Syst. Biol. 57(3):466–482, 2008

ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150802172168

The Reticulate History of *Medicago* (Fabaceae)

IVÁN J. MAUREIRA-BUTLER,[1,4] BERNARD E. PFEIL,[1,5] AMORNTIP MUANGPROM,[2] THOMAS C. OSBORN,[3] AND JEFF J. DOYLE[1]

[1]*Department of Plant Biology, Cornell University, Ithaca, NY 14853, USA; E-mail: jjd5@cornell.edu (J.J.D.)*
[2]*National Center for Genetic Engineering and Biotechnology, Klong Luang, Pathumthani 12120, Thailand*
[3]*Seminis Vegetable seeds (A Division of Monsanto), State Highway 16, Woodland, CA 95695, USA*
[4]*Agro aquaculture Nutritional Genomic Center (CGNA), Plant Biotechnology Unit INIA-Carillanca P.O. Box 58-D, Temuco, Chile*
[5]*CSIRO Plant Industry, GPO Box 1600, Canberra, ACT 2601, Australia*
I.J.M.-B. and B.E.P. contributed equally to this work

Abstract.—The phylogenetic history of *Medicago* was examined for 60 accessions from 56 species using two nuclear genes (CNGC5 and *β-cop*) and one mitochondrial region (*rp*S14-*cob*). The results of several analyses revealed that extensive robustly supported incongruence exists among the nuclear genes, the cause of which we seek to explain. After rejecting several processes, hybridization and lineage sorting of ancestral polymorphisms remained as the most likely factors promoting incongruence. Using coalescence simulations, we rejected lineage sorting alone as an explanation of the differences among gene trees. The results indicate that hybridization has been common and ongoing among lineages since the origin of *Medicago*. Coalescence provides a good framework to test the causes of incongruence commonly seen among gene trees but requires knowledge of effective population sizes and generation times. We estimated the effective population size at 240,000 individuals and assumed a generation time of 1 year in *Medicago* (many are annual plants). A sensitivity analysis showed that our conclusions remain unchanged using a larger effective population size and/or longer generation time. [Bayesian analysis; coalescence; Fabaceae; hybridization; incongruence; lineage sorting; low copy nuclear genes; Medicago; nDNA.]

One of the most exciting results of the increase in DNA sequence availability for plant systematics research is the ability to dissect the history of fragments of the genome separately from one another. Phylogenetic analysis of sequence data can provide high resolution by virtue of the large number of characters potentially available in any one region of the genome. Although phylogenetic analyses using large concatenated data sets have robustly resolved relationships in several taxonomic groups (Baldauf et al., 2000; Bapteste et al., 2002; Rokas et al., 2003b, 2005; Driskel et al., 2004), the history of a single region (i.e., a gene tree) can be uncoupled from that of the whole organism (Nei, 1987). The majority of the genome may be tracking one history, whereas various processes can cause a single region to track (actually or apparently) another history (Doyle, 1992; Maddison, 1997; Wendel and Doyle, 1998; and references within each). A natural question that arises when the history of parts of the genome are uncoupled from other parts is: what does a "species" tree represent (Maddison, 1997)? If only a small fragment of the genome contradicts the remainder, the answer to this question is probably the straightforward one—a species tree represents the genealogical history of the species. However, if significant fractions of the genome track different histories, a single species tree, even one constructed from numerous genes, may be an unrealistic representation of the history of the species (Maddison, 1997), especially if the underlying cause is hybridization. One of the best documented examples of genome uncoupling is observed in *Helianthus* L., where molecular evidence has indicated that three wild sunflower species, *H. anomalus* S.F. Blake, *H. deserticola* Heiser, and *H. paradoxus* Heiser, are the products of independent hybridization events and later genome restructuring between *H. annuus* L. and *H. petiolaris* Nutt (Rieseberg, 1991; Rieseberg et al., 1996, 2003; Ungerer et al., 1998). Therefore, two significant phylogenetic signals coexist in these species, making phylogeny reconstruction dependent upon the DNA fragment used to study these lineages.

Not all incongruent patterns found in sequence data necessarily indicate different histories of parts of the genome. Wendel and Doyle (1998) list three categories of processes that may lead to incongruent patterns, including technical causes, organism-level processes, and gene- or genome-level processes. The alternative possibilities need to be excluded before any one cause of incongruence can be reasonably inferred.

Medicago L. is a genus comprising 46 to 86 taxa (Lesins and Lesins, 1979; Small and Jomphe, 1989; Small, 1990a, 1990b; Small and Brookes, 1991) and includes the crop species *M. sativa*, alfalfa, and the biological model species *M. truncatula*, barrel medik (author names not in text are in Table 1). *Medicago* belongs to the tribe Trifolieae (Fabaceae), subtribe Trigonellinae, which includes *Medicago*, *Trigonella*, and *Melilotus* Mill. Lesins and Lesins (1979) suggested that the area of origin of *Medicago* was the northern coast of the Mediterranean, although previous studies placed it in the Caucasus (Ivanov, 1977). Most species are currently found in countries bordering or close to the Mediterranean Sea, the Arabian peninsula, Iraq, and the eastern Balkans (many are endemic to restricted subsets of these areas); only some members of the *M. sativa* complex, the three species in the *M. platycarpa* clade, and *M. edgeworthii* extend well beyond these areas to central, northern, and eastern Asia (summarized in Small and Jomphe, 1989, and Lesins and Lesins, 1979).

Using morphological traits from fruit, flowers, and seedlings, Small and Jomphe (1989) developed the most recent *Medicago* classification. The authors proposed 12 sections and 8 subsections. Relationships among *Medicago* species have also been studied using molecular and cytological characters (Baum, 1968; Lesins and Lesins,

TABLE 1. *Medicago, Trigonella*, and *Trifolium* accessions included in the study. Species are listed following sectional and subsectional classification proposed by Small and Jomphe (1989).

Taxa	Accession no. or variety name	Country of origin	Seed source[b]	Breeding[c] behavior
Medicago				
Section Dendrotelis				
M. arborea L.	PI 199254	Greece	Bingham	Cross
Section Medicago				
M. sativa L. subsp. *coerulea* (Less. ex Ledeb.) Schm	PI 15798	n.i.[a]	Bingham	Cross
M. sativa L. subsp. *sativa*	PI 536535	Peru	USDA	Cross
M. sativa L. subsp. *glomerata* (Balbis) Rouy	PI 577567	Italy	USDA	Cross
M. sativa L. subsp. *falcata* (L.) Arcangeli	PI 560333	USA	Bingham	Cross
M. sativa L. subsp. *Xvaria*	PI 577530	Russian Federation	USDA	Cross
M. prostrata Jacq.	PI 577446	Italy	USDA	Cross
M. rhodopea Velen.	W6 19154	Bulgaria	USDA	Cross
M. pironae Vis.	PI 577372	Italy	USDA	Cross
M. suffruticosa Ramond ex DC.	AUST 32534	Morocco	AMGRC	Cross
M. marina L.	AUST 30791	France	AMGRC	Cross
Section Carstiensae				
M. carstiensis Wulf.	MED 152/91	n.i.	IPK	Cross
Section Spirocarpos				
Subsection Pachyspireae				
M. soleirolii Duby	PI 537242	Algeria	USDA	Self
M. italica (Miller) Fiori	PI 566864	Spain	USDA	Self
M. littoralis Rohde ex Lois	PI 537222	Morocco	USDA	Self
M. truncatula Gaertn.	Jemalong	Australia	Bingham	Self
M. doliata Carming.	PI 495278	Lebanon	USDA	Self
M. turbinata (L.) All.	PI 441943	Syria	USDA	Self
M. rigidula (L.) All.	PI 495517	Greece	USDA	Self
M. constricta Durieu	PI 534177	Bulgaria	USDA	Self
M. lesinsii E. Small	PI 516720	Morocco	USDA	Self
M. murex Willd.	PI 495350	Italy	USDA	Self
Subsection Rotatae				
M. blancheana Boiss.	PI 495215	Germany	USDA	Self
M. rotata Boiss.	PI 495576	Canada	USDA	Self
M. noeana Boiss.	PI 495407	Turkey	USDA	Self
M. shepardii Post	PI 459132	Turkey	USDA	Self
Subsection Intertextae				
M. intertexta (L.) Miller	PI 498826	United Kingdom	Bingham	Self
M. ciliaris (L.) Krocker	PI 498785	Morocco	USDA	Self
M. muricolepsis Tin.	PI 495401	Italy	USDA	Self
M. granadensis Willd.	PI 498812	Turkey	USDA	Self
Subsection Leptospirae				
M. sauvagei Negre	PI 499152	Morocco	USDA	Self
M. laciniata (L.) Miller	PI 498916	Morocco	USDA	Self
M. minima (L.) Bart	PI 499072	Italy	USDA	Self
M. praecox DC.	PI 495429	Greece	USDA	Self
M. coronata (L.) Bart	PI 498805	Lebanon	USDA	Self
M. polymorpha L.	PI 566880	Belgium	USDA	Self
M. laxispira Heyn	AUST 32302	Morocco	AMGRC	Self
M. arabica (L.) Huds.	PI 495212	Hungary	USDA	Self
M. tenoreana Ser.	PI 499161	Italy	USDA	Self
M. disciformis DC.	PI 487317	Bulgaria	USDA	Self
M. lanigera Winkl. & Fedtsch	PI 498930	Former Soviet Union	USDA	Self
Section Lupularia				
M. lupulina L.	Line 1	USA	Bingham	Self
M. secundiflora Durieu	PI 537239	Morocco	USDA	Self
Section Heynianae				
M. heyniana Greuter	PI 537136	Greece	USDA	Self
Section Orbicularis				
M. orbicularis (L.) Bart.	PI 566871	Italy	USDA	Self
Section Hymenocarpos				
M. radiata L.	PI 459146	Turkey	USDA	Self
Section Platycarpae				
M. plicata (Boiss.) Sirjaev	AUST 14950	Turkey	AMGRC	Self
M. platycarpa (L.) Trautv.	PI 577374	Russian Federation	USDA	Cross
M. ruthenica (L.) Ledebour	PI 568100	China	USDA	Cross
M. popovii (E. Kor.) Sirjaev	PI 150564	Former Soviet Union	USDA	Cross
M. edgeworthii Sirjaev	USDA #26	n.i.	Campbell	Cross

(Continued on next page)

TABLE 1. *Medicago, Trigonella*, and *Trifolium* accessions included in the study. Species are listed following sectional and subsectional classification proposed by Small and Jomphe (1989). *(Continued)*

Taxa	Accession no. or variety name	Country of origin	Seed source[b]	Breeding[c] behavior
M. cretacea M. Bieb.	W6 18315	Russian Federation	USDA	Cross
Section Lunatae				
M. biflora (Griseb) E.Small	AUST 32466	Turkey	AMGRC	Self
M. brachycarpa M. Bieb.	PI 352705	Turkey	USDA	Self
M. huberi E. Small	AUST 14947	Turkey	AMGRC	Self
Section Buceras				
Subsection Erectae				
M. astroides (Fisch. & Mey) Trautv.	AUST 34568	Syria	AMGRC	Self
M. phrygia (Boiss, & Bal.) E.Small	AUST 16107	Iran	AMGRC	Self
M. fischeriana (Ser.) Trautv.	PI 568201	Turkey	USDA	Self
M. crassipes (Boiss.) E. Small	AUST 32848	Turkey	AMGRC	Self
Subsection Reflexae				
M. monspeliaca (L.) Trautv.	PI 419435	Greece	USDA	Self
Trigonella L.				
T. foenum-graecum L.	PI 199264	Greece	USDA	Self
T. mesopotamica L.	IG 16303	Syria	ICARDA	n.i.
T. spruneriana Boiss.	IG 16259	Syria	ICARDA	n.i.
T. anguina L.	PI 517185	Morocco	USDA	n.i.
T. arabica Delile	PI 292502	Israel	USDA	n.i.
T. balansae L.	PI 222211	Afghanistan	USDA	n.i.
T. caerulea L.	PI 186283	Australia	USDA	Self
T. calliceras Fisch.	PI 340801	Canada	USDA	n.i.
T. corniculata L.	PI 244289	Spain	USDA	n.i.
T. cretica L.	PI 415833	Switzerland	USDA	n.i.
T. glabra L.	PI 340803	United Kingdom	USDA	n.i.
T. macrorrhynca Boiss.	PI 222232	Iran	USDA	n.i.
T. spicata L.	PI 206284	Turkey	USDA	n.i.
T. stellata Forssk.	PI 284676	Israel	USDA	n.i.
T. suavissima Lindl.	PI 198170	Australia	USDA	n.i.
Trifolium L.				
T. pratenses L.	PI 304842	Chile	USDA	Cross
T. ambiguum L.	Endura	n.i.	Albrecht	Cross

[a] n.i. = no information.
[b] Bingham = E. T. Bingham, University of Wisconsin, Madison WI 53706; Campbell = T. A. Campbell, Soybean Genomics and Improvement Laboratory, USDA/Agricultural Research Service, Beltsville Agricultural Research Center, Beltsville, MD 20705; Albrecht = K. Albrecht, University of Wisconsin, Madison, WI 53706; USDA = U.S. Department of Agriculture, Western Regional Plant Introduction Station, Pullman, WA 99164; AMGRC = Autralian Medicago Genetic Resource Center, SARDI Waite Research Precinct, Hartley Grove, Urrbrae SA 5064; ICARDA = International Centre for Agriculture Research in Dry Area, Syria; IPK = Institute of Plant Genetics and Crop Research, Germany.
[c] Breeding behavior was assessed by comparing previous reports (Quiros and Bauchan, 1988; Lesins and Gillies, 1972; Lesins and Lesins, 1979; Small and Jomphe, 1989) and visual inspection of fruit setting on undisturbed flowers of plants grown under greenhouse conditions (Table 1).

1979; Small, 1981; Small et al., 1981, 1999; Small and Jomphe, 1989; Brummer et al., 1995; Mariani et al., 1996; Valizadech et al., 1996; Bena et al., 1998a, 1998b, 1998c; Downie et al., 1998; Bena, 2001). Despite the low number of shared taxa, two main points could be extracted from these studies: (i) phylogenetic relationships among taxa have not been fully resolved and (ii) clear incongruence exists between molecular phylogenetic inferences and the earlier generic subdivision based on morphology. A recurrent explanation for these observations in other taxa is the low phylogenetic power associated with single nuclear genes in the recovery of true species relationships (Bapteste et al., 2002; Rokas et al., 2003a, 2003b). However, some studies including data sets covering entire genomes have also failed to recover fully congruent phylogenetic reconstructions (Holland et al., 2004, 2006), pointing out the necessity of alternative explanations. Morphological characters are the reflection of genes scattered across the genome—incongruence between single genes and morphological classification could be explained, at least in part, by the existence of several phylogenetic signals within the taxa under study. If multiple signals exist, phylogenetic analysis of several genes may uncover this phenomenon.

In this study, we examine the phylogenetic history of diploid and some autopolyploid species of the legume genus *Medicago* using sequences from one mitochondrial and two nuclear genes. After observing widespread incongruence, we attempt to determine the likely causes of this pattern.

MATERIALS AND METHODS

Taxon Sampling

A total of 77 plant accessions were acquired from various sources: 60 belonging to 56 species of *Medicago*, 15 belonging to *Trigonella*, and 2 representatives of *Trifolium* that were included as outgroups (Table 1). Only one plant per accession was used as a representative of the species. All accessions used were diploid (either $2n = 16$ or $2n = 14$) except for the following: *M. arborea*, *M. sativa* ssp. *sativa*, *M. sativa* ssp. *Xvaria*, *M. sativa* ssp.

falcata ($4x = 2n = 32$). Plant ploidy was obtained from extensive *Medicago* karyotype data previously published (Clement and Stanford, 1963; Gillies, 1968, 1971, 1972a, 1972b, 1972c; Ho and Kasha, 1972; Lesins and Gillies, 1972). Some species deliberately not included in this study include polyploids of putative hybrid origin (see McCoy and Bingham, 1988; Lesins and Lesins, 1979). We chose to focus on diploid species in the first instance to reduce complexity. The polyploids included from the *M. sativa* complex show tetrasomic inheritance (Quiros, 1982; Stanford, 1951) and are therefore genetic autopolyploids. Breeding behavior was assessed by comparing previous reports (Lesins and Gillies, 1972; Lesins and Lesins, 1979; Quiros and Bauchan, 1988; Small and Jomphe, 1989) and visual inspection of fruit set on undisturbed flowers of plants grown under greenhouse conditions (Table 1). No voucher specimens were created; however, public accession numbers are provided in Table 1.

Gene Primer Development and DNA Amplifications

One or two young leaflets were collected from individual plants, and total DNA was isolated as previously described (Michaels and Amasino, 2001). One mitochondrial and two nuclear genes were amplified using PCR. Primers for the mitochondrial *rp*S14-*cob* region have been previously described (Demesure et al., 1995).

Ten conserved orthologue set (COS) markers that were highly conserved among tomato, *Arabidopsis thaliana*, and *Medicago truncatula* were proposed by Fulton et al. (2002) as possible sources of data in comparative genome and phylogenetic studies. Sequences of the 10 *M. truncatula* COS markers were obtained from the TIGR database (TIGR; http://www.tigr.org/docs/tigr-scripts/tgi/tc_report.pl, as accessed in September 2002) and used to search an *A. thaliana* database (TAIR; http://www.Arabidopsis.org/cgi-bin/Blast/TAIRblast.pl, as accessed in September 2002) using BLASTn. The resulting *A. thaliana* gene sequences showing highest similarity to the *M. truncatula* ESTs and the sequences of the *M. truncatula* ESTs were used to design degenerate primers predicted to amplify orthologous *Medicago* sequences.

Preliminary results showed that primers designed based on two *M. truncatula* EST contigs, TC5734 and TC8858 (COS1850 and COS1039, respectively), were able to amplify a wide range of *Medicago* and *Trigonella* samples, and direct sequencing was possible from the PCR products. In addition, a single-strand conformation polymorphism (SSCP) analysis was carried out (as described by Muangprom et al., 2005) to confirm the presence of only one gene copy and/or allele for these two COS markers. Results from BLASTn showed that *M. truncatula* EST contig TC5734 had highest similarity to At5g57940, a cyclin nucleotide-gated channel (CNGC5), with score 139 (E-value of $9e^{-32}$), and TC8858 had highest similarity to At4g31480, a putative coatomer beta subunit (β-*cop* protein), similar to β-*cop* from *Rattus norvegicus*, *Mus musculus*, and *Homo sapiens*, with score 238 (E-value of $2e^{-61}$). We refer to these two genes as CNGC5 and β-*cop*-like.

The primers used for CNGC5 and β-*cop*-like were forward 5′-TCATCTCTGTYTGGCTTTAGTG-3′ and reverse 5′-AAGCAGCCCARGTYCTCCAT-3′ for CNGC5, and forward 5′-CCACAYCCWATTGATAATGATTC-3′ and reverse 5′-GTGAGYTGAAGAATGCGGTTA-3′ for β-*cop*-like, respectively. PCR reactions were conducted as previously reported (Seah et al., 1998), with the following modifications: the reactions were performed using 20 μL reaction with 2 μL each of 10 $\times$ buffer, 2 mM dNTPs, 10 μM of the forward and reverse primers, and 1.6 μL of 25 mM $MgCl_2$, 0.4 μL (2.0 units) of *Taq* DNA polymerase (Promega, Madison, WI), 0.5 μL BSA, 6.5 μL H_2O, and 3 μL of DNA. Thermal cycling consisted of 94°C for 5 min and 38 cycles of 30 s at 95°C, 30 s at 60°C for *rp*S14-*cob* and β-*cop*-like or 56°C for CNGC5, 1 min at 72°C, and a final step of 72°C for 7 min. Amplification success was determined by separating products on 1.5% agarose gel and visualizing with ethidium bromide. PCR products were excised from the gel and purified using the GFX PCR DNA and gel band isolation kit (Amersham Biosciences, Piscataway, NJ). Direct sequences were produced as described previously (Lukens et al., 2003). All PCR products from CNGC5 and β-*cop*-like were checked using SSCP. Chromosomal locations were inferred using the best matches of a Cvit BLASTn search against the *M. truncatula* pseudomolecule (http://www.medicago.org/genome/cvit_blast.php).

Each PCR product was sequenced twice using both forward and reverse primers in separate sequencing reactions. Forward and reverse sequences were aligned using the BLAST Two Sequences (bl2seq) tool from the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov). Reading error differences between forward and reverse sequences were resolved by visual inspection of chromatograms. Sequences from each set of primers were initially aligned using ClustalW (seqtool.sdsc.edu/CGI/BW.cgi; using default parameters). However, all alignments were confirmed by visual inspection, with manual modifications where necessary. Alignments are available on request. The DNA sequences were deposited in GenBank under accessions numbers DQ662600 to DQ662827.

Phylogenetic Analysis

Assessment of combinability.—We checked for incongruence length differences using the maximum parsimony (MP) criterion to assess whether the three genes were carrying differing signals. The partition homogeneity test implemented in PAUP* (Swofford, 1998) was performed with pairwise and a three-way partition comparison with 100 replicates using only informative characters (Lee, 1998). Searching was done using two random addition sequence (RAS) replicates (with a maximum of 100 trees per RAS replicate) per partition homogeneity replicate. Because significant incongruence was detected, we employed further methods to examine the nature of the incongruence. We checked MP bootstrap scores (using 1000 replicates, with two RAS per replicate, saving a maximum of 10 trees per replicate) in separate

analyses of each partition to see if the incongruence suggested by the partition homogeneity test was robust. We used reverse successive weighting (Trueman, 1998) for each partition separately with 500 bootstrap replicates, searches limited to 10,000 trees but otherwise default parameters to assess whether contradictory secondary signals exist within each of the separate partitions and, if so, whether the characters contributing to the primary or secondary signals are scattered or localized in the sequences.

We arbitrarily selected the GTR + G model for each region in separate analyses using MrBayes version 3.1.1 (Ronquist and Huelsenbeck, 2003) and compared these results to those obtained by equal-weight parsimony. We wanted to test whether the incongruence is only a function of the uniform model (equal-weight parsimony) being applied to all partitions in the homogeneity test. Finding the same topologies across genes by separate analyses would indicate that model misspecification, rather than different histories, is probably the cause of incongruence, rather than different histories. Using flat priors and 10 chains, we ran the Bayesian analysis (BA) for five million generations (sampled every 1000, but excluding a burn-in of one million generations based on the likelihood score over generation plot). We visually examined the likelihood score, total tree length, alpha parameter (using Excel, Microsoft), and topology (via posterior probabilities of clades, using TreeView) and found that each had converged between runs. Where any clade posterior probability was above 0.95, the variation among runs for each gene was no more than 0.02. The standard deviation of split frequencies between chains was also below 0.01 for each gene, indicating adequate mixing within a run. Because the separate BA confirmed the incongruence found in the partition homogeneity test, we proceeded with further refinements to the model choice and analyses for each partition separately.

Refinement of models for each gene.—CNGC5 and *β-cop*-like contain exons and introns, so we focused mostly on mixed models that analyzed the exons and introns as separate partitions, with more or fewer parameters unlinked across data partitions. *rp*S14-cob is predominantly a mitochondrial intergenic spacer and appeared to have evolved more slowly than CNGC5 and *β-cop*-like. Therefore we used only a homogeneous model for this data set, but incorporated more alternative homogeneous models than in CNGC5 or *β-cop*-like, including an invariant sites parameter.

We ran three separate analyses for each model listed below (see Table 2). The first analysis for each model was run to five million generations, the other two analyses to two million generations. Convergence within and between analyses was checked as above, with the addition of the kappa or one of the GTR substitution parameters to those we examined. Although convergence and stability of the likelihood score alone is not enough to guarantee an overall convergent and stable solution, because this value can stabilize whereas other parameters do not (Nylander et al., 2004), convergence among multiple parameters most likely does.

We examined the Bayes factors (BFs, defined as $2\ln B_{10}$, where B_{10} is the ratio of likelihoods of the compared models) for each analysis, and also checked how many parameters produced how much difference in $-\ln L$ among models close to the best model selected by BFs. Interpretation of BFs was done following criteria of Kass and Raftery (1995), where $2\ln B_{10}$ values larger than 10 are considered strong evidence against the simpler model (model 0).

Splits tree display of Bayesian analysis.—We used the trees produced by the best Bayesian analysis of CNGC5 and *β-cop*-like as input for a consensus network display of these results using Splits Tree 4 (Huson and Bryant, 2005). Due to apparent software limitations, we used 200 arbitrarily selected trees for each of these genes (after the burn-in period). We set the display threshold at 0.475, which will generally allow only those clades (and reticulations) supported by 95% or more trees from either gene to be included (where the clades are robustly resolved in each gene). The reticulations in the consensus network can therefore be considered to have 95% or better posterior probability (PP) in these circumstances.

Estimation of Dates

Using the published estimation of the dates of legume divergences based on chloroplast *matK* sequences (Lavin et al., 2005), we inferred the time of the divergence between *Medicago* and *Trigonella* as ca. 15.9 Ma (million years ago) from the published chronogram. The uncertainty surrounding this node was not reported in Lavin et al. (2005), but based on the average of the closest four nodes the 100% credibility interval may be around 12 Myr (million years; 9.9 to 21.9 Ma, or around a 2.2-fold range). This age provided a fixed calibration point at the root of the *Medicago* tree to estimate the dates of internal nodes with a penalized likelihood procedure (Sanderson, 2003). Cross-validation to find the optimal smoothing parameter (10^k) was done using increments of k of 0.1, from $k = -3$ to 3 (using a random tree from the stable posterior distribution of each gene). Chronograms were produced and used in the coalescence simulation (to provide an estimate of the time depth of each branch in the tree).

Coalescence Simulations

Coalescence simulations were carried out to elucidate which was the most likely cause of the incongruence observed among the nuclear gene trees. We used the "Coalescence Contained within Current Tree" module of Mesquite version 1.06 (http://www.mesquiteproject.org) to simulate 200 gene trees, using as species tree the topologies (as chronograms) of each of the two nuclear genes. For these analyses we assumed a generation time of one year, panmixis, and a constant effective population size (N_e) of 240,000. We selected an N_e of 240,000 because (i) empirical N_e estimations for gene CNGC5 using diploid taxa from the *M. sativa* complex have yielded values around 240,000 (sequences from *M. sativa* sp. *caerulea* [2×] and *M. sativa* sp. *falcata* [2×] were used to estimate the mutation rate

TABLE 2. Bayesian analysis results by gene and model.

Region and model	No. additional parameters	Unlinked parameters	Unstable parameters	−ln *L* harmonic mean
CNGC5				
Homogeneous models				
1. HKY	—		All stable	−5589
2. HKY+G	1		TL unstable	−5868
3. GTR	5		All stable	−5585
4. GTR+G	6		ln *L* not converged at 2M; TL stable only after 2M, α unstable after 2M	−5617
Heterogeneous models				
5. HKY+G	2	α	TL and α unstable	−5565
6. HKY+G	2	κ	TL unstable	−5701
7. HKY+G	3	α, κ	Only 2 of 3 ln *L* converged at 2M; TL and α unstable	−5560
8. HKY+G	6	α, κ, base frequency	Only 2 of 3 ln *L* converged at 2M; TL and α unstable	−5546
9. GTR+G	7	α	Only 2 of 3 ln *L* converged at 2M; TL and α unstable	−5560
10. GTR+G	12	revmat	TL unstable	−5700
11. GTR+G	13	α, revmat	TL and α unstable	−5558
12. GTR+G	16	α, revmat, base frequency	Only 2 of 3 ln *L* converged at 2M; TL and α unstable	−5541
β-cop-like				
Homogeneous models				
1. HKY	—		All stable	−5490
2. HKY+G	1		All stable	−5407
3. GTR	5		All stable	−5475
4. GTR+G	6		All stable	−5395
Heterogeneous models				
5. HKY+G	2	α	α Unstable	−5337
6. HKY+G	2	κ	TL and α unstable	−5373 (3.5M only)
7. HKY+G	3	α and κ	α Unstable (not all runs stable and convergent in TL)	−5326
8. HKY+G	6	α, κ, base frequency	α Unstable	−5316
9. GTR+G	7	α	α Unstable	−5330
10. GTR+G	12	revmat	All stable	−5349
11. GTR+G	13	α and revmat	TL and α unstable	−5369
12. GTR+G	16	α, κ, base frequency	TL and α unstable	−5338
rps4-cob				
1. HKY	—		All stable	−2771
2. HKY+I	1		All stable	−2767
3. HKY+G	1		TL unstable	−2977
4. HKY+I+G	2		TL unstable, not all converged at 2M	−2820
5. GTR	5		All stable	−2763
6. GTR+I	6		All stable	−2775
7. GTR+G	6		TL unstable	−2977
8. GTR+I+G	7		TL unstable	−2859

α = alpha, the shape parameter of the gamma distribution of rate variation; κ = kappa, ratio between transition and transversion rates; revmat = rate substitution matrix for the GTR model; TL = tree length; I = proportion of invariable sites.

and N_e using *M. truncatula* as the outgroup following Fischer et al., 2004); (ii) the *M. sativa* complex contains the most wide-ranging taxa that also readily outcross, both factors that indicate it may have the largest N_e in the genus; (iii) possible introgression from other *Medicago* species as a result of alfalfa breeding may have caused an increase in genetic diversity leading to an inflated estimate of N_e; and (iv) using a value of N_e several times higher than commonly assumed estimations (Maddison and Knowles, 2006) allows more stringent comparisons between allele sorting and hybridization as a cause of incongruence by favoring the null hypothesis of allele sorting (see Discussion). We then compared the tree-to-tree distance of the original nuclear gene tree ("species tree") with the simulated trees (each treated as unrooted) using the partition metric (Penny and Hendy, 1985) implemented in PAUP* (Swofford, 1998) as the symmetric distance and checked whether the distance between the two nuclear gene trees was contained within the distribution of tree-to-tree distances of the simulated gene trees.

If the distance between the two nuclear gene trees was larger than 95% of the distribution of tree-to-tree distances of simulated trees from their respective gene trees, we interpreted this as making lineage sorting of ancestral polymorphisms *alone* an unlikely explanation for the incongruence observed among the real gene trees. The coalescence model assumes that the gene trees produced under a given species tree are evolving neutrally.

Because of the uncertainty in age estimates in the phylogeny (which affects the estimated number of generations), the use of only a single locus to estimate effective population size, and the assumption of one year per

generation being unlikely over the history of the genus, we explored the sensitivity of the coalescence analysis to varying parameters. We doubled the N_e and tripled the years per generation and explored the effect on our conclusions separately and in combination.

Simulations were also carried out across a range of N_e with an assumed species tree for 15 taxa (Supplemental Data; available online at http://www.systematicbiology.org). The species tree for these tests was somewhat pectinate and contained varying branch lengths reasonably similar to many published molecular phylogenies. Fifty simulated gene trees were compared in all pairwise combinations and the same test (above) implemented. From these results, we were able to ascertain that the type 1 error rate is less than 5% when lineage sorting alone is responsible for intergenic differences, suggesting our test is conservative when rejecting the null hypothesis (Supplemental Data).

RESULTS

Sequences from each gene resulted in traces without double peaks. For all individuals, only a single allele at each locus was observed at both nuclear genes on SSCP gels. The aligned lengths of the sequences were as follows: CNGC5 955 nucleotides, *β-cop*-like 936 nucleotides and *rp*S14-*cob* 1096 nucleotides. CViT BLASTn searches (http://www.medicago.org) found the best match to pseudochromosome 8 for CNGC5 and pseudochromosome 7 for *β-cop*-like in *M. truncatula*. We therefore assumed a lack of linkage between these loci in all taxa. Matrices and main trees were submitted to TreeBase (http://www.treebase.org) as study number S1917.

Tests of Incongruence

Partition homogeneity tests of pairwise combinations among the three genes and a three-way test suggested severe incongruence among partitions (all $P < 0.01$). Bootstrap support using MP showed that several taxa had well-supported alternative phylogenetic positions (BS > 70%; data not shown), and despite removal of these taxa, incongruence in the partition homogeneity test remained ($P < 0.01$). Removal of the two outgroup *Trifolium* sequences to check for long branch attraction and reanalysis with MP did not fundamentally change the topology of the ingroup (not shown).

Reverse successive weighting of each partition separately found no robust secondary signal. The characters contributing to the primary signal were fairly evenly spread within each data partition. These results suggested that chimeric sequences or convergence shared among many taxa do not explain the incongruence among data partitions. Convergence or chimeras may be present among only a few taxa, although the observation that removal of several (up to six) incongruent taxa had no appreciable affect on the incongruence (data not shown) suggests that this is unlikely.

Simple Bayesian separate analyses using GTR+G produced trees for each gene that were similar to the parsimony trees, most notably containing strong agreement in the position and support of the incongruent taxa (not shown). Therefore, it is unlikely that the incongruence among these genes was solely due to differing model/parameter requirements.

Separate Analyses: Model Results

Results for the separate analyses using BA are shown in Table 2. Of the three analyses run for each model, usually two or three converged before two million generations. In only one case (CNGC5 model 4) did all three fail to converge after two million generations, so an additional five million–generation analysis was conducted. This additional run also failed to converge, so we assumed that the best of four runs represented the optimal estimation of PP. Two million generations appear to be generally adequate to reach convergence for these data, as judged from runs with five million generations for each model.

For CNGC5, the most complex model (with the most parameters) had the best $-\ln L$ score. However, the improvement in likelihood gained by adding parameters to the model was not linear. Although model 12 for CNGC5 had the best $-\ln L$, the improvement over model 8 was minimal, given that these models differ by 10 extra parameters. It is noteworthy that there was only minor improvement from the simplest model to model 5 with 2 extra parameters and then to model 12 with 17 extra parameters. The addition of two unlinked alpha parameters to model among-site rate heterogeneity within the exons and introns in model 5 separately appears to provide the greatest likelihood improvement at the cost of the minimum number of extra parameters within this gene.

With Bayes factors as a guide to model selection (Table 3 and Supplemental Data, available online at http://www.systematicbiology.org), the most complex model was preferred for CNGC5, because the improvement over simpler models is 10 BFs or greater, an amount regarded as "strong" to "very strong" evidence against the simpler model (Nylander et al., 2004).

For *β-cop*-like, the most complex model did not have the best $-\ln L$ score. As in CNGC5, the improvement from model 1 to model 5, with two extra parameters, was the most marked given the number of additional parameters involved. BFs indicate that model 8 improves on all simpler models and is not improved upon by more complex models and is therefore preferred by this method.

For *rp*S14-*cob*, the most complex model did not have the best $-\ln L$ score. Most improvement per extra parameter was found between model 1 and model 2; however, the best model selected by BFs was model 5.

Unstable parameters were found in several of the more complex models in each gene (not shown). In all of our partitioned analyses, the $-\ln L$ was stable, and the topology and clade PPs were fairly consistent. The consistency of topology, of clade PPs, and among MP and BA analyses allow us to conclude that model and analysis differences have little to no effect on the topologies inferred with each gene.

TABLE 3. Bayes factors among models (from Table 2) for *CNGC5* and *β-cop*-like. Values above and below the diagonal correspond to model comparisons for *CNGC5* and *β-cop*-like, respectively.

Model	1	2	3	4	5	6	7	8	9	10	11	12
1		8	−558	−56	48	−224	58	86	58	−222	62	96
2	30		−566	−64	40	−232	50	78	50	−230	54	88
3	166	136		502	606	334	616	644	616	336	620	658
4	190	160	24		104	−168	114	142	114	166	118	152
5	306	276	140	116		−272	10	38	10	−270	14	48
6	234	204	68	44	−76		282	310	282	2	286	320
7	328	298	162	138	22	94		28	0	−280	4	38
8	348	318	182	158	42	114	20		−28	−308	−24	10
9	320	290	154	130	14	86	−8	−28		−280	4	38
10	282	252	116	92	−24	48	−46	−66	−38		284	318
11	242	212	76	52	−64	8	−86	−106	−78	−40		34
12	304	274	138	114	−2	70	−24	−44	−16	22	62	

Separate Analyses: Phylogenetic Results

Separate analyses using a variety of models supported our initial result that the three genes, and especially CNGC5 and *β-cop*-like, are carrying different phylogenetic signals for a large number of taxa (Fig. 1, Supplemental Data, available online at http://www.systematicbiology.org). The lower level of resolution of *rp*S14-*cob* limited the detection of incongruent clades when compared with the two nuclear regions; however, a number of incongruent clades were found (Supplemental Fig. 1). MP bootstrap support for clades (not shown) with high PP (≥0.95) was usually high (≥80%) and no cases of high MP bootstrap were found for clades contradicting those with high PP. In one case, a robustly supported disagreement between *rp*S14-*cob* and both nuclear genes was found regarding the grouping *M. granadensis* with *M. intertexta*, *M. muricolepsis*, and *M. ciliaris* (a clade of four species identified by both nuclear genes; Clade 8 in Fig. 1). The trees based on *rp*S14-*cob* instead placed *M. arabica* sister to the latter three species. We focus most of our discussion on the two nuclear genes because they display the greatest incongruence.

The depth and breadth of the incongruence between the two nuclear genes shown in Figure 1 was displayed as a consensus network, which shows the considerable supported incongruence among taxa (Fig. 2). Groups having the same circled number in Figure 1 are common to both nuclear genes and, as expected, form clades in Figure 2, referred to as common clades hereafter (although three of these groups do not form a clade in one gene, the clade found in the second gene is not contradicted by the first). However, the relationships among these common clades are entangled. Most of these common clades contain from two to four species, although the clade containing *M. truncatula* is an exception with eight species.

Of nine common clades, the majority (seven) only contain species that readily produce selfed seed (identified with an S in Fig. 1). The remaining two common clades (containing *M. marina* + *M rhodopea* and *M. platycarpa* + *M. popovii* + *M. ruthenica*) and the remaining members of the *M. sativa* complex (broadly defined) are the only groups of taxa that do not readily form selfed seed, even when hand-pollinated (identified with an × in Fig. 1).

The incongruence in relationships among these common clades and among other taxa penetrates almost to the deepest parts of the network. Whereas *Medicago* and *Trigonella* are resolved as monophyletic sister groups in each gene, indicating a common pattern of genus membership supported by each data set, the reticulation within *Medicago* confounds standard phylogenetic inference from the earliest divergence within the genus. This reticulating pattern also extends to the present day within the *M. sativa* complex, where other studies have shown current gene flow between some of the named species/subspecies in this group (Lesins and Lesins, 1979; Quiros and Bauchan, 1988; Brummer et al., 1991; Kidwell et al., 1994; Muller et al., 2006).

Removal of the two outgroup *Trifolium* sequences (with the longest branch and least sampled clade in the analyses), to test for the effect of long branch attraction, and reanalysis with MP and BA did not change the topology of the ingroup significantly (data not shown).

Coalescence Simulations

Under the assumption that alleles from unlinked loci assort independently from ancestral to descendant species, we carried out coalescent simulations to test if the incongruent pattern between our nuclear genes could be explained by random chance alone (lineage sorting hypothesis). The distribution of tree-to-tree distances was simulated under a neutral coalescence model for each nuclear gene, on the assumption that the gene tree was the true species topology. The distance between the two gene trees was then compared to these distributions, and in both cases was found to lie far outside the distribution for either gene (Fig. 3). This result indicates that under the assumptions of this analysis (panmixis, $N_e = 240{,}000$), lineage sorting alone cannot account for the degree of incongruence between the two nuclear gene trees. With $N_e = 480{,}000$ (a twofold increase), or 3 years per generation (threefold increase), or both—increases that each favor the null—lineage sorting alone was still rejected (Fig. 3).

Morphological Classification versus Gene Trees

There was little concordance between the current morphology-based subgeneric classification (Small and

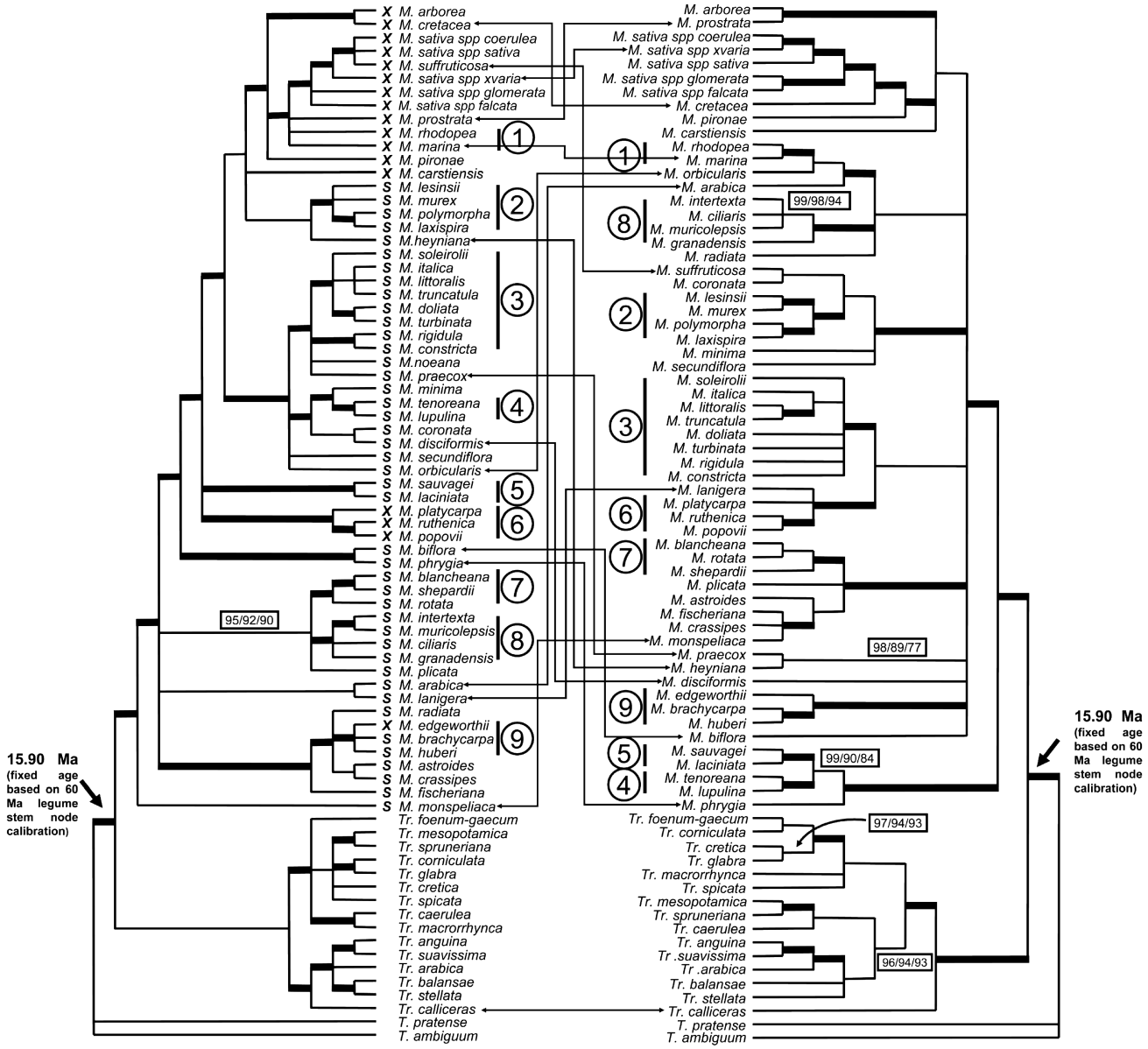

FIGURE 1. Summary of phylogenetic relationships of *Medicago* species for two nuclear genes. CNGC5 (nDNA, left), *β-cop*-like (nDNA, right). Strongly supported clades (≥ 95% PP) under three models have bold subtending branches (CNGC5: HKY; HKY + G with α unlinked; GTR + G with α, substitution matrix, and base frequencies unlinked. *β-cop*: HKY; HKY + G with α unlinked; HKY + G with α, κ, and base frequencies unlinked). Clades where alternative models disagree on support are indicated by boxed support values for those models (in the same order as above for each gene). Taxa with supported alternative placements between CNGC5 and *β-cop*-like are connected by lines across the center of the figure. Clades in agreement between CNGC5 and *β-cop*-like are indicated by circled numbers (1 to 9). **S** and **X** identify selfing and outcrosing breeding behavior, respectively.

Jomphe, 1989; Table 1) and our phylogenetic results. For example, *M. lupulina* and *M. tenoreana*, which differ greatly in their morphology from one another and were previously classified in section *Lupularia* and section *Spirocarpos* subsection *Leptospirae*, respectively (Small and Jomphe, 1989), were strongly grouped by both nuclear genes. Only one section, *Medicago* (grouped by a dotted line) and subsections *Pachyspirae* (group 3), *Intertextae* (group 8), and *Rotatae* (group 7) of section *Spirocarpos* were concordant, although not perfectly, to clusters found by our network analysis (Fig. 2).

DISCUSSION

How Reasonable Are Various Causes of Incongruence?

The incongruence we found was robustly supported under alternative methods of analysis (MP and BA) and largely consistent among a wide variety of models within BA. Technical causes listed in Wendel and Doyle (1998) as possible causes of incongruence (insufficient data, the choice of a gene that changes too slowly) can be ruled out because of these robust results. Sequencing errors are unlikely to have caused strongly supported different

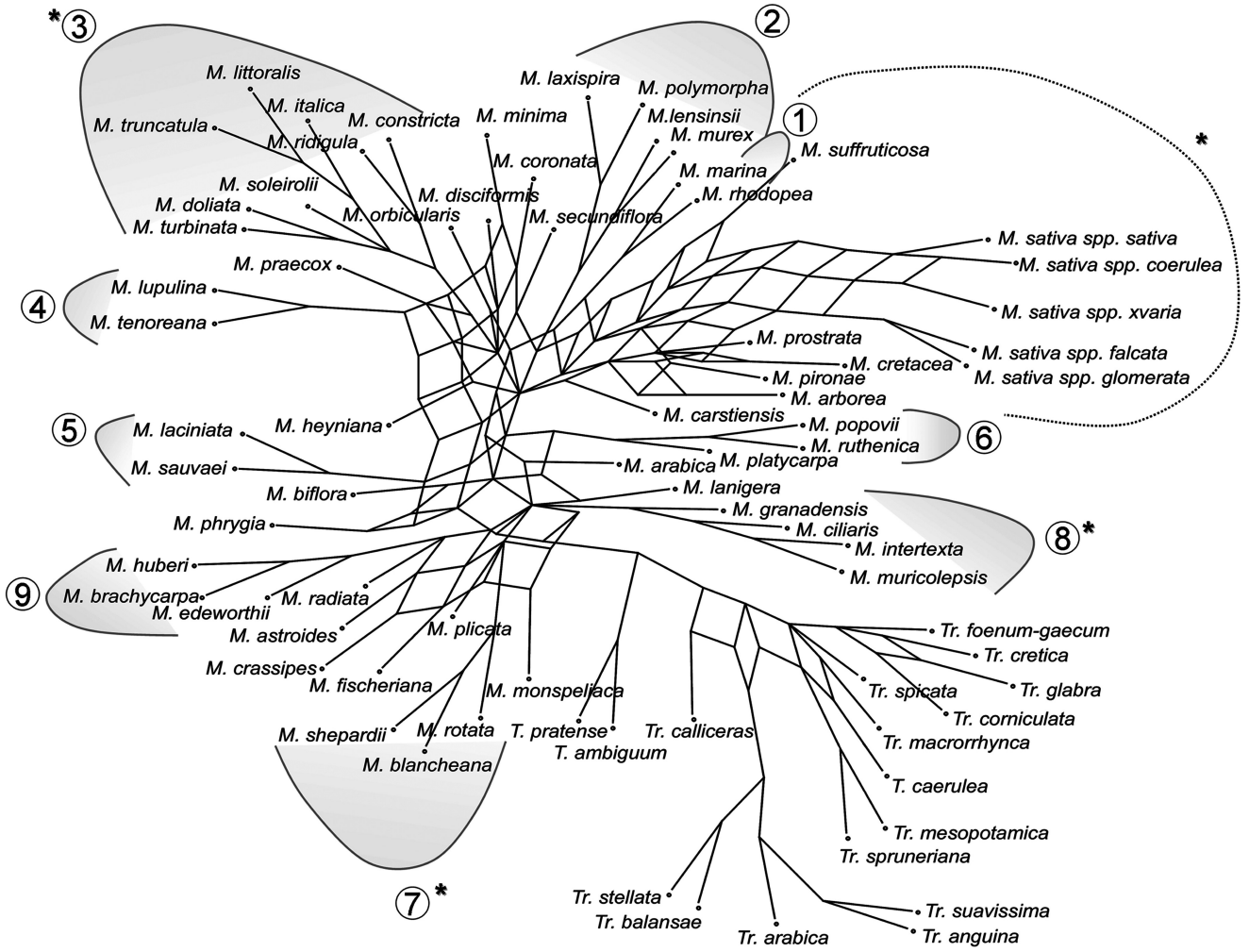

FIGURE 2. Consensus network of phylogenetic trees of *Medicago* derived from CNGC5 and *β-cop*-like Bayesian analysis. Two hundred trees were sampled from each of the stable posterior distributions of CNGC5 (HKY + G with α unlinked) and *β-cop*-like (HKY + G with α unlinked) using Splits Tree 4. A threshold of 0.475 was used to ensure that only clades with 0.95 PP or better from either set of gene trees contribute to resolving either clades or reticulations within the network. Circled numbers indicate common clades shown in Figure 1. Asterisks (*) indicate clades that show good agreement with morphological classifications (Small and Jompe, 1989; Table 1). Clade 3: subsection *Pachyspirae* (lacking *M. lesinsii* and *M. murex*); clade 7: subsection *Rotatae*; clade 8: subsection *Intertextae*; segmented line: includes species and subspecies mostly classified within section *Medicago* (plus *M. arborea* and *M. cretacea*).

placements of taxa among genes, because all sequences were sequenced in both directions and checked thoroughly (see Materials and Methods). Sequences were generated from the same DNA isolation for each gene examined, ruling out misidentified accessions as a cause of incongruence. Insufficient taxon sampling, due either to poor sampling of extant species or to extinction, is also unlikely to be a factor of incongruence. We sampled all sections and subsections of *Medicago* except *Medicago hypogaea*, the sole member of section *Geocarpa*. Extinction can produce long branches, which in general pose greater problems for parsimony than for model-based analyses (Swofford et al., 2001). Given that we observed incongruence in both parsimony and model-based analyses, we conclude that long-branch attraction, and hence undersampling due to extinction, is unlikely to be the cause of incongruence. Removal of the outgroup sequences did not change the topologies of the ingroup significantly (data not shown).

Convergent evolution is unlikely to be driving the incongruence. Functional convergence would mainly involve the coding sequences, whereas the majority of variable sites in the nuclear genes used here are in introns, thought to be neutrally evolving. Rapid diversification can also be ruled out, because we have numerous robustly resolved clades containing different taxa among genes.

Horizontal transfer of genes can also cause incongruence. Several horizontal transfers of mitochondrial genes from plant to plant between distantly related species have been reported (Bergthorsson et al., 2003). There are also evolutionarily recent cases of transfer of genes from mitochondria into the nucleus (Adams and Palmer, 2003). The nuclear genes used here are not related to

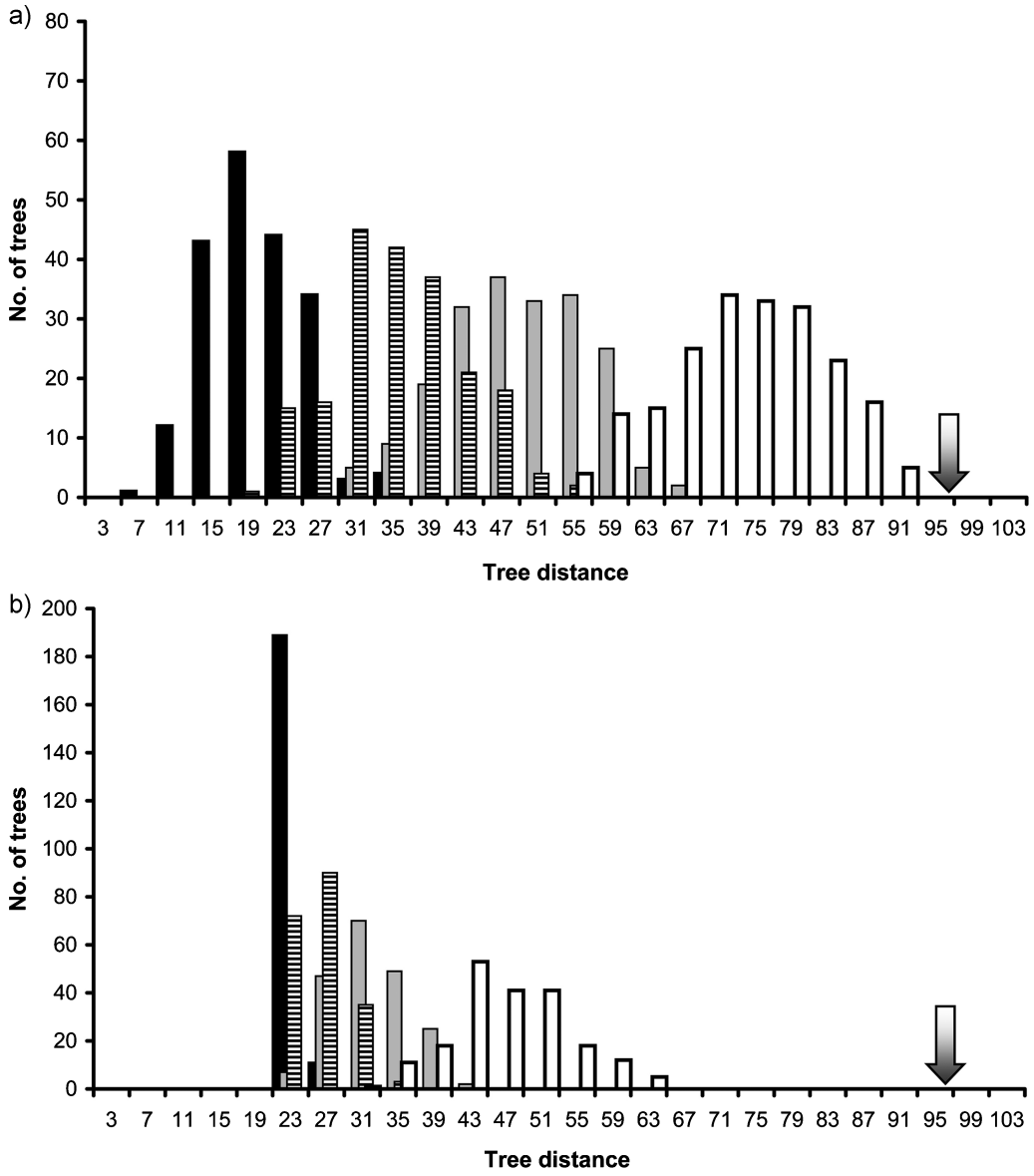

FIGURE 3. Frequency distribution of tree-to-tree distances between each nuclear gene tree and 500 simulated gene trees using coalescence simulations. (a) Distances of simulated trees from the CNGC5 and (b) from *β-cop*-like gene trees; the distance from the CNGC5 to the *β-cop*-like gene trees is marked by an arrow. Black, dashed, grey, and white bars represent distances of simulated trees under the assumption of 240,000 N_e/1 year per generation, 480,000 N_e/1 year per generation, 240,000 N_e/3 years per generation, and 480,000 N_e/3 years per generation, respectively.

proteins encoded by mitochondrial genes; therefore, they are unlikely to have entered the nucleus via mitochondria, although transfer by other mechanisms cannot be ruled out.

Some gene- and genome-level processes that could cause incongruence can also be ruled out. These include orthology/paralogy conflation, interlocus interactions, and concerted evolution. The use of the same primers across taxa for each nuclear gene coupled with SSCP determination of a single amplification product from each individual strongly argues for the presence of a single orthologous locus within each individual. Ancient duplicate copies (paralogues) being confused with orthologues is very unlikely given the congruence among gene trees at the generic level (non-monophyly of the genera would be likely if ancient paralogues were sorting out among *Medicago* and *Trigonella* species). Gene duplications within *Medicago* would have to have been accompanied by retention over several cladogenic events, followed by paralogous losses in every member of a clade containing the duplication (to return each to a single copy). That *no* duplicate copies were recovered in a sample of 77 taxa (60 from *Medicago*) makes this scenario unlikely. The lack of duplicate copies renders moot interlocus interactions, including concerted evolution.

Rate heterogeneity among sites and common unequal base frequencies are two potential causes of incongruence that are accounted for in our model-based analyses. A related phenomenon is heterotachy, where parts of a sequence can be evolving quickly or slowly but with different rates in different lineages rather than generally quickly or slowly (this might include effects such as lineage-specific base compositional bias). Heterotachy can have an effect on phylogenetic reconstruction when the branches separating heterotachous lineages are short and may affect model-based methods slightly more than parsimony (Kolaczkowski and Thornton, 2004). Our consistent results within each gene among methods indicate that our data do not fall into the small zone where model-based methods fail, whereas parsimony does not. Heterotachy—caused by shifts in selection pressure on groups of sites among taxa (Lopez et al., 2002)—is unlikely to apply to our data, which are derived mainly from noncoding DNA.

Hybridization versus the Sorting of Ancestral Polymorphisms Tested by Coalescence Simulation

The phylogenetic pattern produced by ancestral polymorphism with subsequent lineage sorting is difficult to distinguish from hybridization (e.g., Doyle et al., 1999; Sang and Zhong, 2000; Peters et al., 2007) and therefore both need to be considered as possible causes of incongruence. To some extent these are the extremes of a continuum. At the one end of the spectrum is hybridization among fully differentiated species that have subsequent fixation of some nuclear genes and possibly organellar genomes. At the other end is the capacity of a large near-panmictic population to carry multiple alleles for many loci followed by the breakdown of panmixis to allow population differentiation and subsequent sorting (random fixation) of specific alleles at many loci independently within the differentiating populations.

Nei and Kumar (2000) showed that the probability of incongruence, due to incomplete sorting of ancestral alleles, between gene topologies and species topologies will increase when (i) the time between species splitting measured in number of generations is short and (ii) when the effective population size (N_e) is high. Assessing the effect of N_e is an intricate problem given the difficulty of estimating the ancestral population sizes for each of the 56 *Medicago* species included here. Although several methods have been developed to estimate ancestral population sizes (Yang, 1997; Rannala and Yang, 2003; Wall, 2003), they require data from numerous orthologous genes and the sampling of several individuals per species. Moreover, most of these methodologies are designed to deal with few species, making their use less feasible when analyzing a large number of taxa.

As an alternative approach, we carried out coalescence simulations to assess the effect of N_e on the likelihood that the sorting of ancestral alleles is a major cause of incongruence. Our gene tree-to-tree distance histograms (Fig. 3) show that under the permissive assumption of a large N_e (240,000) and 1 year per generation, ancestral allele sorting alone was not a reasonable sole explanation for the incongruence observed between CNGC5 and *β-cop*-like across the whole genus.

A key outcome of hybridization is that the tree of population divergences is no longer being tracked by all genes for all species in the genome. This could be thought of as producing multiple species trees. The sorting of ancestral polymorphisms still operates, even if there are multiple species trees, to produce a cloud of gene trees representing the different outcomes of mutation, segregation, and sorting among alleles among each species tree. Because we cannot know the species tree(s) in advance, or how many there might be, our coalescence simulation-based test attempts to ascertain whether differences among gene trees are too great to be explained by lineage sorting alone. In effect, we are asking whether there is one cloud of gene trees from a single species tree, with variation produced by lineage sorting alone, that contains both of our sampled genes or whether more than one nonoverlapping cloud exists.

We made the assumption that each gene tree represents the species tree that produced it. Clearly this assumption is unreasonable when lineage sorting is prevalent, but the key point is that the coalescence simulation provides a framework for assessing the variation induced by lineage sorting alone around the gene tree. It approximates the size of the cloud of trees that the real species tree(s) could produce and allows assessment of whether that cloud overlaps between the two gene trees sampled here.

We tested how well this works by taking a hypothetical species tree, simulating gene trees and sampling all paired combinations of these gene trees and then running the test. When lineage sorting is low, most gene trees look like the species tree, as do most trees simulated from the gene trees. Even when lineage sorting is high and few or no gene trees look like the species tree, the variation in gene trees estimated by the simulations using gene trees as species trees still allows the overlap of gene tree-to-tree distances to show that only one species tree is required to explain the gene tree similarities. The gene trees are different, but it is the variation around the trees estimated by coalescence simulation that is important in the success of the test.

We assessed the type 1 error rate and found it to be less than 5% under a range of levels of lineage sorting with a critical value of 95% (Supplemental Data). When lineage sorting reaches the level where no simulated gene trees match the species tree, the critical value required to keep a 5% or less type 1 error rate rises to near 100% (i.e., no overlap between the frequency distribution of tree-to-tree distances from simulations compared to the gene tree distances is required to reject the null; Supplemental Data). This difference in the critical value and the gene tree distances is maintained in our results under the original parameters and the parameter variation chosen in the sensitivity analysis (Fig. 3).

However, despite these tests suggesting that, on average, a pair of genes sampled at random compared with an appropriate critical value in the way we describe will have an acceptable type 1 error rate, clearly for

this particular pair of genes we cannot know for certain whether the result is correct. It is possible that the genes chosen are unrepresentative of the original cloud of gene trees around the species tree(s). Outliers may be more distant from one another by chance and therefore produce a lack of overlap in simulated tree distances, even though only a single species tree (i.e., no hybridization) adequately represents the history of these organisms. A future improvement to this test might be to use more unlinked genes to reduce the effect of sampling outlying genes.

Another limitation of the test is that there may be a tendency for gene trees containing deeply coalescing alleles (i.e., that do not track speciation) to produce very similar trees under simulation, because the branch lengths (and therefore inferred duration of branches) are more often greater than in the species tree. An unrepresentative gene tree—one containing many deeply coalescing alleles—may therefore underestimate the variation due to lineage sorting alone that the species' population parameters would suggest, thereby overemphasizing the difference between gene trees (although the topological difference between the gene trees may still be representative of whatever process formed them). Our test has been designed to minimize this effect in the following way. If one gene tree underestimates variation due to lineage sorting, but the second does not, we fail to reject the null hypothesis if either gene's simulated trees variation is high enough (i.e., for the distribution of tree-to-tree distances to overlap the gene tree distance beyond the critical value). In this way, one gene (but not both) may be unrepresentative, but despite this the test may still work appropriately. Further testing would clearly be useful.

We have also assumed what we believe is an unrealistically large N_e that thereby favors lineage sorting. A large N_e is conservative with respect to excluding lineage sorting as an explanation for incongruence in *Medicago*. Although a higher N_e estimate (~940,000) than this has been reported for *Zea maize* ssp. *parviglumis* (Eyre-Walker et al., 1998) using polymorphism at *adh1*, more recent studies using microsatellite mutation rates have reported an N_e of only 38,500 (Vigouroux et al., 2002), a value significantly smaller than the single-gene estimation. Vigouroux et al. (2002) suggested that the disagreement between the estimations is probably due to the utilization of an inadequate rate of substitution that does not account for an apparent rate acceleration observed in the maize lineage (Gaut and Clegg, 1993; White and Doebley, 1999). In fact, N_e estimates based on isozyme data for a wide range of inbreeders and outcrossers reported means of 3500 and 7000, respectively (Schoen and Brown, 1991), suggesting that very large N_e may not be common among plant species. More recently, estimates have been published for three species of *Pinus*, ranging between 17,000 and 120,000 individuals (Syring et al., 2007). Although this suggests that outcrossing trees growing in large stands can achieve high N_e, the highest of these estimates is still half of our estimate for *Medicago*, reinforcing the probability that our estimate is unrealistically high.

Our estimate of N_e based on the *M. sativa* complex is unlikely to hold for the whole genus throughout its history, especially given the propensity to self-fertilize and the occurrence of population bottlenecks that may occur due to climate change and stochastic events. This argument suggests that lineage sorting is much less likely than hybridization for many of the observed incongruences.

Because our coalescence simulation approach estimates how likely a lineage is to hold multiple alleles from ancestral polymorphisms through to the next speciation event, it reduces the need for sampling numerous individuals per species. If N_e is too small and/or the number of generations between speciation events for a lineage too large, then all neutrally evolving alleles present in an extant species should be monophyletic and show phylogenetic coalescence that is younger than the species. Therefore, a large sample size of individuals from each species is not needed to reject lineage sorting, although the degree of hybridization is likely to be underestimated with a small sample (it should be noted that the estimation of N_e may use a large sample from one species with extrapolation to all species, as we have done here).

Our test indicates that hybridization is required to explain the incongruence observed among the two nuclear gene trees. However, we do not have a clear indication of how much hybridization has occurred. If our original parameter values are more accurate that the modifications made during the sensitivity analysis, then there is likely to have been numerous hybridization events—the gene tree difference compared to the background noise of lineage sorting is large. However, we cannot say precisely how many events have occurred.

To summarize, although the sorting of ancestral polymorphisms cannot be ruled out entirely, hybridization is the best explanation for most of the incongruence we report. Therefore, hybridization is supported as a pervasive and ongoing process throughout the history of *Medicago*.

The hybridization pattern observed in our data could be limited only to the genomic regions we studied. *Drosophila* hybridization studies have shown that gene flow between species is heterogeneous across the genome and bidirectional; however, at the single-locus level, gene flow seems to be unidirectional (Wang et al., 1997; Machado et al., 2002; Hey and Nielsen, 2004; Llopart et al., 2005). Variation of gene flow across the genome has also been observed in plants. Although analyses of multilocus variation between *Arabidopsis halleri* and *A. petraea* have shown gene flow between these species, haplotype sharing between them was observed only at the *GS* locus among eight loci (Ramos-Onsins et al., 2004). Thus, sampling much more of the *Medicago* genome is necessary to understand the extent of incongruence and the relationships among genomic regions following the same evolutionary history.

Other Data on Hybridization in Medicago

If hybridization in *Medicago* is as common as our results indicate, we would expect that other genes from these taxa might show patterns of either agreeing with one or the other of the two nuclear genes for the placement of some taxa, or identifying yet other relationships. Sequences from the ITS and ETS regions have also been used to infer the phylogeny of *Medicago,* although often with different sampling among studies (Bena et al., 1998a, 1998b, 1998c; Downie et al., 1998; Bena, 2001). Although not very resolved, ITS-ETS phylogenies show a number of supported relationships that are notable. For instance, the clade including *M. minima, M. tenoreana, M. lupulina, M. coronata,* and *M. disciformis* in CNGC5 was also observed in the ITS-ETS phylogeny (although the relationship among these taxa was not identical) but not in the *β-cop*-like tree. In contrast, *M. praecox* was part of the clade containing *M. truncatula* (clade 3) in the CNGC5 trees, but not in *β-cop*-like or ITS-ETS phylogenies. Instead, *M. praecox* was close to *M. heyniana* in the *β-cop*-like and ITS-ETS, although not strongly supported in the latter. Hybridization is a reasonable explanation for these patterns.

Hybridization explains the incongruence observed in our data, but the hybridization does not appear to be very recent because we observed only a single PCR product (therefore allele) in each individual, suggesting that sufficient time had elapsed for loci to become homozygous, thereby fixing alleles from either one progenitor or the other. Introgressive hybridization (Seehausen, 2004; Grant et al., 2005) and/or homoploid hybrid speciation (Ungerer et al., 1998; Gross et al., 2003; Rieseberg et al., 2003) will combine different phylogenetic signals within the same individual, making the inference of evolutionary histories a challenge (Grant and Grant, 1992). Several species of *Medicago* have shown signs of natural hybridization (Lesins and Lesins, 1979; Small et al., 1999; Baquerizo-Audiot et al., 2001). Attempts to transfer favorable variation into cultivated *M. sativa* have suggested the presence of a complicated gene-flow network among species of section *Medicago* and beyond (Oldenmeyer, 1956; Simon, 1965; Simon and Millington, 1967; Lesins, 1970, 1972; Lesins and Lesins, 1979; McCoy and Bingham, 1988, 2005; Haas and Bingham, 2005). Many of the interfertile species also have overlapping ranges and may once have grown in sympatry at the local level (Lesins, 1969; Lesins et al., 1971; Lesins and Lesins, 1979; Small and Jomphe, 1989; Small et al., 1999). Most of the pollinators of cross-pollinated *Medicago* are ground-nesting bees (Lesins and Lesins, 1979). An important representative of these bees, *Megachile rotundata*, is currently distributed worldwide (Bohart, 1972) and has been partially domesticated as an alfalfa (*M. sativa* spp. *sativa*) pollinator (Goulson, 2003). Originally from Eurasia (Bohart, 1972), *Megachile rotundata* has been described as a polylectic species, visiting a broad range of species in Fabaceae and Asteraceae. These factors taken together strongly suggest a present-day pollination biology that is likely to result in hybridization that may also have operated historically. Further, genetic barriers to gene flow do not appear to be strong. The presence of floral mechanisms associated with insect pollination in *Medicago* inbreeders also suggests that these lineages were ancestrally outcrossers or at least once had higher rates of outcrossing. The only morphological character that discriminates *Medicago* from closely related genera is an explosive tripping pollination mechanism (Small et al., 1987) that in other taxa is associated with insect pollination. This floral syndrome is present in cross- and self-pollinated taxa, probably allowing the latter species to exchange genes at low frequencies. Even *Medicago lupulina,* which is the only small-flowered selfer that lacks flower tripping, possesses vestigial floral morphology associated with the floral tripping mechanism (Small et al., 1987). Thus, *Medicago* species that are not outcrossing at present may have been at an earlier stage of speciation. Gene flow does not have to be recent to produce a footprint of hybridization or introgression. For instance, although *Drosophila pseudoobscura* and *D. persimilis* are reproductively isolated and have not recently exchanged genes, analyses of patterns of linkage disequilibrium have shown that gene flow between them continued for some time after these species split (Machado et al., 2002).

CONCLUSIONS

Incongruence among data sets is well documented in studies of plants (Vriesendorp and Bakker, 2005) but is also being found more frequently in studies of other eukaryotes (Rokas et al., 2003b). Although either hybridization or lineage sorting is usually invoked to explain this phenomenon once technical or analytical explanations have been rejected, the evidence does not always clearly favor one explanation over another, and often no firm conclusion can be reached (Near et al., 2004; Wanntorp et al., 2006). Many cases in plants have been attributed to hybridization (Rieseberg and Ellstrand, 1993; Vriesendorp and Bakker, 2005); likewise, cases of hybridization among animal species are also accumulating, e.g., in birds (Grant and Grant, 1992), insects (Buckley et al., 2006), and mammals (Ropiquet and Hassanin, 2006), suggesting that hybridization is probably more widespread than currently appreciated.

Given the extent of hybridization within *Medicago*, it is clear that a bifurcating topology is a grossly unrealistic representation of the origins of taxa in this genus. Network methods allow reticulation among taxa to be displayed, but summaries such as Figure 2 do not solve a more fundamental problem. The ability to use trees to understand character evolution, to time speciation events, and to make predictions about taxa from their nearest relatives is confounded by a reticulate history. How can we predict whether a given species might possess a certain character if it is known to have a hybrid origin, whether recent or ancient? When lineages appear to have multiple hybridizations occurring throughout their history between different groups, as appears to be the

case for many *Medicago* species, this problem is further compounded. Finally, given that many (if not most) morphological characters are underlain by multiple genes, each of which may reflect a different history in hybrid species, it is not surprising that classifications based on morphology often disagree with gene tree topologies.

ACKNOWLEDGEMENT

We thank the anonymous reviewers and editors for comments that improved the manuscript and Anna Monro for editing the final draft. Part of this work was carried out by using the resources of the Computational Biology Service Unit from Cornell University, which is partially funded by Microsoft Corporation. Work was partially supported by NSF DEB-0516673 to J.J.D.

REFERENCES

Adams, K. L., and J. D. Palmer. 2003. Evolution of mitochondrial gene content: Gene loss and transfer to the nucleus. Mol. Phylogenet. Evol. 29:380–395.

Baldauf, S. F., A. J. Roger, I. Wenk-Siefert, and W. F. Doolittle. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 290:972–977.

Bapteste, E., H. Brinkmann, J. A. Lee, D. V. Moore, C. W. Sensen, P. Gordon, L. Duruflé, T. Gaasterland, P. Lopez, M. Müller, and H. Phillippe. 2002. The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyotelium, Entamoeba, and Mastigamoeba. Proc. Natl. Acad Sci. USA 99:1414–1419.

Baquerizo-Audiot, E., B. Desplanque, J. M. Prosperi, and S. Santoni. 2001. Characterization of microsatellite loci in the diploid legume *Medicago truncatula* (barrel medic). Mol. Ecol. Notes 1:1–3.

Baum, B. R. 1968. A clarification of the generic limits of *Trigonella* and *Medicago*. Can. J. Bot. 46:741–746.

Bena, G. 2001. Molecular phylogeny supports the morphologically based taxonomic transfer of the "medicagoid" *Trigonella* species to the genus *Medicago* L. Plant Syst. Evol. 229:217–236.

Bena, G., M. F. Jubier, I. Olivieri, and B. Lejeune. 1998a. Ribosomal external and internal transcribed spacers: Combined use in the phylogenetic analysis of *Medicago* (Leguminosae). J. Mol. Evol. 46:299–306.

Bena, G., B. Lejeune, J.-M. Prosperi, and I. Olivieri. 1998b. Molecular phylogenetic approach for studying life-history evolution: The ambiguous example of the genus *Medicago* L. Proc. R. Soc. Lond. Biol. 265:1141–1151.

Bena, G., J. M. Prosperi, B. Lejeune, and I. Olivieri. 1998c. Evolution of annual species of the genus *Medicago*: A molecular phylogenetic approach. Mol. Phylogenet. Evol. 9:552–559.

Bergthorsson, U., K. L. Adams, B. Thomason, and J. D. Palmer. 2003. Widespread horizontal gene transfer of mitochondrial genes in flowering plants. Nature 424:197–201.

Bingham, E. T. 2005. Field observations on progeny of sac plants. Medicago Genet. Rep. 5 (http://www.medicago-reports.org/)

Bohart, G. E. 1972. Management of wild bees for the pollination of crops. Annu. Rev. Entomol. 17:287–312.

Brummer, E. C., J. H. Bouton, and G. Kochert. 1995. Analysis of annual *Medicago* species using RAPD markers. Genome 38:362–367.

Brummer, E. C., G. Kochert, and J. H. Bouton. 1991. RFLP variation in diploid and tetraploid alfalfa. Theor. Appl. Genet. 83:89–96.

Buckley, T. R., M. Cordeiro, D. C. Marshall, and C. Simon. 2006. Differentiating between hypotheses of lineage sorting and introgression in New Zealand alpine cicadas (*Maoricicada* Dugdale). Syst. Biol. 55:411–425.

Clement, W. M., and E. H. Stanford. 1963. Pachytene studies at the diploid level in *Medicago* . Crop Sci. 3:147–150.

Demesure, B., N. Sodzi, and R. J. Pettit. 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. Mol. Ecol. 4:124–131.

Downie, S. R., D. S. Katz Downie, E. J. Rogers, H. L. Zujewski, and E. Small. 1998. Multiple independent losses of the plastid *rpoC1* intron in *Medicago* (Fabaceae) as inferred from phylogenetic analyses of nuclear ribosomal DNA internal transcribed spacer sequences. Can. J. Bot. 76:791–803.

Doyle, J. J. 1992. Gene trees and species trees: Molecular systematics as one-character taxonomy. Syst. Bot. 17:144–163.

Doyle, J. J., J. L. Doyle, and A. H. D. Brown. 1999. Incongruence in the diploid B-genome species complex of *Glycine* (Leguminosae) revisited: Histone H3-D alleles versus chloroplast haplotypes. Mol. Biol. Evol. 16:354–362.

Driskel, A. C., C. Ané, J. G. Burleigh, M. M. McMahon, B. C. O'Meara, and M. J. Sanderson. 2004. Prospects for building the tree of life from large sequence databases. Science 306:1172–1174.

Eyre-Walker, A., R. L. Gaut, H. Hilton, D. L. Feldman, and B. S. Gaut. 1998. Investigation of the bottleneck leading to the domestication of maize. Proc. Natl. Acad. Sci. USA 95:4441–4446.

Fisher, A., V. Wiebe, S. Pääbo, and M. Przeworski. 2004. Evidence for a complex demographic history of chimpanzees. Mol. Biol. Evol. 21:799–808.

Fulton, T. M., R. Van der Hoeven, N. T. Eannetta, and S. D. Tanksley. 2002. Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14:1457–1467.

Gaut, B. S., and M. T. Clegg. 1993. Molecular evolution of the *Adh1* locus in the genus *Zea*. Proc. Natl. Acad. Sci. USA 90:5095–5099.

Gillies, C. B. 1968. The pachytene chromosomes of diploid *Medicago sativa*. Can. J. Genet. Cytol. 10:788–793.

Gillies, C. B. 1971. Pachytene studies in 2n = 14 species of *Medicago*. Genetica 42:278–298.

Gillies, C. B. 1972a. Pachytene chromosomes of perennial *Medicago* species. I. Species closely related to *M. sativa*. Hereditas 72:277–288.

Gillies, C. B. 1972b. Pachytene chromosomes of perennial *Medicago* species. II. Distantly related species whose karyotypes resemble *M. sativa*. Hereditas 72:289–302.

Gillies, C. B. 1972c. Pachytene chromosomes of perennial *Medicago* species. III. Unique karyotypes of *M. hybrida* Trautv. and*M. suffruticosa* Ramond. Hereditas 71:303–310.

Goulson, D. 2003. Effects of introduced bees on native ecosystems. Annu. Rev. Ecol. Evol. Syst. 34:1–26.

Grant, P. R., and B. R. Grant. 1992. Hybridization of bird species. Science 265:193–197.

Grant, P. R., B. R. Grant, and K. Petren. 2005. Hybridization in the recent past. Am. Nat. 166:56–67.

Gross, B. L., A. E. Schwarzbach, and L. H. Rieseberg. 2003. Origin(s) of the diploid hybrid species *Helianthus deserticola* (Asteraceae). Am. J. Bot. 90:1708–1719.

Haas, T., and E. T. Bingham. 2005. Large flowers on sac plants in winter greenhouse. Medicago Genet. Rep. 5 (http://www.medicago-reports.org/)

Hey, J., and R. Nielsen. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with application to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167:747–760.

Ho, K. M., and K. J. Kasha. 1972. Chromosome homology at pachytene in diploid *Medicago sativa*, *M. falcata* and their hybrids. Can. J. Genet. Cytol. 14:829–838.

Holland, B. R., K. T. Huber, V. Moulton, and P. J. Lockhart. 2004. Using consensus networks to visualize contradictory evidence for species phylogeny. Mol. Biol. Evol. 21:1459–1461.

Holland, B. R., L. S. Jermiin, and V. Moulton. 2006. Improved consensus network techniques for genome-scale phylogeny. Mol. Biol. Evol. 23:848–855.

Huson, D. H., and D. Bryant. 2005. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23:254–267.

Ivanov, A. I. 1977. History, origin and evolution of the genus *Medicago*, subgenus *Falcago*. Bull. Appl. Genet. Plant Breed. 59:3–40.

Kass, R. E., and A. E. Raftery. 1995. Bayes factors. J. Am. Stat. Assoc. 90:773–795.

Kidwell, K. K., D. F. Austin, and T. C. Osborn. 1994. RFLP evaluation of nine *Medicago* accessions representing the original germplasm sources for the North American alfalfa cultivars. Crop Sci. 34:230–236.

Kolaczkowski, B., and J. W. Thornton. 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. Nature 431:980–984.

Lavin, M., P. S. Herendeen, and M. F. Wojciechowski. 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. Syst. Biol. 54:575–594.
Lee, M. S. Y. 1998. Uninformative characters and apparent conflict between molecules and morphology. Mol. Biol. Evol. 18:676–680.
Lesins, K. A. 1969. Relationship of taxa in genus *Medicago* as revealed by hybridization. IV. *M. hybrida* x *M. suffruticosa*. Can. J. Genet. Cytol. 11:340–345.
Lesins, K. A. 1970. Interspecific crosses involving alfalfa. V. *Medicago saxatilis* x *M. sativa* with reference to *M. cancellata* and *M. rhodopea*. Can. J. Genet. Cytol. 12:80–86.
Lesins, K. A. 1972. Interspecific hybrids involving alfalfa. VII. *Medicago sativa* x *M. rhodopea*. Can. J. Genet. Cytol. 14:221–226.
Lesins, K. A., and C. B. Gillies. 1972. Taxonomy and cytogenetics of *Medicago*. Agron. Monogr. 15:53–86.
Lesins, K. A., and I. Lesins. 1979. Genus *Medicago* (Leguminosae). A taxogenetic study. Dr. W. Junk, The Hague.
Lesins, K. A., S. M. Singh, and A. Erac. 1971. Relationship of taxa in the genus *Medicago* as revealed by hybridization. V. Section Intertextae. Can. J. Genet. Cytol. 13:335–346.
Llopart, A., D. Lachaise, and J. Coyne. 2005. Multilocus analysis of introgression between two sympatric species of *Drosophila*: *Drosophila yakuba* and *D. santomea*. Genetics 171:197–210.
Lopez, P., D. Casane, and H. Philippe. 2002. Heterotachy, an important process of protein evolution. Mol. Biol. Evol. 19:1–7.
Lukens, L., F. Zou, D. Lydiate, I. Parkin, and T. C. Osborn. 2003. Comparison of *Brassica oleracea* genetic map with the genome of *Arabidopsis thaliana*. Genetics 164:359–372.
Machado, C. A., R. M. Kliman, J. E. Market, and J. Hey. 2002. Infering the history of speciation from multilocus DNA sequence data: The case of *Drosophila pseudoobscura* and close relatives. Mol. Biol. Evol. 19:472–488.
Maddison, W. P. 1997. Gene trees in species trees. Syst. Biol. 46:523–536.
Maddison, W. P., and L. L. Knowles. 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55:21–30.
Mariani, A., F. Pupilli, and O. Calderini. 1996. Cytological and molecular analysis of annual species of the genus *Medicago*. Can. J. Bot. 74:299–307.
McCoy, T. J., and E. T. Bingham. 1988. Cytology and cytogenetics of alfalfa. Agron. Monogr. 29:737–776.
Michaels, S., and R. Amasino. 2001. High throughput isolation of DNA and RNA in 96-well format using a paint shaker. Plant Mol. Biol. Rep. 19:227–233.
Muangprom, A., S. G. Thomas, T. P. Sun, and T. C. Osborn. 2005. A novel dwarfing mutation in a green revolution gene from *Brassica rapa*. Plant Physiol. 137:931–938.
Muller, M. H., C. Poncet, J. M. Prosperi, S. Santoni, and J. Ronfort. 2006. Domestication history in the *Medicago sativa* species complex: Inferences from nuclear sequence polymorphism. Mol. Ecol. 15:1589–1602.
Near, T. J., D. I. Bolnick, and P. C. Wainwright. 2004. Investigating phylogenetic relationships of sunfishes and black basses (Actinopterygii: Centrarchidae) using DNA sequences from mitochondrial and nuclear genes. Mol. Phylogenet. Evol. 32:344–357.
Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.
Nei, M., and S. Kumar. 2000. Molecular evolution and phylogenetics. Oxford University Press, New York.
Nylander, J. A. A., F. Ronquist, J. P. Huelsenbeck, and J. L. Nieves-Aldry. 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53:47–67.
Oldenmeyer, R. K. 1956. Distant relatives of cultivated alfalfa, *Medicago ruthenica* and *M. platycarpa*. Agron. J. 48:583–584.
Penny, D., and M. D. Hendy. 1985. The use of tree comparison metrics. Syst. Zool. 34:75–82.
Peters, J. L., Y. Zhuravlev, I. Fefelov, A. Logie, and K. E. Omland. 2007. Nuclear loci and coalescent methods support ancient hybridization as cause of mitochondrial paraphyly between gadwall and falcated duck (*Anas* spp.). Evolution. 61:1992–2006.
Quiros, C. F. 1982. Tetrasomic segregation for multiple alleles in alfalfa. Genetics 101:117–127
Quiros, C. F., and G. R. Bauchan. 1988. The genus *Medicago* and the origin of the *Medicago sativa* complex. Agron. Monogr. 29:737–776.
Ramos-Onsins, S. E., B. E. Stranger, T. Mitchell-Olds, and M. Aguadé. 2004. Multilocus analysis of variation in the closely related species *Arabidopsis halleri* and *A. lyrata*. Genetics 166:373–388.
Rannala, B., and Z. Yang. 2003. Bayes estimations of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics 164:1645–1656.
Rieseberg, L. H. 1991. Homoploid reticulate evolution in *Helianthus* (Asteraceae): Evidence from ribosomal genes. Am. J. Bot. 78:1218–1237.
Rieseberg, L. H., and N. C. Ellstrand. 1993. What can molecular and morphological markers tell us about plant hybridization? Crit. Rev. Plant Sci. 12:213–241.
Rieseberg, L. H., O. Raymond, D. M. Rosenthal, Z. Lai, K. Livingstone, T. Nakazato, J. L. Durphy, A. E. Schwarzbach, L. A. Donovan, and C. Lexer. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. Science 301:1211–1216.
Rieseberg, L. H., B. Sinervo, C. R. Linder, M. C. Ungerer, and D. M. Arias. 1996. Roles of gene interactions in hybrid speciation: Evidence from ancient and experimental hybrids. Science 272:741–745.
Rokas, A., N. King, J. Finnerty, and S. B. Carroll. 2003a. Conflicting phylogentic signals at the base of the metazoan tree. Evol. Dev. 5:346–359.
Rokas, A., D. Krüger, and S. B. Carroll. 2005. Animal evolution and the molecular signature of radiations compressed in time. Science 310:1933–1938.
Rokas, A., B. L. Williams, N. King, and S. B. Carroll. 2003b. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425:798–804.
Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.
Ropiquet, A., and A. Hassanin. 2006. Hybrid origin of the Pliocene ancestor of wild goats. Mol. Phylogenet. Evol. 41:395–404.
Sanderson, M. J. 2003. r8s; inferring absolute rates of evolution and divergence times in the absence of a molecular clock. Bioinformatics 19:301–302.
Sang, T., and Y. Zhong. 2000. Testing hybridization hypotheses based on incongruent gene trees. Syst. Biol. 49:422–434.
Schoen, D. J., and A. H. D. Brown. 1991. Intraspecific variation in population gene diversity and effective population size correlates with mating system in plants. Proc. Natl. Acad. Sci. USA 88:4494–4497.
Seah, S., K. Sivasithamparam, A. Karakousis, and E. S. Lagudah. 1998. Cloning and characterization of a family of disease resistance gene analogs from wheat and barley. Theor. Appl. Genet. 97:937–945.
Seehausen, O. 2004. Hybridization and adaptive radiation. Trends Ecol. Evol. 19:198–207.
Simon, J. P. 1965. Relationship in annual species of *Medicago*. II. Interspecific crosses between *M. tornata* (L.) Mill. and *M. littoralis* Rhode. Aust. J. Agric. Res. 16:51–60.
Simon, J. P., and A. J. Millington. 1967. Relationship in annual species of *Medicago*. III. The complex *M. littoralis* Rhode-*M. truncatula* Gaertn. Aust. J. Bot. 15:35–73.
Small, E. 1981. A numerical analysis of major groupings in *Medicago* employing traditionally used characters. Can. J. Bot. 59:1553–1577.
Small, E. 1990a. *Medicago rigiduloides*, a new species segregated from *M. rigidula*. Can. J. Bot. 68:2614–2617.
Small, E. 1990b. *Medicago syriaca*, a new species. Can. J. Bot. 68:1473–1478.
Small, E., and B. Brookes. 1991. A clarification of *Medicago sinkiae*. Can. J. Bot. 69:100–106.
Small, E., C. W. Crompton, and B. S. Brookes. 1981. The taxonomic value of floral characters in tribe Trigonelleae (Leguminosae), with special reference to *Medicago*. Can. J. Bot. 59:1578–1598.
Small, E., and M. Jomphe. 1989. A synopsis of the genus *Medicago* (Leguminosae). Can. J. Bot. 67:3260–3294.
Small, E., P. Lassen, and B. S. Brookes. 1987. An expanded circumscription of *Medicago* (Leguminosae, Trifolieae) based on explosive flower tripping. Willdenowia 16:415–437.
Small, E., S. I. Warwick, and B. S. Brookes. 1999. Allozyme variation in relation to morphology in *Medicago* Sect. *Spirocarpos* subsect. *Intertextae* (Fabaceace). Plant Syst. Evol. 214:29–47.
Stanford, E. H. 1951. Tetrasomic inheritance in alfalfa. Agron. J. 43:222–225.

Swofford, D. L. 1998. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Version 4.0b10. Sinauer Associates, Sunderland, Massachusetts.
Swofford, D. L., P. J. Waddell, J. P. Huelsenbeck, P. G. Foster, P. O. Lewis, and J. S. Rogers. 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. Syst. Biol. 50:525–539.
Syring, J., K. Farrell, R. Businsky, R. Cronn, A. Liston. 2007. Widespread genealogical nonmonophyly in species of *Pinus* subgenus *Strobus*. Syst. Biol. 56:163–181.
Trueman, J. W. H. 1998. Reverse successive weighting. Syst. Biol. 47:733–737.
Ungerer, M. C., S. J. E. Baird, J. Pan, and L. H. Rieseberg. 1998. Rapid hybrid speciation in wild sunflowers. Proc. Natl. Acad. Sci. USA 95:11757–11762.
Valizadech, M., K. K. Kang, A. Kanno, and T. Kameya. 1996. Analysis of genetic distance among nine *Medicago* species by using DNA polymorphisms. Breed. Sci. 46:7–10.
Vigouroux, Y., J. S. Jaqueth, Y. Matsuoka, O. S. Smith, W. D. Beavis, J. S. C. Smith, and J. F. Doebley. 2002. Rate and pattern of mutation at microsatellite loci in maize. Mol. Biol. Evol. 19:1251–1260.
Vriesendorp, B., and F. T. Bakker. 2005. Reconstructing patterns of reticulate evolution in angiosperms: What can we do? Taxon 54:593–604.
Wall, J. 2003. Estimating ancestral population sizes and divergence times. Genetics 163:395–404.
Wang, R. L., J. Wakeley, and J. Hey. 1997. Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. Genetics 147:1091–1106.
Wanntorp, L., A. Kocyan, and S. S. Renner. 2006. Wax plants disentangled: A phylogeny of *Hoya* (Marsdenieae, Apocynaceae) inferred from nuclear and chloroplast DNA sequences. Mol. Phylogenet. Evol. 39:722–733.
Wendel, J. F., and J. J. Doyle. 1998. Phylogenetic incongruence: Window into genome history and molecular evolution. Pages 265–296 *in* Molecular systematics of plants II: DNA sequencing (D. E. Soltis, P. S. Soltis, and J. J. Doyle, eds.). Kluwer Academic Publishers, Norwell.
White, S. E., and J. F. Doebley. 1999. The molecular evolution of terminal *ear1*, a regulatory gene in the genus *Zea*. Genetics 153:1455–1462.
Yang, Z. 1997. On the estimation of ancestral population sizes of modern humans. Genet. Res. 69:111–116.

First submitted 22 February 2007; reviews returned 7 May 2007; final acceptance 14 March 2008
Associate Editor: Susanne Renner
Editors in Chief: Rod Page and Jack Sullivan

A sample of the pod diversity observed in *Medicago*. Pods are organized from left to right then top to bottom: *M. italica* (Miller) Fiori, *M. soleirolii* Duby, *M. lesinsii* E.Small, *M. shepardii* Post, *M. suavagei* Negre, *M .murex* Willd, *M. rotata* Boiss, *M. lanigera* Winkl. & Fedtsch, *M. secundiflora* Durieu, *M. heyniana* Greuter, *M. orbicularis* (L.) Bart, *M. radiata* L., *M. plicata* (Boiss) Sirjaev, *M. platicarpa* (L.) Trautv., *M. lupulina* L., *M. tenoreana* Ser., *M. huberi* E. Small, *M. biflora* (Griseb) E. Small, *M. laxispira* Heyn, and *M. truncatula* Gaertn.