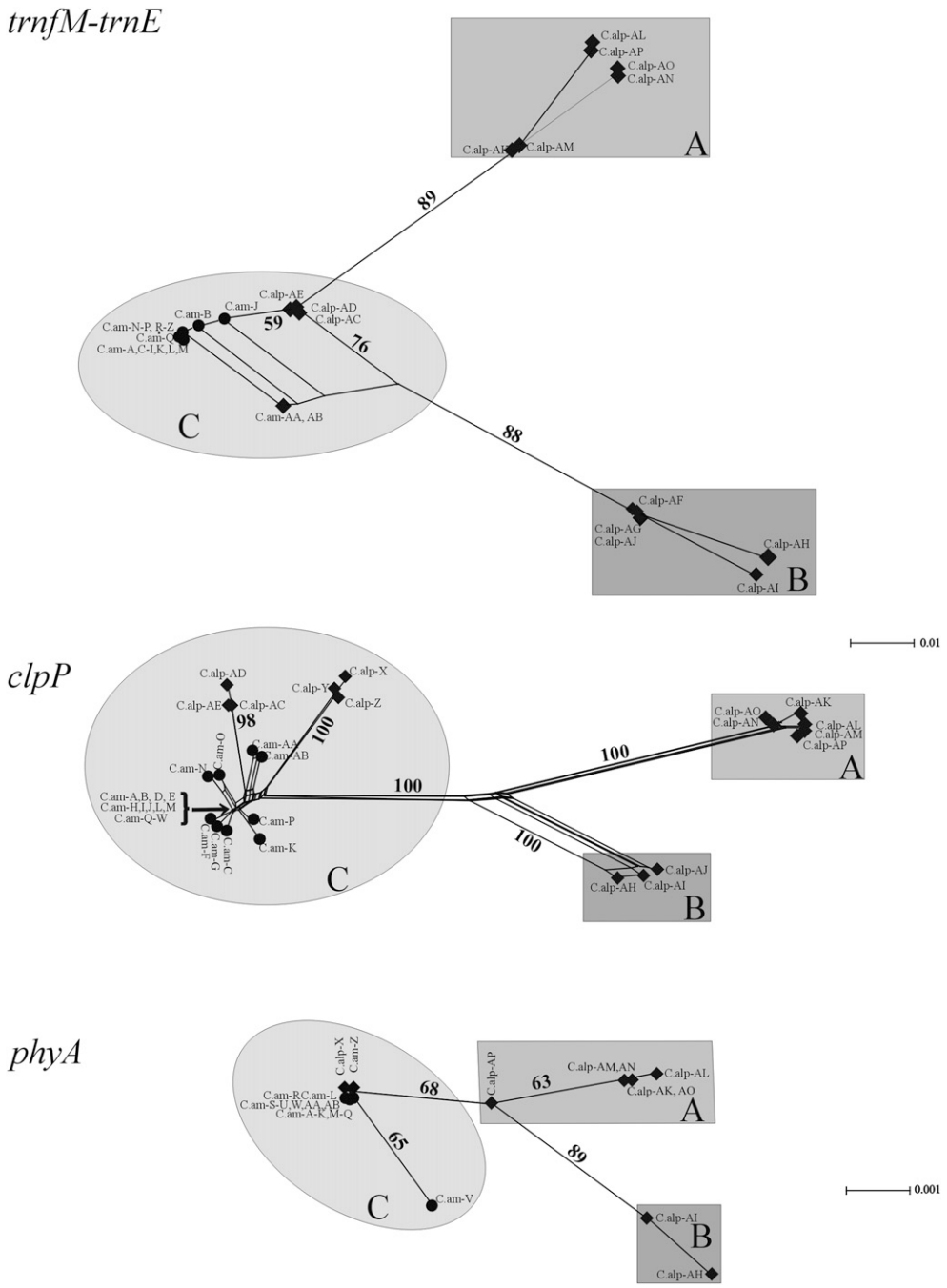

Fig. 2. Phylogenetic networks (neighbor-net split graphs) obtained from plastid (*trnfM-E* and *clpP*) and nuclear (*PHYA*) sequences of *Conopholis*. Major groups recovered and discussed in this study are highlighted and labeled A–C. Numbers represent bootstrap values ≥50% (1000 replicates). Closed circles represent individuals traditionally identified as *C. americana*; diamonds represent those of *C. alpina*. Taxon labels are indicated in Appendix 1. *Note:* plastid networks are at the same scale.

Phylogenetic analyses of combined data—Trees produced from the combined analyses had better resolution and overall support compared to those produced from individual analyses. The Bayesian analyses from each of the two runs starting from a random tree reached an asymptotic plateau no later than 150 000 generations, and all trees obtained prior to the plateau were excluded from the assemblage of a consensus tree. Figure 3 shows the majority-rule consensus tree resulting from the Bayesian analysis of all available data, combined plastid and nuclear sequences as well as coded gaps, obtained from *Conopholis* and *Epifagus* accessions. The topology is consistent with the results from the separate data set analyses using distance (Fig. 2) and parsimony (trees not shown). Identical backbone clades (outgroup plus three ingroup clusters labeled A–C) were recovered from the combined analyses, all with ≥90% bootstrap support (BS) in parsimony and with 1.0 posterior probability (PP) in Bayesian analysis (Fig. 3). The majority of the southwestern North American specimens included in this study

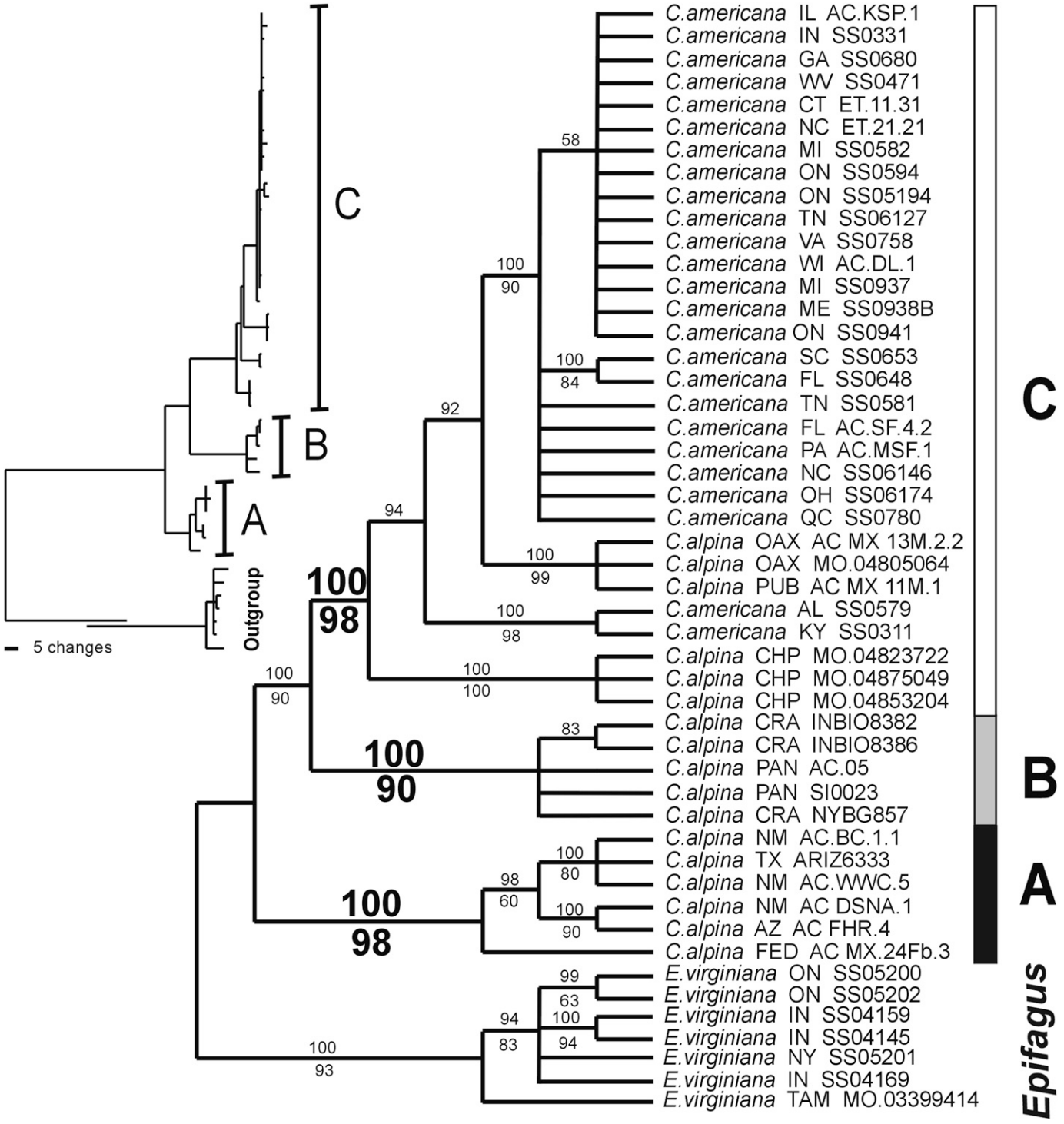

Fig. 3. Majority rule consensus tree resulting from the partitioned Bayesian analysis of the combined plastid (*clpP*, *trnfM-E*) and nuclear (*PHYA*) sequence data plus coded gaps showing phylogenetic relationships among and between populations of *Conopholis*. The tree is rooted using individuals from the sister genus *Epifagus* as outgroups. The MP search resulted in a strict consensus tree with almost identical topology (326 steps in length). Bayesian posterior probabilities are indicated above branches; parsimony bootstrap values (≥50%) are indicated below branches. Major clades recovered from analyses and discussed in this study are labeled A–C. Species names are followed by abbreviations of states/provinces in which they were collected and their respective DNA accession numbers (Appendix 1). Inset shows one of the equally parsimonious trees chosen to illustrate branch lengths.

(*C. alpina* var. *alpina* and *C. alpina* var. *mexicana*) are found in two distinct clades (A and B). Clade A contains all individuals from northern Mexico and the southwestern portion of the USA. Clade B comprises the lineage found in Costa Rica and Panama. The eastern North American individuals (*C. americana*) are all found within the third well-supported group, clade C. However,

as one of the most surprising results of this study, six accessions of *C. alpina* var. *alpina* sampled from the southern Mexican states of Chiapas, Oaxaca, and Puebla are also found nested within clade C, more closely related to *C. americana* than the other specimens of *C. alpina*. The seven accessions of *Epifagus* cluster with each other to form a separate outgroup clade, sister to *Conopholis*.

Parsimony analyses of the combined plastid and nuclear data resulted in essentially the same relationships between populations and species as described above. The inset in Fig. 3 shows a phylogram of one of the MP trees resulting from the analysis of the combined data matrix. This phylogram also depicts three substantial branches subtending the three major clades A–C, while individuals within each of these major clades appear more homogeneous with relatively shorter branch lengths.

Molecular clock and placement of the root—All phyogenetic analyses recovered the same major ingroup lineages within *Conopholis* (clades A–C; Fig. 3). In addition, the combined data analyses, which included outgroups, indicated that the first split occurred between clade A on one side and clades B plus C on the other. All of these backbone relationships received strong internal support. However, two problems are raised by the analysis. First, because *Conopholis* includes only three major clades and a root (Fig. 3), even a simple topological distortion, such as nearest-neighbor interchange (NNI), would result in trees with different placements of the root (e.g., clades A + B sister to clade C or clades A + C sister to clade B). Such competing rooting solutions could be caused by the artifact of attraction involving long outgroup and ingroup branches (Felsenstein, 1978). Second, we were able to achieve unambiguous alignment between ingroup and outgroup accessions only for the *clpP* region; yet, a significant proportion of missing data could lead to inaccurate phylogenetic reconstruction (Scotland et al., 2003).

To explore the influence of these potentially adverse conditions on our results, we reanalyzed the same combined data set but with the exclusion of *Epifagus* sequences, thereby eliminating the possibly misleading long outgroup branch as well as a relatively large number of missing data cells. Instead, we used the maximum likelihood to produce a rooted phylogeny. Results of the likelihood ratio test revealed that the null hypothesis (molecular clock enforced) could not be rejected ($df = 40$, $\chi^2_{obs} = 12.0824$, $p = 0.0001$). By enforcing the molecular clock, the ML search results in an ultrametric tree intrinsically rooted by all members of *C. alpina* var. *mexicana* (Appendix S1; see Supplemental Data online at http://www.amjbot.org/cgi/content/full/ajb.1000375/DC1). These ingroup-only analyses not only recovered the same major clades, with the same composition, but also resulted in identical inference of the root node within *Conopholis*, the same as when the outgroup *Epifagus* was included (compare Fig. 3 to online Appendix S1).

To test whether the molecular clock is in effect more broadly, between *Conopholis* and its sister *Epifagus*, the *clpP* data matrix (the most complete data set where the ingroup taxa were alignable with the outgroup) was analyzed using ML approach as well. Results of the likelihood ratio test revealed that the null hypothesis (molecular clock enforced) could not be rejected ($df = 45$, $\chi^2_{obs} = 22.9202$, $p = 0.0026$). The strict consensus tree resulting from ML analysis with the molecular clock enforced is shown in online Appendix S2.

Tests of alternate topologies—Given that neither species of *Conopholis* was found to be monophyletic on the optimal (unconstrained) trees, we wanted to determine the cost in parsimony and its significance when enforcing monophyly of species as circumscribed by Haynes (1971). These tests were done using the combined data matrix, containing both *Conopholis* and *Epifagus* sequences. When the topologies were constrained so that all eastern North American individuals were monophyletic (i.e., the monophyly of *C. americana* s.s.), this resulted in trees of 327 steps, only one step longer than the optimal tree (Fig. 3). Not surprisingly, this result proved not to be significantly different from the optimal tree (SH test $P = 0.782$; AU test $P = 0.353$). However, when the topology was constrained so that *C. alpina* was monophyletic, this produced trees of 342 steps, 16 steps longer than the MP tree. This result provides strong evidence in support of the optimal topologies as it rejects the monophyly of *C. alpina* as a significantly worse solution (SH test $P = 0.004$; AU test $P = 2 \times 10^{-4}$).

DISCUSSION

This work represents the only fine-scale molecular phylogenetic study for *Conopholis*. It is based on a combination of plastid and nuclear DNA sequences obtained from individuals sampled across the entire taxonomic and geographic range of the two presently recognized species. The resulting phylogenetic inferences are robust and show significant support for the composition and relationships between major clades in the tree. Figure 4 summarizes our present understanding of phylogenetic relationships among populations of *Conopholis*, the relationship between the traditional taxonomy and putative phylogenetic classification suggested here, and a biogeographic scenario is proposed to explain present-day distribution of the genus.

Phylogenetic and taxonomic implications—Prior to the publication of the monograph by Robert R. Haynes (1971), upward of five species of *Conopholis* had been described: *C. americana*, *C. alpina*, *C. mexicana*, *C. panamensis*, *and C. sylvatica*. However, due to the severe reduction in morphology of parasitic plants, there is a limited number of characters that can be potentially relied upon to differentiate between these species, creating uncertainty as to the number of species in the genus in early floristic treatments, ranging from one to four (e.g., Beck-Mannagetta, 1930; Small, 1933; Fernald, 1950; Gleason, 1952). Following his seminal work, Haynes (1971) concluded that only two of these taxa warranted recognition at the species level, *C. americana* and *C. alpina*, with the latter being further subdivided into two varieties (var. *alpina* and var. *mexicana*). Specifically, after studying the relevant type specimens, he concluded that the individuals assigned to *C. alpina*, *C. sylvatica*, and *C. panamensis* represented only intraspecific variability and did not warrant separation in three different species. Therefore, these taxa were reduced to a single species, to which the specific epithet of *C. alpina* was assigned based on priority.

Regardless of the data set used or phylogenetic methodology employed, none of our analyses lend support for the strict subdivision of the genus into the two presently recognized species. Instead, molecular data presented in the present study provide strong evidence for three distinct lineages within *Conopholis* (Figs. 2, 3) having various degrees of overlap with previously proposed taxa (Fig. 4). In addition, these analyses also reveal substantial branches subtending each of three clades (see insets in Fig. 3 and online Appendices S1 and S2). These three branches are comparable in length to each other, and in all three

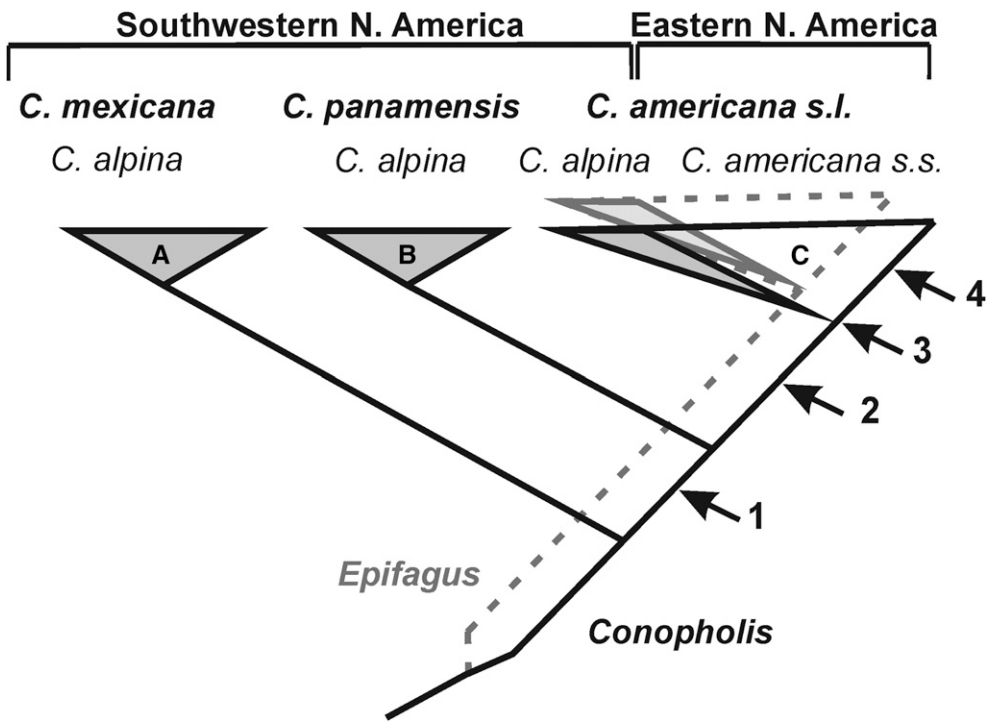

Fig. 4. Schematic overview of the evolutionary hypothesis for *Conopholis* derived from plastid and nuclear sequence data. Correspondence between the major clades (A–C) recovered in this study, the traditional classification, new classification (in bold) presented here, and current geographic distribution is indicated. Shaded clade (dotted lines) represents the sister genus *Epifagus* and its current distribution. See Discussion for full description of proposed four-step biogeographic scenario (1–4).

cases they are substantially longer than the branch lengths observed within the major clades. Assuming that these relative lengths are indicative of the overall amount of genetic diversification, this suggests that the three lineages have been reproductively isolated from each other for a period of time long enough to allow them to accumulate greater genetic differences among lineages compared to within. In aggregate, the composition of the clades as well as the branches subtending those clades lends support for the recognition of three distinct lineages within *Conopholis*, possibly at the species level (Fig. 4).

Clade A corresponds entirely to *C. alpina* var. *mexicana*, containing all sampled individuals found in the southwestern portion of the USA and north of the TMVB. Therefore, taxonomically this clade corresponds to *C. mexicana*, a taxon originally recognized as a separate species by Watson (1882). He deemed it to be distinguished from *C. americana* by its longer and more rigid lanceolate acuminate scales as well as having a larger corolla and a less deeply toothed calyx. Aside from this clade, the remaining individuals traditionally assigned to *C. alpina* are also found in clades B and C. Clade B is composed of specimens occurring solely in Costa Rica and Panama, although according to the scheme of Haynes (1971), individuals obtained from this geographic region should be more closely related to those found in clade A. Instead, clade B may correspond to another previously described species, *C. panamensis* (Woodson and Seibert, 1939), morphologically distinct from both *C. mexicana* and *C. americana*. According to its original description, *C. panamensis* can be distinguished from *C. mexicana*, a taxon that is geographically in a relatively close proximity to the Central American populations, based on its shallow and broadly obtuse calyx. In addition, this putative species can be distinguished from both *C. mexicana* and *C. americana* by its seeds that are about half the size of those in the other two taxa. However, it shares with *C. americana* broad bracts concealing its calyx, while the loss of style in fruit resembles that of *C. mexicana*. The morphological distinction between the Central American populations and those found in northern Mexico and the southwestern USA is consistent with the two genetically distinct lineages of *C. alpina* (clades A and B; Figs. 2 and 3) recovered by phylogenetic networks and tree approaches. Also, for clade B to correspond to the description of *C. alpina* var. *alpina*, it would have to contain individuals sampled from the southern Mexican states. However, the six accessions of *C. alpina* sampled from Chiapas, Oaxaca, and Puebla are found interspersed within clade C. Excluding these six accessions from southern Mexico, clade C otherwise contains all sampled individuals of *C. americana* from eastern North America. While the monophyly of *C. americana* is not strictly recovered according to the optimal trees, the possibility of its monophyly added only one step to the MP trees and could not be rejected with confidence by the SH and AU tests.

Given the composition of clade C, we hypothesize that a comprehensive re-evaluation of morphological characters within this genus will likely reveal morphological features shared between the members of *C. americana* and those of *C. alpina* distributed in the southern Mexican states. The observation that no single character can be relied upon to distinguish the eastern from western species as noted by Haynes (1971, p. 252) may thus be explained by the fact that some *Conopholis* populations found in Mexico are actually disjunct members of *C. americana* rather than of *C. alpina*, as expected by their distribution. A similar case of intraspecific disjunction is indeed observed in

Epifagus virginiana. This species, sister to *Conopholis*, also occurs predominantly throughout eastern North America, but it has disjunct populations located in known relict temperate forests in Mexico such as Rancho del Cielo (Thieret, 1969), mirroring closely the distribution of individuals found in clade C (Fig. 4). Other Mexican disjunct lineages that are considered conspecifics with their eastern United States counterparts include *Nyssa sylvatica* (Miranda and Sharp, 1950) and *Fagus grandifolia* (Morris et al., 2010). Also, two members of the *Corallorhiza striata* species complex (*C. bentleyi* and *C. striata* var. *invulata*) show a similar pattern and are presumed relicts left from a once broader distribution of the ancestor of that clade (Barrett and Freudenstein, 2009). However, contrary to these other cases, which showed no or little genetic variation in the Mexican accessions studied (e.g., *Fagus grandifolia*; Morris et al., 2010), we observed substantial variation among the disjunct Mexican samples of *C. alpina* found in clade C, even though our sampling of those populations is limited (Fig. 3).

The incomplete lineage sorting of ancestral polymorphisms (i.e., deep coalescence) could be seen as an alternative explanation for the phylogenetic patterns observed in our study, in particular regarding the nonmonophyly of *C. alpina*. However, several lines of evidence suggest this as an unlikely scenario. First, two independent sources of data were used, plastid and nuclear DNA sequences, resulting in congruent gene trees, both producing three separate and well-supported groups (Figs. 2 and 3). Second, because the plastid genome has a significantly smaller effective population size compared to nuclear loci (Moore, 1995), the phylogenetic relationships resulting from the use of plastid regions have a higher probability of a faster coalescence times, leading to more rapid eliminations of any polymorphisms. Third, in the case of *C. alpina* whose members are found in all three of these well-differentiated clades (as evidenced by the branch lengths subtending those clades; see inset, Fig. 3) the polymorphisms would have had to persist through multiple cladogenesis events. Phylogenetic analyses of additional, independently inherited nuclear sequence data, as well as consideration of faster-evolving markers, such as microsatellite loci, will help us to further test our current taxonomic and phylogeographic hypotheses (Hare, 2001).

Historical biogeography—Taxa such as *Conopholis* that exhibit an east–west geographic disjunction in North America are considered to be either: (1) tertiary relicts of the mixed broadleaf (mesophytic) forest, or (2) disruptions in the continuous ranges caused by Pleistocene glaciations, or (3) the result of more recent long-distance dispersal events (Graham, 1964; Wood, 1972, Graham, 1993; Soltis et al., 2006). Oaks, the hosts of *Conopholis*, are documented to have been part of the broadleaf forest that was found across North America from the late Eocene through the Miocene (40–5 Ma; Braun, 1947; Axelrod, 1983). The range of *Conopholis* is thought to have been continuous during the latter portion of this time period with that of oaks. However, with the appearance of widespread prairie vegetation and the aridification of the midcontinental North America during the Pliocene (5–2 Ma), a major east–west disruption of the broadleaved deciduous forest was created (Graham, 1993).

Given the composition of and rooted relationships among three separate lineages inferred within *Conopholis* (Fig. 3; online Appendices S1 and S2), we propose a four-step biogeographic scenario to explain the current distribution of the genus and place it in a historical context (Fig. 4). The first, and therefore the oldest, split seems to have occurred between clade A and the rest of the genus, with the Trans-Mexican volcanic belt functioning as a barrier. The formation of this belt began in the late Miocene (Ferrari et al., 1999) and continues today, forming the tallest mountain range in Mexico that runs from east to west in the central region of the country (Rzedowski, 1978). This mountain range is a recognized center for biodiversity (*Quercus*, Nixon, 1993; *Pinus*, Styles, 1993) and is established as a known vicariant barrier for a variety of organisms (insects, Halffter, 1964, 1976; mosses, Delgadillo, 1987). For *Conopholis*, this volcanic belt can be seen as an effective barrier to migration and gene flow, creating a north–south divide in Mexico and effectively separating populations to the north of the belt from the rest of *Conopholis* (clade A; Figs. 2, 3).

The second step is the separation of high-mountain populations from Costa Rica and Panama from those that lay further north and east, around the Gulf of Mexico. Members of clade B are geographically isolated from the nearest population of *Conopholis* reported to be in Guatemala (Haynes, 1971). There are no known populations occurring in Honduras, El Salvador, or Nicaragua. Hence, these populations have presumably existed in isolation from other members of the genus for an extended period of time, resulting in accumulation of genetic differences between them and their nearest relatives along the Gulf Coast.

The third split in *Conopholis* can be explained by the east–west North American disjunction. In addition to the aridification of midcentral North America during the Pliocene described above, there is also the possibility that the range of *Quercus* and *Conopholis* remained continuous along the Gulf Coast during the last glacial period occurring in the Pleistocene, 110 000-10 000 yr ago (Jackson et al., 2000). However, the harsh climate during the Pleistocene glaciations can be assumed to have eliminated the north-central portion of the range (Wood, 1972), and in so doing, created an east–west divide along the Gulf Coast. This geographic divide of more than 1400 km in the case of *Conopholis* resulted in the genetic differentiation between populations found today in northeastern North America from those in southern Mexico (Figs. 1, 4).

Finally, the fourth step would involve repeated range expansion and contraction of eastern North American *Conopholis* populations following the glaciation minima and maxima. In eastern North America during the peak Pleistocene glaciation, plants and animals presumably survived primarily in several glacial refugia located in the southern portions of the United States along the Gulf coast (Pielou, 1991). Fossil data suggests that pockets of hardwood forests existed in the Lower Mississippi Valley during the last glacial maximum forming a glacial refugium in the southern USA for temperate taxa (Delcourt and Delcourt, 1984). As the ice retreated, populations migrated northwards to their ranges present day. The "Southern Refugia hypothesis" postulates higher diversity in the southern nonglaciated regions and loss of this diversity by populations moving northwards (Hewitt, 1996, 2000; Petit et al., 2002; McLachlan et al., 2005). The results of our present analyses offer an initial support for this hypothesis in *Conopholis*. Specifically, within clade C, we observed greater diversity in populations collected from the southern Mexican and USA states (e.g., Chiapas, Oaxaca, Puebla, Alabama, and Kentucky) compared to the central and northern parts of its range (Fig. 3 and online Appendices S1 and S2). To gain a better understanding of relationships within clade C as well as the amount and geographic distribution of genetic diversity within this lineage, we need a much more dense sampling strategy, in combination with faster-evolving markers.

Given the present distribution of *Epifagus*, the last two steps (steps three and four) appear to apply equally well on populations of this genus, sister to *Conopholis*. Although not as dramatic as in *Conopholis*, an east–west North American geographic disjunction clearly exists in *Epifagus* (Thieret, 1969; Fig. 1), and this divide is also recovered through molecular phylogenetic analyses (Fig. 3 and online Appendix S2). The representative of the *Epifagus* population found in Mexico is genetically distinct from those located in northeastern North America. In addition, from the molecular clock trees (inset, online Appendix S2), we can deduce that populations of *Epifagus* have very similar diversification times (rates) compared to those of *C. americana* in eastern North America. The branches subtending populations within these clades are very short and indicative of low levels of sequence divergence within these two clades, particularly in the northern range, compatible with more comprehensive results from plastid and microsatellite data in *Epifagus* (Tsai and Manos, 2010).

A more integrative approach to historical phylogeny-based biogeography has been strongly advocated (e.g., Donoghue and Moore, 2003 and references therein), in particular regarding the need for an explicit incorporation of temporal information in such studies (Ree et al., 2005). However, primarily because Orobanchaceae as a family has no fossils that can be used to set a reference date (Cronquist, 1988), we feel that efforts to estimate the absolute timing of the diversification of lineages within *Conopholis* remain premature. Nevertheless, an initial attempt was made by Wolfe et al. (2005) to put a timeframe on the origin of Orobanchaceae within a broader phylogenetic context, by taking estimates for the Lamiales crown-clade age based on sequence data (71–74 Ma; Wikström et al., 2001) and fossil data (37 Ma for Oleacaceae; Magallón et al., 1999), and averaging these estimates to create a reference node time of 55.5 Ma. Wolfe et al. (2005, p. 125) acknowledged the limitations of their approach, stating that "not having a fossil for Orobanchaceae, and calculating divergence times based on an average estimated age of Lamiales […] means that our inference of divergence times should be considered as a baseline for future studies." Aside from these calibration issues, their divergence estimates within Orobanchaceae were deduced from a phylogenetic tree with relatively poor support for many of the backbone relationships in this family and confidence intervals were not provided for any of the calculated point estimates, making it altogether difficult to further critically evaluate those results or use them here.

Conclusions—Altogether, these analyses reveal three distinct lineages lending support to the possibility of there being three species within the genus. A fine-scale morphometric analysis is needed to determine if there are morphological features that could further corroborate our molecular results. In addition, research with multiple individuals per population should be conducted to provide more accurate estimates of population-level genetic diversity within *C. americana* s.l. and to draw conclusions about its postglacial migration.

LITERATURE CITED

ALBACH, D. C., K. YAN, S. R. JENSEN, AND H.-Q. LI. 2009. Phylogenetic placement of *Triaenophora* (formerly Scrophulariaceae) with some implications for the phylogeny of Lamiales. *Taxon* 58: 749–756.

APG III [ANGIOSPERM PYLOGENY GROUP III]. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161: 105–121.

AXELROD, D. I. 1983. Biogeography of oaks in the Arcto-Tertiary province. *Annals of the Missouri Botanical Garden* 70: 629–657.

BAIRD, V. W., AND J. L. RIOPEL. 1986a. The developmental anatomy of *Conopholis americana* (Orobancaceae) seedlings and tubercles. *Canadian Journal of Botany* 64: 710–717.

BAIRD, V. W., AND J. L. RIOPEL. 1986b. Life history studies of *Conopholis americana* (Orobanchaceae). *American Midland Naturalist* 116: 140–151.

BARRETT, C. F., AND J. V. FREUDENSTEIN. 2009. Patterns of morphological and plastid DNA variation in the *Corallorhiza striata* species complex (Orchidaceae). *Systematic Botany* 34: 496–504.

BAUM, D. A. 1992. Phylogenetic species concepts. *Trends in Ecology & Evolution* 7: 1–2.

BAUM, D. A., AND M. J. DONOGHUE. 1995. Choosing among alternative "phylogenetic" species concepts. *Systematic Botany* 20: 560–573.

BAUM, D. A., AND K. L. SHAW. 1995. Genealogical perspectives on the species problem. *In* P. C. Hoch and A. G. Stephenson [eds.], Experimental and molecular approaches to plant biosystematics, 289–303. Missouri Botanical Garden Press, St. Louis, Missouri, USA.

BECK-MANNAGETTA, G. 1930. Orobanchaceae. *In* A. Engler [ed.], D. Pflanzenrich IV. 261, Heft 96, 1–348.

BENNETT, J. R., AND S. MATHEWS. 2006. Phylogeny of the parasitic plant family Orobanchaceae inferred from phytochrome A. *American Journal of Botany* 93: 1039–1051.

BRAUN, E. L. 1947. Development of the deciduous forests of eastern North America. *Ecological Monographs* 17: 211–219.

BRYANT, D., AND V. MOULTON. 2004. Neighbor-net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution* 21: 255–265.

COLWELL, A. E. 1994. Genome evolution in a non-photosynthetic plant, *Conopholis americana*. Ph.D. dissertation, Washington University, St. Louis, Missouri, USA.

CRONQUIST, A. 1988. The evolution and classification of flowering plants, 433. New York Botanical Garden, Bronx, New York, USA.

DEBRY, R. W., AND R. G. OLMSTEAD. 2000. A simulation study of reduced tree-search effort in bootstrap resampling analysis. *Systematic Biology* 49: 171–179.

DELCOURT, P. A., AND H. R. DELCOURT. 1984. Ice age heaven for harwoods. *Natural History* 93: 22–25.

DELGADILLO, M. C. 1987. Moss distribution and the phytogeographical significance of the Neovolcanic Belt in Mexico. *Journal of Biogeography* 14: 69–78.

DEPAMPHILIS, C. W., AND J. D. PALMER. 1990. Loss of photosynthetic and chlororespiratory genes from the plastid fenome of a parasitic flowering plant. *Nature* 348: 337–339.

DE QUEIROZ, K., AND M. J. DONOGHUE. 1990. Phylogenetic systematics or Nelson's version of cladistics? *Cladistics* 6: 61–75.

DONOGHUE, M. J., AND B. R. MOORE. 2003. Toward an integrative historical biogeography. *Integrative and Comparative Biology* 43: 261–270.

DOYLE, J. J., J. I. DAVIS, R. J. SORENG, D. GARVIN, AND M. J. ANDERSON. 1992. Chloroplast DNA inversions and origin of the grass family (Poaceae). *Proceedings of the National Academy of Sciences, USA* 89: 7722–7726.

DOYLE, J. J., AND L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* 27: 401–410.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution* 17: 368–376.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies—An approach using the bootstrap. *Evolution* 39: 783–791.

FELSENSTEIN, J. 2004. Inferring phylogenies. Sinauer, Sunderland, Massachusetts, USA.

FENG, Y., S. H. OH, AND P. S. MANOS. 2005. Phylogeny and historical biogeography of the genus *Platanus* as inferred from nuclear and chloroplast DNA. *Systematic Botany* 30: 786–799.

FERNALD, M. I. 1950. Grey's manual of botany, 8th ed. American Book Co., New York, New York, USA.

FERRARI, L., M. LÓPEZ-MARTÍNEZ, G. AGUIRRE-DÍAZ, AND G. CARRASCO-NÚÑEZ. 1999. Space-time patterns of Cenozoic arc volcanism in central Mexico: From the Sierra Madre Occidental to the Mexican Volcanic Belt. *Geology* 27: 303–306.

GLEASON, H. A. 1952. The new Britton and Brown illustrated flora of the northeastern United States and adjacent Canada, vol. 3. Lancaster Press, Lancaster, Pennsylvania, USA.

GOLDMAN, N. 1993. Statistical tests of models of DNA substitution. *Journal of Molecular Evolution* 36: 182–198.

GOLDMAN, N., J. P. ANDERSON, AND A. G. RODRIGO. 2000. Likelihood-based tests of topologies in phylogenetics. *Systematic Biology* 49: 652–670.

GÓMEZ, L. D. 1980. Notes on the biology of Central American Orobanchaceae. *Besnia* 17: 389–396.

GRAHAM, A. 1964. Origin and evolution of the biota of Southeastern North America: Evidence from the fossil plant record. *Evolution* 18: 571–585.

GRAHAM, A. 1993. History of the vegetation: Cretaceous (Maastrichtian)—Tertiary flora of North America, 57–70. Oxford University Press, New York, New York, USA.

HALFFTER, G. 1964. La entomofauna americana, ideas acerca de su origen y distribución. *Folia Entomológica Mexicana* 6: 1–108.

HALFFTER, G. 1976. Distribución de los insectos en la zona de transición mexicana. Relaciones con la entomofauna de Norteamérica. *Folia Entomológica Mexicana* 35: 1–64.

HARE, M. P. 2001. Prospects for nuclear phylogeography. *Trends in Ecology & Evolution* 16: 700–706.

HAYNES, R. R. 1971. A monograph of the genus *Conopholis* (Orobanchaceae). *SIDA* 4: 246–264.

HEWITT, G. M. 1996. Some genetic consequences of ice ages and their role in divergence and speciation. *Biological Journal of the Linnaean Society* 58: 247–276.

HEWITT, G. M. 2000. The genetic legacy of Quaternary ice ages. *Nature* 405: 907–913.

HUSON, D. H., AND D. BRYANT. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23: 254–267.

JACKSON, S. T., R. S. WEBB, K. H. ANDERSON, J. T. OVERPECK, T. WEBB III, J. W. WILLIAMS, AND B. C. S. HANSEN. 2000. Vegetation and environment in eastern North America during the last glacial maximum. *Quaternary Science Reviews* 1 : 489–508.

JENSEN, S. R., H.-Q. LI, D. C. ALBACH, AND C. H. GOTFREDSEN. 2008. Phytochemistry and molecular systematics of *Triaenophora rupestris* and *Oreosolen watii* (Scrophulariaceae). *Phytochemisty* 69: 2162–2166.

KUIJT, J. 1969. The biology of parasitic flowering plants. University of California Press, Berkeley, California, USA.

LEWIS, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50: 913–925.

LI, W. H. 1993. So what about the molecular clock hypothesis? *Current Opinion in Genetics & Development* 3: 896–901.

MAGALLÓN, S., P. R. CRANE, AND P. S. HERENDEEN. 1999. Phylogenetic pattern, diversity, and diversification of eudicots. *Annals of the Missouri Botanical Garden* 86: 297–372.

MANOS, P. S., Z.-K. ZHOU, AND C. H. CANNON. 2001. Systematics of Fagaceae: Phylogenetic tests of reproductive trait evolution. *International Journal of Plant Sciences* 162: 1361–1379.

MCLACHLAN, J. S., J. S. CLARK, AND P. S. MANOS. 2005. Molecular indicators of tree migration capacity under rapid climate change. *Ecology* 86: 2088–2098.

MIRANDA, F., AND A. J. SHARP. 1950. Characteristics of the vegetation in certain temperate regions of eastern Mexico. *Ecology* 31: 313–333.

MOORE, W. S. 1995. Inferring phylogenies from mtDNA variation: Mitochondrial-gene trees versus nuclear-gene trees. *Evolution* 49: 718–726.

MORRIS, A. S., C. H. GRAHAM, D. E. SOLTIS, AND P. S. SOLTIS. 2010. Reassessment of phylogeographical structure in eastern North American trees using Monmonier's algorithm and ecological niche modelling. *Journal of Biogeography* 37: 1657–1667.

MÜLLER, K. 2005. SeqState: Primer design and sequence statistics for phylogenetic DNA datasets. *Applied Bioinformatics* 4: 65–69.

MÜLLER, K. 2006. Incorporating information from length-mutational events into phylogentic analysis. *Molecular Phylogenetics and Evolution* 38: 667–676.

MYLECRAINE, K. A., J. E. KUSER, P. E. SMOUSE, AND G. L. ZIMMERMANN. 2004. Geographic allozyme variation in Atlantic white-cedar, *Chamaecyparis thyoides* (Cupressaceae). *Canadian Journal of Forest Research* 34: 2443–2454.

NICKRENT, D. L. 2010. The parasitic plant connection [online]. Website http://www.parasiticplants.siu.edu/ [accessed 1 June 2010].

NIXON, K. C. 1993. El género *Quercus* en México. *In* T. P. Ramamoorthy, R. Bye, A. Lot, and J. Fa [eds.], Diversidad biológica de México: Orígenes y distribución, 435–448. Instituto de Biología, Universidad Nacional Autónoma de México, México, D. F., México.

OLMSTEAD, R. G., C. W. DEPAMPHILLIS, A. D. WOLFE, N. D. YOUNG, W. J. ELISONS, AND P. A. REEVES. 2001. Disintegration of the Scrophulariaceae. *American Journal of Botany* 88: 348–361.

PETIT, R. J., S. BREWER, S. BORDACS, K. BURG, R. CHEDDADE, E. COART, J. COTTRELL, ET AL. 2002. Identification of refugia and postglacial colonization routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management* 156: 49–74.

PIELOU, E. C. 1991. After the ice age: The return of life to glaciated North America. University of Chicago Press, Chicago, Illinois, USA.

POSADA, D., AND K. A. CRANDALL. 1998. ModelTest: Testing the model of DNA substitution. *Bioinformatics* 14: 817–818.

RAMBAUT, A. 2002. Se-Al sequence alignment editor, version 2.0a11 [computer program]. University of Oxford, Oxford, UK.

REE, R. H., B. R. MOORE, C. O. WEBB, AND M. J. DONOGHUE. 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution* 59: 2299–2311.

RONQUIST, F., AND J. P. HUELSENBECK. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.

RZEDOWSKI, J. 1978. Vegetación de México. Editorial Limusa, México, D.F., México.

SANDERSON, M. J. 2002. Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. *Molecular Biology and Evolution* 19: 101–109.

SCOTLAND, R. W., R. G. OLMSTEAD, AND J. R. BENNETT. 2003. Phylogeny reconstructions: The role of morphology. *Systematic Biology* 52: 539–548.

SHIMODAIRA, H. 2002. An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* 51: 492–508.

SHIMODAIRA, H., AND M. HASEGAWA. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* 16: 1114–1116.

SHIMODAIRA, H., AND M. HASEGAWA. 2001. CONSEL: For assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17: 1246–1247.

SIMMONS, M. P., AND H. OCHOTERNA. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49: 369–381.

SMALL, J. K. 1933. Manual of the southeastern flora. University of North Carolina Press, Chapel Hill, North Carolina, USA.

SOLTIS, D. E., A. B. MORRIS, J. S. MCLACHLAN, P. S. MANOS, AND P. S. SOLTIS. 2006. Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology* 15: 4261–4293.

STEFANOVIĆ, S., D. W. RICE, AND J. D. PALMER. 2004. Long branch attraction, taxon sampling, and the earliest angiosperms: *Amborella* or monocots? *BMC Evolutionary Biology* 4: 35.

STYLES, B. T. 1993. Genus *Pinus*: A Mexican purview. *In* T. P. Ramamoorthy, R. Bye, A. Lot, and J., Fa [eds.], Biological diversity of Mexico: Origins and distribution, 397–420. Oxford University Press, New York, New York, USA.

SWOFFORD, D. L. 2002. Phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sinauer, Sunderland, Massachusetts, USA.

THIERET, J. W. 1969. Notes on *Epifagus. Castanea* 34: 397–402.

TSAI, Y.-H. E., AND P. S. MANOS. 2010. Host density drives the postglacial migration of the tree parasite, *Epifagus virginiana. Proceedings of the National Academy of Sciences, USA* 107: 17035–17040.

WATSON, S. 1882. Contributions to American botany. I. List of plants from southwestern Texas and northern Mexico. II. Gamopetalae to Acotyledons. *Proceedings of the American Academy of Arts and Sciences* 18: 96–191.

WAKASUGI, T., M. SUGITA, T. TSUDZUKI, AND M. SUGIURA. 1998. Updated gene map of tobacco chloroplast DNA. *Plant Molecular Biology Reporter* 16: 231–241.

WIKSTRÖM, N., V. SAVOLAINEN, AND M. W. CHASE. 2001. Evolution of the angiosperms: Calibrating the family tree. *Proceedings. Biological Sciences* 268: 2211–2220.

WOLFE, A. D., C. P. RANDLE, L. LIU, AND K. E. STEINER. 2005. Phylogeny and biogeography of Orobanchaceae. *Folia Geobotanica* 40: 115–134.

WOLFE, K. H., C. W. MORDEN, AND J. D. PALMER. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proceedings of the National Academy of Sciences, USA* 89: 10648–10652.

WOOD, C. E. J. 1972. Morphology and phytogeography: The classical approach to the study of disjunctions. *Annals of the Missouri Botanical Garden* 59: 107–124.

WOODSON, R. E. JR., AND R. J. SEIBERT. 1939. Contributions towards a flora of Panama II. Miscellaneous collections during 1936–1938. *Annals of the Missouri Botanical Garden* 25: 823–840.

XIA, Z., Y.-Z. WANG, AND J. F. SMITH. 2009. Familial placement and relations of *Rehmannia* and *Triaenophora* (Scrophulariaceae s.l.) inferred from five gene regions. *American Journal of Botany* 96: 519–530.

YOUNG, N. D., K. E. STEINER, AND C. W. DEPAMPHILIS. 1999. The evolution of parasitism in Scrophulariaceae/Orobanchaceae plastid gene sequences refute an evolutionary transition series. *Annals of the Missouri Botanical Garden* 86: 876–893.

APPENDIX 1. Taxa, DNA accession numbers, voucher information, locality from where specimen were collected, geographic coordinates, labels for names used in text, and GenBank accession numbers for sequences used in this study.

Species	DNA accession [a]/voucher [b]	Locality [c]	Geographic coordinates [d]	Label [e]	GenBank accessions		
					clpP	*trnfM-E*	*PHYA*
Conopholis alpina Liebm.	AC.MX.11M.1/ 6257861; MO	Coxcatlan, Puebla, Mexico	18°22′N 97°00′W	Z	HQ895610	HQ895687	HQ895645
	AC.MX.13M.2.2/ 6257860; MO	La Carbonera, Oaxaca, Mexico	17°35′N 97°00′W	X	HQ895605	HQ895682	HQ895644
	AC.05.Panama; N/A	Chiriqui, Panama	08°33′N 82°24′W	AH	HQ895602	HQ895677	HQ895642
	INBIO-8382; INBIO	Limón, Talamanca, Costa Rica	09°06′N 82°58′W	AF	N/A	HQ895678	N/A
	INBIO-8386a; INBIO	Puntarenas, Coto Brus, Costa Rica	08°57′N 82°49′W	AG	N/A	HQ895679	N/A
	MO.04805064; MO	Huajuapan, Oaxaca, Mexico	17°48′N 97°46′W	Y	HQ895608	HQ895685	N/A
	MO.04823722; MO	Reserva de la Biosfera, Chiapas, Mexico	16°46′N 93°06′W	AC	HQ895606	HQ895683	N/A
	MO.04875049; MO	Cerro Quetzal, Chiapas, Mexico	16°47′N 93°04′W	AD	HQ895607	HQ895684	N/A
	MO.04853204; MO	Reserva de la Biosfera, Chiapas, Mexico	16°45′N 93°09′W	AE	HQ895609	HQ895686	N/A
	NYBG-857; NY	San Jose, Costa Rica	09°56′N 84°03′W	AJ	HQ895604	HQ895681	N/A
	SI-0023A; US	Chiriqui, Panama	08°36′N 82°22′W	AI	HQ895603	HQ895680	HQ895643
Conopholis alpina Liebm. var. *mexicana* (A.Gray, ex S. Watson) R. R. Haynes	AC.AZ-FHR.4/ 6257876; MO	Gila Co., Arizona, USA	34°22′N 111°6′W	AO	HQ895600	HQ895675	HQ895640
	AC.NM-WWC.5/ 6257868; MO	Otero Co., New Mexico, USA	32°53′N 105°57′W	AM	HQ895598	HQ895673	HQ895638
	AC.NM-BC.1.1; N/A	Colfax Co., New Mexico, USA	36°33′N 105°03′W	AK	HQ895596	HQ895671	HQ895636
	AC.NM.DSNA.1; N/A	Town of Las Cruces, New Mexico, USA	32°18′N 106°46′W	AN	HQ895599	HQ895674	HQ895639
	ARIZ-6333b; ARIZ	Van Horn Rural, Texas, USA	31°54′N 104°50′W	AL	HQ895597	HQ895672	HQ895637
	AC-MX-24Fb.3/ 6257866; MO	Aculco, Districto Federal, Mexico	19°20′N 99°30′W	AP	HQ895601	HQ895676	HQ895641
Conopholis americana (L.) Wallr.	AC.IL-KSP.1; N/A	Vermillion Co., Illinois, USA	40°07′N 87°44′W	A	HQ895621	HQ895698	HQ895656
	SS0331; TRTE	Hickory Ridge Lookout, Monroe Co., Indiana, USA	39°02′N 86°19′W	B	HQ895622	HQ895699	HQ895657
	SS0680; TRTE	Cloudland Canyon, Dade Co., Georgia, USA	34°50′N 85°27′W	C	HQ895623	HQ895700	HQ895658
	SS0471; TRTE	Kanawha Co., West Virginia, USA	39°11′N 81°27′W	D	HQ895624	HQ895701	HQ895659
	ET.CT.11.31; N/A	Hubbard Park, New Haven Co., Connecticut, USA	41°33′N 72°50′W	E	HQ895625	HQ895702	HQ895660
	ET.NC.21.21; N/A	Hot Springs, Madison Co., North Carolina, USA	35°53′N 82°49′W	F	HQ895626	HQ895703	HQ895661
	SS0582; TRTE	Holland, Ottawa Co., Michigan, USA	42°47′N 80°06′W	G	HQ895627	HQ895604	HQ895662
	SS0594; TRTE	Halton Co., Ontario, Canada	43°25′N 79°52′W	H	HQ895628	HQ895605	HQ895663
	SS05194; TRTE	Township of Archipelago, Parry Sound, Ontario, Canada	45°20′N 80°02′W	I	HQ895629	HQ895606	HQ895664
	SS06127; TRTE	Gatlinburg, Blount Co., Tennessee, USA	35°42′N 83°30′W	J	HQ895630	HQ895707	HQ895665
	SS0758; TRTE	Shenandoah, Madison Co., Virginia, USA	38°41′N 78°19′W	K	HQ895631	HQ895708	HQ895666

APPENDIX 1. Continued

Species	DNA accession [a]/voucher [b]	Locality [c]	Geographic coordinates [d]	Label [e]	GenBank accessions		
					clpP	*trnfM-E*	*PHYA*
	SS0937; TRTE	Cheboygan Co., Michigan, USA	45°33′N 84°40′W	L	HQ895633	HQ895710	HQ895668
	SS0941; TRTE	Port Severn, Ontario, Canada	44°55′N 79°44′W	M	HQ895635	HQ895712	HQ895670
	SS0653; TRTE	Lake Warren, Hampton Co., South Carolina, USA	32°49′N 81°10′W	N	HQ895614	HQ895691	HQ895649
	SS0648; TRTE	Wakulla Spring, Wakulla Co., Florida, USA	30°07′N 84°21′W	O	HQ895615	HQ895692	HQ895650
	SS0581; TRTE	Huntland, Franklin Co., Tennessee, USA	35°03′N 86°16′W	P	HQ895613	HQ895690	HQ895648
	AC.FL-SF.4.2; N/A	Alachua Co., Florida, USA	29°44′N 82°26′W	Q	HQ895616	HQ895693	HQ895651
	AC.PA-MSF.1; N/A	Franklin Co., Pennsylvania, USA	39°55′N 77°26′W	R	HQ895617	HQ895694	HQ895652
	SS06146; TRTE	Swain Co., North Carolina, USA	35°25′N 83°27′W	S	HQ895618	HQ895695	HQ895653
	SS06174; TRTE	Granville, Licking Co., Ohio, USA	40°04′N 82°31′W	T	HQ895619	HQ895696	HQ895654
	SS0780; TRTE	Montarville Quebec, Canada	45°32′N 73°21′W	U	HQ895620	HQ895697	HQ895655
	AC.WI-DL.1; N/A	Devil's Lake State Park, Sauk Co., Wisconsin, USA	43°24′N 89°42′W	V	HQ895632	HQ895709	HQ895667
	SS0938B; TRTE	Industry Town, Franklin Co., Maine, USA	44°45′N 70°04′W	W	HQ895634	HQ895711	HQ895669
	SS0579A; TRTE	Huntsville, Madison Co., Alabama, USA	34°43′N 86°35′W	AA	HQ895611	HQ896788	HQ895646
	SS0311; TRTE	Gulf Bottom Trail, McCreary County, Kentucky, USA	36°41′N 84°28′W	AB	HQ895612	HQ895689	HQ895647
Epifagus virginiana (L.) W.P.C. Barton	MO.03399414; MO	Sierra de Guatemala, Tamaulipas, Mexico	23°5′N 99°15′W		HQ895590	N/A	N/A
	SS05200; TRTE	Mississauga, Peel, Ontario, Canada	43°32′N 79°39′W		HQ895591	N/A	N/A
	SS05202; TRTE	Moon River, Ontario, Canada	45°05′N 79°56′W		HQ895593	N/A	N/A
	SS04159; TRTE	Martin Co, Indiana, USA	38°42′N 86°44′W		HQ895589	N/A	N/A
	SS04145; TRTE	Bloomington, Indiana, USA	39°12′N 86°30′W		HQ895595	N/A	N/A
	SS05201; TRTE	Herkimer Co. New Yourk, USA	43°31′N 74°47′W		HQ895592	N/A	N/A
	SS04169; TRTE	Huntington Co, Indiana, USA	40°50′N 85°26′W		HQ895594	N/A	N/A

Note: in column DNA accession/voucher: SS, Sasa Stefanović; AC, Alison Colwell; ET, Erica Tsai; N/A, voucher not available. In column GenBank Accessions: N/A, sequences not available.

[a] Extraction labels for the specimen indicated on the phylogenetic trees (Fig. 3; Appendices S1, S2).

[b] Abbreviations of herbaria where vouchers are deposited follow Index Herbariorum.

[c] Geographic areas where the specimen were collected. When known, the lower administrative units within a country are listed (e.g., states or provinces, counties or townships).

[d] Approximate geographic coordinates for the localities from which the specimens were obtained.

[e] Letter(s) corresponding to node labels on the networks (Fig. 2).