Multiple independent losses of the plastid *rpo*C1 intron in *Medicago* (Fabaceae) as inferred from phylogenetic analyses of nuclear ribosomal DNA internal transcribed spacer sequences

Stephen R. Downie, Deborah S. Katz-Downie, Erica J. Rogers, Heidi L. Zujewski, and Ernest Small

Abstract: A previous polymerase chain reaction based survey for the occurrence of the intron in chloroplast gene *rpo*C1 revealed its absence in one of the eight species of *Medicago* (Fabaceae; Trifolieae) examined. We extend the survey of *Medicago* to include 65 of the 86 species, representing all 12 sections and seven of the eight subsections recognized in the most recent comprehensive treatment of the genus. Our results indicate that 17 species from five sections lack the intron and that three of these sections are heterogeneous with regard to intron content. DNA sequencing across the *rpo*C1 intron–exon boundary in three of these species reveals the precise excision of the intron from the gene. Phylogenies derived from nuclear ribosomal DNA internal transcribed spacer sequences, estimated using maximum parsimony and maximum likelihood methods, suggest that most of the previously recognized sections in *Medicago* are not monophyletic as currently circumscribed. Furthermore, these results suggest that the *rpo*C1 intron has been lost independently a minimum of three times during the evolution of the group. The occurrence of multiple independent intron losses severely reduces the utility of this character as a phylogenetic marker in *Medicago*.

Key words: *Medicago*, Fabaceae, chloroplast DNA, *rpo*C1 intron, nuclear ribosomal DNA internal transcribed spacer (ITS) sequences.

Résumé : Un premier survol basé sur la polymérisation en chaîne pour déceler la présence d'intron dans le gène chloroplastique *rpo*C1 a révélé son absence dans une des huit espèces de *Medicago* (Fabaceae; Trifoliae) examinées. L'étude a été poursuivie en effectuant un examen du genre *Medicago* incluant 65 des 86 espèces, représentant l'ensemble des 12 sections et sept des huit sous-sections reconnues dans le regroupement le plus récent du genre. Les résultats indiquent que 17 espèces appartenant à cinq sections ne possèdent pas d'intron, et que trois sections sont hétérogènes quant à la présence d'introns. Le séquençage de l'ADN sur l'ensemble de la région limitrophe *rpo*C1 intron–exon, chez trois de ces espèces, révèle le site précis d'excision de l'intron du gène. Les phylogénies dérivées des séquences de l'espaceur interne transcrit de l'ADN ribosomal nucléaire, évaluées par les méthodes de parcimonie et de probabilité de ressemblance maximum, suggèrent que la plupart des sections précédemment reconnues du genre *Medicago* ne sont pas monophylétiques comme on le défini couramment. De plus, ces résultats suggèrent que l'intron du *rpo*C1 a été perdu indépendamment un minimum de trois fois au cours de l'évolution du groupe. L'incidence de pertes indépendantes multiples de l'intron réduit fortement l'utilité de ce caractère comme marqueur phylogénétique dans le genre *Medicago.*

Mots clés : *Medicago,* Fabaceae, ADN chloroplastique, intron *rpo*C1, séquences de l'espaceur interne transcrit (ITS) de l'ADN ribosomal nucléaire.

[Traduit par la rédaction]

Introduction

Major structural rearrangements of the chloroplast genome, such as changes in gene order and intron presence, can provide useful characters for phylogenetic reconstruction (Downie and Palmer 1992). Intron loss characters have been used to circumscribe taxa at the generic, tribal, familial, and ordinal levels (Downie et al. 1991, 1996; Doyle et al. 1995; Bailey et al. 1997; Lai et al. 1997; Campagna and Downie 1998). At lower taxonomic ranks, however, the utility of intron losses in demarcating monophyletic groups is not as clear. Owing to the conservative nature of chloroplast DNA (cpDNA) structural evolution among photosynthetic angiosperms (Palmer 1991), often only one exemplar from a genus is included in a survey for such characters.

A previous PCR (polymerase chain reaction) based survey

S.R. Downie,[1] D.S. Katz-Downie, E.J. Rogers, and H.L. Zujewski. Department of Plant Biology, University of Illinois, Urbana, IL 61801, U.S.A.
E. Small. Eastern Cereal and Oilseed Research Centre, Agriculture Canada, Central Experimental Farm, Ottawa, ON K1A 0C6, Canada.

[1] Author for correspondence. e-mail: sdownie@uiuc.edu

© 1998 NRC Canada

for the occurrence of the intron in chloroplast gene *rpo*C1 across a broad representation of angiosperm species revealed its absence in a minimum of six independent lineages, including one of the eight species of *Medicago* L. examined (Downie et al. 1996). The genus *Medicago* (Fabaceae; Trifolieae) consists of 86 annual and perennial species (and 18 infraspecific taxa) divided most recently into 12 sections and 8 subsections (Small and Jomphe 1989; Small 1990*a*, 1990*b*; Small and Brookes 1991). The genus is important economically and includes many forage and cover crops, such as alfalfa. The variable presence of the *rpo*C1 intron in *Medicago* cpDNAs provides an opportunity to assess the utility of intron loss in documenting monophyly of taxa below the generic level.

In this study, we report results of a survey of the occurrence of the *rpo*C1 intron in *Medicago*. To assess the utility of intron loss as a phylogenetic marker, the distribution of this character is evaluated against a phylogeny for the group inferred on the basis of phylogenetic analyses of nuclear ribosomal DNA internal transcribed spacer (ITS) sequences. Relationships within *Medicago* are inferred based on these ITS sequence data and the distribution of *rpo*C1 intron loss.

Methods

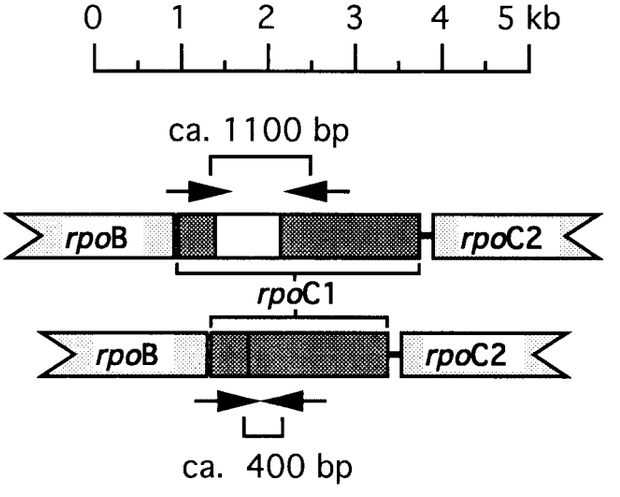
Plant material

A total of 68 accessions (65 species) of *Medicago*, representing all 12 sections and seven of the eight subsections recognized by Small and Jomphe (1989), was obtained from various sources (Table 1). Material was unavailable for the rare *Medicago retrorsa* (Boiss.) E. Small, the only member of section *Buceras* subsection *Deflexae*. Plant material was obtained either directly from the field, from plants cultivated from seed at our respective institutions, or from our herbarium specimens. Total genomic DNAs were isolated from fresh leaf or herbarium material using a slightly modified version of Doyle and Doyle's (1987) CTAB procedure. Upon the addition of 1.0% sodium bisulfite and 1.0% polyvinylpyrrolidone (PVP) to the 2× CTAB isolation buffer, approximately 30 mg of leaf tissue was ground in a mortar with pestle. The grindate was incubated at 60°C for 30 min prior to treatment with chloroform – isoamyl alcohol. Dried herbarium leaf tissue from samples as old as 25 years yielded DNAs suitable for PCR and sequencing.

*Rpo*C1 intron analysis

All 68 available accessions of *Medicago* were surveyed for the presence or absence of the plastid *rpo*C1 intron. Included were *M. sativa* ssp. *sativa*, previously reported to contain the intron, and *M. suffruticosa* ssp. *leiocarpa*, previously reported not to contain the intron (Downie et al. 1996; Fig. 1). For each genomic DNA, the entire *rpo*C1 intron (if present) and portions of its flanking exons were amplified using the PCR technique (Fig. 1). Details of the amplification reactions were the same as those presented in Downie and Katz-Downie (1996), with the exception of an increase in the $MgCl_2$ concentration from 1.5 to 3.0 mM. Information on primer construction and their precise locations in the tobacco (*Nicotiana tabacum* L.) chloroplast genome is presented in Downie et al. (1996); both primers were designed to anneal to the highly conserved 5′ and 3′ exon regions of the *rpo*C1 gene. Each PCR reaction proceeded as follows: (*i*) 1 min at 94°C; (*ii*) 1 min at 53°C; and (*iii*) 1 min at 72°C. The first cycle was preceded by an initial denaturation step of 30 s at 94°C. A 10-min 72°C extension period followed completion of the 35 thermal cycles. Each set of reactions was monitored by the inclusion of positive (*M. sativa* ssp. *sativa* and *M. suffruticosa* ssp. *leiocarpa*) and negative (no template) controls. The ensuing PCR fragments were separated by electrophoresis in 1% agarose gels, stained with ethidium bromide,

Fig. 1. Structural organization of the chloroplast *rpo*C1 gene and flanking regions; only portions of genes *rpo*B and *rpo*C2 are indicated. Coding regions are indicated by shaded boxes; the intron in gene *rpo*C1 is indicated by an open box. Scale units are in kilobase pairs (kb). In *Medicago sativa* ssp. *sativa* and the vast majority of angiosperm chloroplast genomes examined, the *rpo*C1 gene is interrupted by an intron of about 700 base pairs (bp). The two arrows indicate the relative positions of the two PCR primers used to assess the presence or absence of the intron. When the intron is present, the resultant PCR product is approximately 1.1 kb in size. When the intron is absent, such as in *Medicago suffruticosa* ssp. *leiocarpa*, the resultant PCR product is reduced by the size of the intron and is about 400 bp in size. Forward and reverse primers, written 5′ to 3′, are ACGTCTTCCTAGYTAYATHGC and AATAGACAYAANACCATCCA, respectively.



and sized against *Eco*RI–*Hin*dIII digested lambda DNA standards and (or) the positive controls. Successful PCR amplifications resulted in a single DNA band of about 1100 base pairs (bp) when the intron was present or about 400 bp when the intron was absent and precisely excised (Fig. 1). To characterize the precise point of excision of the intron, small-sized PCR products were sequenced for two taxa (Table 1). The PCR product from a third intronless taxon, *M. suffruticosa* ssp. *leiocarpa*, was sequenced previously (Downie et al. 1996). The sequencing strategy used was the same as outlined in Downie and Katz-Downie (1996). In the present study, only material of *Medicago* was examined for *rpo*C1 intron loss. The plesiomorphic state for *Medicago* is assumed to be presence of the *rpo*C1 intron. This intron is present in the chloroplast genomes of most land plants, including single representatives from Fabaceae genera *Acacia*, *Albizia*, *Caesalpinia*, *Calliandra*, *Enterolobium*, *Glycine*, *Inga*, *Lablab*, *Lysiloma*, *Phaseolus*, *Pisum*, *Pithecellobium*, and *Trifolium* (Downie et al. 1996).

ITS sequence analysis

Complete sequences of the two ITS regions of 18S–26S nuclear ribosomal DNA were obtained for 60 of the 68 *Medicago* accessions examined for *rpo*C1 intron loss (Table 1). Details of the DNA purification, PCR amplification, and sequencing strategies used are the same as described in Downie and Katz-Downie (1996). In summary, the sequence data were obtained through direct manual sequencing of double-stranded templates derived from the PCR procedure. Forward primers ITS 3 and ITS 5 and reverse primers ITS 2 and ITS 4 (White et al. 1990) were each used in the sequencing of each template

© 1998 NRC Canada

Table 1. *Medicago* accessions surveyed for the presence or absence of the cpDNA *rpo*C1 intron.

Taxon	Accession code[a]	*rpo*C1 intron: present (+) absent (—)	GenBank accession numbers for ITS sequences
Section *Dendrotelis* (Vassilcz.) Lassen (2 species)			
M. arborea L. *var. arborea*	M2071	+	AF028367, AF028427
M. strasseri Greuter, Matthäs, & Risse	C6	+	AF028363, AF028423
Section *Medicago* (12 species)			
M. sativa L. ssp. *sativa*	SRD862	+	AF028357, AF028417
M. sativa L. ssp. *caerulea* (Less. ex Ledeb.) Schmalh.	USDA577544	+	AF028358, AF028418
M. sativa L. ssp. *glomerata* (Balbis) Rouy	USDA577566	+	AF028359, AF028419
M. sativa L. ssp. *falcata* (L.) Arcangeli var. *falcata*	M355c	+	AF028360, AF028420
M. papillosa Boiss. ssp. *macrocarpa* (Boiss.) Urban	M1099b	+	AF028397, AF028457
M. prostrata Jacq. ssp. *prostrata*	M279	+	AF028402, AF028462
M. cancellata M. Bieb.	M1024	—*	AF028368, AF028428
M. rhodopea Velen.	GAS4168	+	AF028405, AF028465
M. saxatilis M. Bieb.	H3324	+	Not sequenced
M. daghestanica Rupr.	M1023	+	AF028362, AF028422
M. pironae Vis.	M159	+	AF028400, AF028460
M. hybrida (Pourr.) Trautv.	M57	—*	AF028366, AF028426
M. suffruticosa Ramond ex DC. ssp. *leiocarpa* (Bentham) P. Fourn.	USDA516912	—*	AF028361, AF028421
M. marina L.	M110	+	AF028389, AF028449
Section *Carstiensae* Kozukharov (1 species)			
M. carstiensis Wulf.	M472a	+	AF028377, AF028437
Section *Spirocarpos* Ser.			
Subsection *Pachyspireae* (Urb.) Heyn (13 species)			
M. syriaca E. Small	M1348	+	AF028411, AF028471
M. italica (Miller) Fiori	M1452	?[b]	AF028379, AF028439
M. truncatula Gaertn. forma *truncatula*	USDA577634	+	AF028364, AF028424
M. doliata Carmign.	M368	?[b]	AF028374, AF028434
M. turbinata (L.) All.	USDA577644	+	AF028365, AF028425
M. rigidula (L.) All.	M1298	+	AF028406, AF028466
M. rigiduloides E. Small	M1293	+	AF028407, AF028467
M. sinskiae Uljan.	M1578	+	AF028410, AF028470
M. constricta Durieu	M343	+	AF028372, AF028432
M. lesinsii E. Small	M1149	+	AF028387, AF028447
M. murex Willd.	M660	+	Not sequenced
Subsection *Rotatae* (Urb.) Heyn (6 species)			
M. rugosa Desr.	M741	+	AF028398, AF028458
M. scutellata (L.) Miller	M284	+	Not sequenced
M. blancheana Boiss.	M458	+	AF028376, AF028436
M. rotata Boiss.	M723	+	AF028408, AF028468
M. noeana Boiss.	M338	+	AF028395, AF028455
M. shepardii Post	GW979	+	AF028409, AF028469
Subsection *Intertextae* (Urb.) Heyn (4 species)			
M. intertexta (L.) Miller	M2009	+	AF028384, AF028444
M. ciliaris (L.) Krocker	M1935	+	AF028371, AF028431
M. muricoleptis Tin.	M1944	+	AF028394, AF028454
M. granadensis Willd.	M1951	+	AF028381, AF028441
Subsection *Leptospireae* (Urb.) Heyn (11 species)			
M. sauvagei Nègre	H2989	+	AF029355, AF028415
M. laciniata (L.) Miller	M598	?[b]	AF028356, AF028416
M. minima (L.) Bart.	M24	—	AF028391, AF028451
M. praecox DC.	ES257	+	AF028403, AF028463
M. coronata (L.) Bart.	M12	—	Not sequenced
M. polymorpha L.	USDA577440	+	AF028353, AF028413
M. laxispira Heyn	H51F	+	AF028386, AF028446
M. arabica (L.) Huds.	M364	+	AF028375, AF028435
M. tenoreana Ser.	M341	—	AF028412, AF028472

© 1998 NRC Canada

Table 1 (*concluded*).

Taxon	Accession code[a]	*rpo*C1 intron: present (+) absent (—)	GenBank accession numbers for ITS sequences
M. disciformis DC.	M1012	—	AF028373, AF028433
M. lanigera Winkl. & Fedtsch.	M346	—	AF028385, AF028445
Section *Geocarpa* E. Small (1 species)			
M. hypogaea E. Small	F177	—	AF028383, AF028443
Section *Lupularia* Ser. in DC. (2 species)			
M. lupulina L.	M1231	—	AF028388, AF028448
M. secundiflora Durieu	M339	+	Not sequenced
Section *Heynianae* Greuter (1 species)			
M. heyniana Greuter	M342	+	AF028382, AF028442
Section *Orbicularis* Urb. (1 species)			
M. orbicularis (L.) Bart.	M46	?[b]	AF028396, AF028456
Section *Hymenocarpos* Ser. (1 species)			
M. radiata L.	M225	+	AF028404, AF028464
Section *Platycarpae* E. Small (8 species)			
M. platycarpa (L.) Trautv.	M161	+	Not sequenced
M. ruthenica (L.) Ledebour	USDA568103	+	AF028354, AF028414
M. edgeworthii Sirjaev	M957	+	AF028378, AF028438
M. cretacea M. Bieb.	M348	+	Not sequenced
Section *Lunatae* (Boiss.) E. Small (4 species)			
M. biflora (Griseb.) E. Small	ES189	+	AF028370, AF028430
M. brachycarpa M. Bieb.	GAS3284	+	AF028369, AF028429
Section *Buceras* (Ser.) E. Small			
Subsection *Erectae* (Sirjaev) E. Small (15 species)			
M. fischeriana (Ser.) Trautv.	T154	—	AF028380, AF028440
M. medicaginoides (Retz.) E. Small	T295d	—	AF028390, AF028450
M. monantha (C.A. Meyer) Trautv.	T251	—	AF028393, AF028453
M. orthoceras (Kar. & Kir.) Trautv.	T173	—	AF028399, AF028459
M. polyceratia (L.) Trautv.	T153	—	AF028401, AF028461
Subsection *Reflexae* (Sirjaev) E. Small (1 species)			
M. monspeliaca (L.) Trautv.	T252	—	AF028392, AF028452
Subsection *Isthmocarpae* (Boiss.) E. Small (2 species)			
M. isthmocarpa (Boiss. & Bal.) E. Small	T156	—	Sequenced in part

Note: Those taxa for which nuclear ribosomal DNA ITS data were also obtained, and thus included in the phylogenetic analyses of these data, are indicated by GenBank accession numbers for separate ITS 1 and ITS 2 sequences, respectively. Asterisks show those three species whose intron absence has been confirmed by DNA sequencing. Classification of *Medicago* based on Small and Jomphe (1989), Small (1990*a*, 1990*b*), and Small and Brookes (1991). Numbers of species recognized in each section or subsection are indicated in parentheses. Material of the monotypic *Buceras* subsection *Deflexae* was not available for study.

[a]Abbreviations are as follows: C, ES, F, H, M, T, collection codes of E. Small; SRD, S.R. Downie; GAS, G.A. Stevenson; GW, G. Westmoreland; USDA, U.S. Department of Agriculture, Western Regional Plant Introduction Station, Pullman, Wash. Voucher specimens for USDA and Downie collections are deposited at ILL; all remaining vouchers are deposited at DAO (herbarium abbreviations from Holmgren et al. 1990).

[b]*rpo*C1 intron region could not be PCR amplified.

DNA. Both spacer regions were sequenced in their entirety on both strands. All sequences were aligned manually. Because sequence data for the intervening 5.8S gene were incomplete for many taxa, only the two spacer regions were included in the analysis. These ITS data have been deposited with GenBank (see Table 1 for GenBank accession numbers for separate ITS 1 and ITS 2 sequences) and the complete aligned data matrix can be obtained directly from the authors. Pairwise nucleotide differences were determined using the distance matrix option of PAUP (version 3.1.1; Swofford 1993). These divergence values were calculated simply as the proportion of divergent sites in each direct pairwise comparison with no provision made to account for multiple hits.

Phylogenetic analysis

The resulting DNA sequences, together with the published ITS sequences for putative outgroups *Trifolium nanum* Torr. and *Trifolium longipes* Nutt. var. *neurophyllum* (Greene) Martin ex Isely (Sanderson and Wojciechowski 1996), were analyzed using maximum parsimony (MP) and maximum likelihood (ML) methods. Using PAUP, the length of the shortest trees was determined by initiating 500 random addition replicate searches, with TBR branch swapping and MULPARS selected, but saving no more than five of the shortest trees from each search. These trees were then used as starting trees for TBR branch swapping, with a MAXTREE limit of 5000. All character state changes were weighted equally. Bootstrap values (Felsenstein 1985) were calculated from 100 replicate analyses using a heuristic search strategy, simple addition sequence of the taxa, and TBR branch swapping. To facilitate the bootstrap analysis, a MAXTREE limit of 100 trees per replicate was set, although it is acknowledged that, because of this restriction, the values obtained may not represent the best estimate of branch support. The number of additional steps required to force particular taxa into a monophyletic group was examined using the CON-

© 1998 NRC Canada

STRAINTS option of PAUP. Separate analysis of each spacer region was not done. Previous studies have indicated the high complementarity of spacer data and the greater phylogenetic resolution and internal support achieved in trees when both spacers are considered together than when either spacer is treated alone (reviewed in Baldwin et al. 1995). The distribution of the number of inferred changes per character over a subset of the maximally parsimonious trees was calculated using MacClade (version 3.01; Maddison and Maddison 1992). Gaps in the ITS sequence alignments were incorporated into the parsimony analysis by scoring each indel as a separate presence or absence (i.e., binary) character.

Using the program fastDNAml (version 1.0.6; Olsen et al. 1994), ML trees were inferred using a transition/transversion rate ratio of 2.0, randomizing the input order of sequences, and invoking the global branch swapping search option. The heuristic analysis was repeated until three separate runs, each starting with a different random number seed, produced the same highest (least negative) log likelihood value. Empirical base frequencies were derived from the sequence data and used in the maximum likelihood calculations. Bootstrapping of the maximum likelihood data was computationally prohibitive and was not done.

Outgroup selection

Parsimony analysis of the "temperate herbaceous clade" of legumes based on ITS sequence data revealed a weakly supported clade consisting of *Melilotus* and *Trifolium* as sister to *Medicago polymorpha* and *Medicago lupulina* (Sanderson and Wojciechowski 1996). The genera *Medicago*, *Trigonella*, and *Melilotus* comprise Trifolieae subtribe Trigonellinae; these genera, along with *Trifolium*, constitute tribe Trifolieae (Small 1987). *Medicago* and *Trigonella* are very closely related and, until recently, *Medicago* sections *Buceras* and *Lunatae* were placed in *Trigonella* (Small 1987).

All trees computed in this study were rooted with *Trifolium nanum* and *T. longipes*. When *Melilotus officinalis* (L.) Pallas and *Melilotus alba* Medikus replaced the two *Trifolium* species as outgroups, very minor changes in tree topology resulted. These changes occurred at the base of the trees, an area where branch support and resolution of relationships are poor (see Results). Material of *Trigonella*, arguably a more appropriate outgroup than either *Trifolium* or *Melilotus*, was not included in our investigation.

Evaluating the distribution of *rpo*C1 intron loss

Within the independent context of the phylogenies inferred from nucleotide sequence variation in the ITS region of nuclear ribosomal DNA, the evolutionary pattern of plastid *rpo*C1 intron loss in *Medicago* was hypothesized by Dollo parsimony optimization (Le Quesne 1974; Farris 1977). In the absence of studies suggesting otherwise, we make the assumption that intron loss is irreversible. Indeed, insertion of DNA into the angiosperm chloroplast genome has yet to be documented (Palmer 1991). In contrast, multiple independent intron losses are possible, having been reported for other plastid introns in other groups of flowering plants (Downie et al. 1991, 1996; Doyle et al. 1995; Lai et al. 1997; Campagna and Downie 1998). The STATE CHANGES and STASIS option of MACCLADE, under the assumption of Dollo parsimony, was used to calculate the exact number of intron losses across all 5000 maximally parsimonious topologies.

Results

*Rpo*C1 intron loss

Our PCR-based survey revealed two major size categories of PCR products: the first corresponding to the presence of the intron (1100-bp product), and the second corresponding to its loss (400-bp product; Fig. 1). For those species possessing the intron, detectable size variation was evident in only a few species, differing in size from the *M. sativa* PCR product by about 50 bp or less. Of the 68 *Medicago* accessions surveyed, the intron was present in 47 taxa (44 species) and absent in 17 species (Table 1). Four species (*M. italica*, *M. doliata*, *M. laciniata*, and *M. orbicularis*; Table 1) yielding sufficient DNA could not be amplified with the *rpo*C1 primers, despite the use of four serial dilutions and adjustments to the PCR parameters.

The 17 species of *Medicago* lacking the *rpo*C1 intron fall into 5 of the 12 sections recognized in the genus (Small and Jomphe 1989; Table 1): (*i*) Section *Medicago* (*M. cancellata*, *M. hybrida*, *M. suffruticosa* ssp. *leiocarpa*); (*ii*) Section *Spirocarpos* (*M. minima*, *M. coronata*, *M. tenoreana*, *M. disciformis*, *M. lanigera*); (*iii*) Section *Geocarpa* (*M. hypogaea*); (*iv*) Section *Lupularia* (*M. lupulina*); and (*v*) Section *Buceras* (*M. fischeriana*, *M. medicaginoides*, *M. monantha*, *M. orthoceras*, *M. polyceratia*, *M. monspeliaca*, *M. isthmocarpa*). With the exception of section *Buceras*, where the intron is missing from all examined representatives, and the monotypic section *Geocarpa*, where the intron is also absent, the three remaining sections were heterogeneous with regard to intron content. Specifically, of the 2, 12, and 34 species described for sections *Lupularia*, *Medicago*, and *Spirocarpos*, respectively (Small and Jomphe 1989; Table 1), one, three, and five species, respectively, lack the intron. Within section *Spirocarpos*, the five species lacking the intron are all in subsection *Leptospireae*; examined members of *Spirocarpos* subsections *Pachyspireae*, *Rotatae*, and *Intertextae* all possess the intron.

DNA sequencing of the *rpo*C1 intron region in *M. cancellata* and *M. hybrida* confirmed that the intron is indeed missing from this region in these taxa. A previous sequencing study revealed the absence of this intron in *Medicago suffruticosa* ssp. *leiocarpa* cpDNA (Downie et al. 1996). While these three taxa have been treated in the same section (Small and Jomphe 1989), the ITS-based phylogenies (described below) suggest that *M. cancellata* may be distantly related to *M. suffruticosa* and *M. hybrida*. Owing to similarly sized PCR products, it is expected that all small-sized PCR products represent the loss of the intron. Furthermore, in these three taxa at least, the *rpo*C1 gene has undergone a precise deletion of the intron with the two remaining exons juxtaposed into a single, uninterrupted gene. This precise excision of a plastid intron is not unexpected, being reported wherever an intron is absent (reviewed in Downie et al. 1996). Such precise loss may be explained by reverse transcription of a spliced transcript, followed by homologous recombination between the intronless complementary DNA (cDNA) and the original gene (Downie et al. 1991).

ITS sequence analysis

Of the 68 *Medicago* accessions examined for intron loss, 60 were included in the ITS sequencing study. With the exceptions of *M. coronata* and *M. isthmocarpa*, all species lacking the *rpo*C1 intron were represented. Owing to poor template quality, only partial ITS data were available for *M. isthmocarpa* (a region comprising some 85 bp of sequence from the ITS 1 region did not yield clean data). The ITS region of *M. coronata* could not be PCR amplified. Also included, as outgroups, were two species of *Trifolium* (Sanderson and Wojciechowski 1996). Alignment of all 62 ITS 1 and ITS 2 DNA sequences resulted in a matrix of 467 unambiguously aligned positions; their sequence characteristics are summarized in Table 2. Of these 467 positions, 296 (63.4%) were invariant,

© 1998 NRC Canada

Fig. 2. Strict consensus of five thousand 301-step trees of *Medicago* derived from equally weighted parsimony analysis of nuclear ribosomal DNA ITS 1 and ITS 2 sequences (CIs with and without uninformative characters = 0.661 and 0.547, respectively; RI = 0.788). Values above the nodes indicate the number of times a monophyletic group occurred in 100 bootstrap replicates. Sectional and subsectional classification of *Medicago* is based on Small and Jomphe (1989); complete taxon names are presented in Table 1. When the putative hybrid *M. cancellata* is excluded (indicated by the broken line) and the analysis rerun, the resultant strict consensus tree has the same topology as that depicted here (tree length = 297 steps; CIs with and without uninformative characters = 0.670 and 0.553, respectively; RI = 0.794).

Table 2. Sequence characteristics of the two nuclear ribosomal DNA internal transcribed spacers, separately and combined, in 60 accessions of *Medicago* and two species of *Trifolium*.

Sequence characteristic	ITS 1	ITS 2	ITS 1 and ITS 2
Nucleotide sites			
Length variation (range in bp)	230–242	212–216	444–456
No. aligned	245	222	467
No. constant (%)	150 (61.2)	146 (65.8)	296 (63.4)
No. potentially parsimony informative (%)	53 (21.6)	46 (20.7)	99 (21.2)
No. autapomorphic (%)	42 (17.1)	30 (13.5)	72 (15.4)
Length variation			
No. of unambiguous alignment gaps	10	10	20
Size of gaps (range in bp)	1–7	1–3	1–7
No. of gaps potentially parsimony informative	7	4	11
Sequence divergence (range, %)			
Medicago accessions only	0–8.1	0–7.0	0–6.9
All accessions	0–14.0	0–12.0	0–12.4
G + C content (range, %)	46–51	45–49	46–49

99 (21.2%) were potentially informative for parsimony analysis, and 72 (15.4%) were autapomorphic. In direct pairwise comparisons across all *Medicago* accessions, sequence divergence values ranged from identity to 8.1% of nucleotides in ITS 1 and from identity to 7.0% of nucleotides in ITS 2. Comparisons of sequence pairs across both spacers gave divergence values ranging between identity and 6.9%. The four examined accessions of *Medicago sativa* possessed identical ITS sequences, as did *M. monantha* and *M. orthoceras, M. rigidula* and *M. rigiduloides*, and *M. blancheana* and *M. rotata*. Nucleotide sequence divergence between the two *Trifolium* species was 2.5%; divergence values between *Trifolium* and *Medicago* across both spacer regions ranged between 9.2 and 12.4%. Twenty gaps were introduced into the alignment: 11 were 1 bp in length, six were 2 bp in length, one was 3 bp in length, one was 4 bp in length, and one was 7 bp in length. Eleven gaps were potentially informative for parsimony analysis; five of these distinguished the two species of *Trifolium* from *Medicago*. Although ITS 1 is longer than ITS 2, both spacers contributed comparable numbers of informative nucleotide substitutions and gaps to the phylogenetic analysis. The ratio of terminal taxa (62) to potentially parsimony informative characters across both spacers (99) was 1:1.6. Relative to the number of taxa examined, the number of informative characters is small.

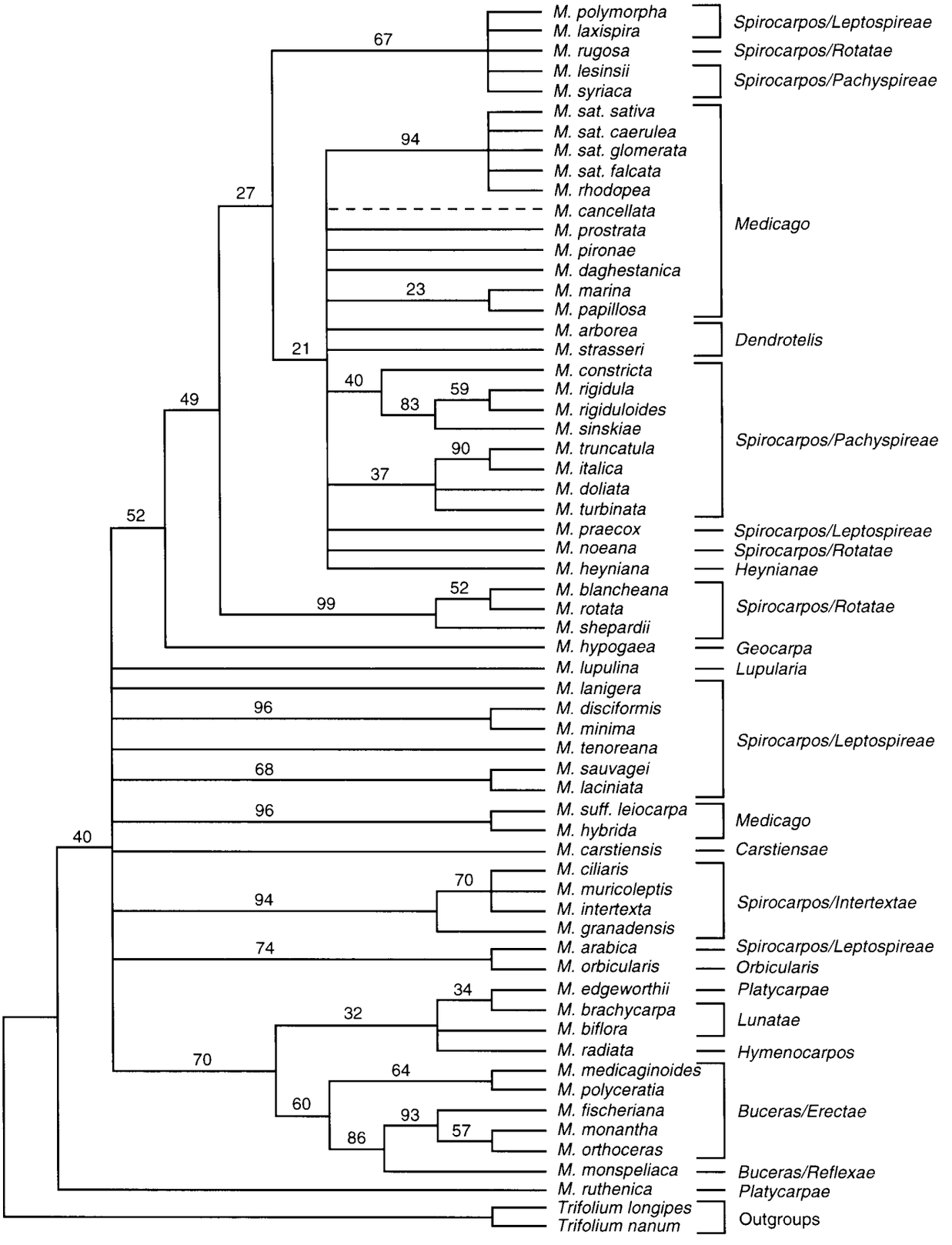
ITS phylogenetic analyses

MP analysis of all 62 ITS 1 and ITS 2 DNA sequences resulted in at least 5000 minimal length trees; the strict consensus of these trees, with accompanying bootstrap values, is presented in Fig. 2. Sectional and subsectional designations of *Medicago*, following Small and Jomphe (1989), are also indicated. These trees have a length of 301 steps, consistency indices (CIs) of 0.661 and 0.547, with and without uninformative characters, respectively, and a retention index (RI) of 0.788. Bootstrap values ranged between 21 and 99%. The average number of steps per character, calculated over all 467 aligned positions and 385 of the 5000 MP topologies saved, was 0.6. Fifteen characters required a minimum of four to seven evolutionary changes over this subset of trees. When partial ITS data for *M. isthmocarpa* (*Buceras* subsection *Isthmocarpae*) was added to the 62-taxon data matrix and the MP analysis rerun, this species fell as sister to the clade of *M. fischeriana, M. monantha,* and *M. orthoceras*, all members of section *Buceras* (not shown).

Optimization of the 20 inferred alignment gaps onto one arbitrarily selected 301-step tree (open and solid circles in Fig. 3) revealed that the distribution of insertion and deletion events (indels) is highly consistent with the inferred phylogeny. Of the two indels distinguishing the ITS sequences of *Medicago sativa*, only one of these occurs in *M. rhodopea*. Reanalyzing the nucleotide data with the inclusion of the 20 inferred alignment gaps coded as binary characters yielded at least 5000 MP trees. These trees each had a length of 321 steps, CIs of 0.682 and 0.568 (with and without uninformative characters, respectively) and a RI of 0.795. With the exception of *M. rhodopea* arising as sister to the *M. sativa* clade, the strict consensus of 5000 of these trees had an identical topology to that seen in Fig. 2.

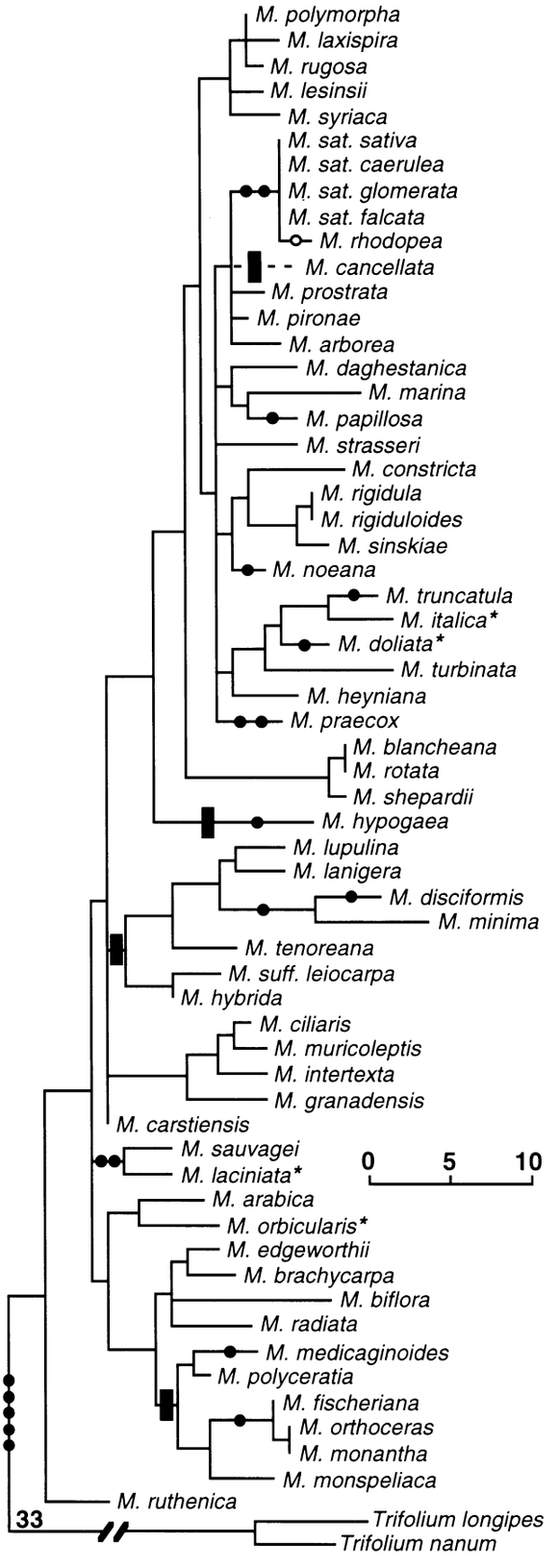
When *Melilotus officinalis* and *M. alba* replaced the two *Trifolium* species as outgroups and the MP analysis rerun, minor changes in tree topology occurred. These changes involved the replacement of *M. ruthenica* at the base of *Medicago* (Fig. 2) with the 10-taxon clade that includes *M. edgeworthii*

© 1998 NRC Canada



© 1998 NRC Canada

Fig. 3. One of five thousand 301-step trees of *Medicago* derived from equally weighted parsimony analysis of nuclear ribosomal DNA ITS 1 and ITS 2 sequences (CIs with and without uninformative characters = 0.661 and 0.547, respectively; RI = 0.788). Lengths of branches are proportional to the number of inferred nucleotide substitutions occurring along them (note scale bar and truncated branch of 33 steps leading to *Trifolium*). The distribution of 21 indels inferred from the 20 alignment gaps is optimized on this tree; solid circles represent uniquely occurring indels, and the open circle represents a reversal. The distribution of *rpo*C1 intron loss (indicated by the solid rectangles) is also optimized on this tree under the assumptions that the presence of the intron is the plesiomorphic condition and that intron loss is irreversible. Based on this tree, the intron in chloroplast gene *rpo*C1 has been lost independently a minimum of four times during the evolution of these taxa. When the putative hybrid *M. cancellata* is removed from the analysis (indicated by the broken line), three such intron losses are inferred. Asterisks show those four taxa whose *rpo*C1 intron presence could not be determined.



and *M. monspeliaca*. This 10-taxon clade formed a trichotomy, together with a clade comprising *M. arabica* and *M. orbicularis* and another consisting of all other *Medicago* species. *Medicago ruthenica* arose in a large polytomy one branch up the tree. These changes due to outgroup selection were limited to the more ancestral nodes of the strict consensus tree, areas where resolution and bootstrap support are low.

The best maximum likelihood tree, calculated with a transition/transversion rate ratio of 2.0 and rooted with *Trifolium*, had a log likelihood of –2561.75 (Fig. 4). While there is much concordance of relationships between the trees derived from the MP and ML analyses, some major differences occurred and include (*i*) the relative placement of *M. ruthenica*, either at the base of the strict consensus tree or positioned intermediately within the ML tree, and (*ii*) the placement of the 10-taxon *M. edgeworthii* through *M. monspeliaca* clade as sister to all other examined *Medicago* accessions in the ML tree. The clade of *M. lupulina, M. lanigera, M. disciformis, M. mimima,* and *M. tenoreana* in the ML tree is also evident in a subset of the trees (e.g., Fig. 3) derived from the MP analysis; in the latter, this clade is sister to *M. suffruticosa* ssp. *leiocarpa* and *M. hybrida*. Similarly, the 10-taxon *M. edgeworthii – M. monspeliaca* clade appears in all phylogenetic trees constructed.

The inclusion of taxa of hybrid origin in phylogenetic analyses has potential to impact the topology of the trees, such as distorting the pattern of hypothesized relationships among related nonhybrid taxa (Hull 1979; Funk 1985; Cronquist 1987; McDade 1992, 1995). Of all the *Medicago* taxa included in our investigation, *M. cancellata* is the only one for which there is appreciable speculation of multispecific origin. *Medicago cancellata* ($2n = 48$) is considered to be an allopolyploid of *M. rupestris* ($2n = 16$) and *M. sativa* ($2n = 32$; Lesins and Lesins 1979; Small 1987); all three species are treated in section *Medicago*, are very similar morphologically, and are undoubtedly closely related (Small and Jomphe 1989). Material of *M. rupestris* was not included in our study. Despite its name, *M. hybrida* is diploid ($2n = 16$), and no serious speculation of its possible hybrid origin exists.

© 1998 NRC Canada

Fig. 4. Maximum likelihood tree of *Medicago* derived from nuclear ribosomal DNA ITS 1 and ITS 2 sequences using a transition/transversion rate ratio of 2.0. Branch lengths are proportional to the number of expected substitutions per site (note scale bar at lower right and truncated branch leading to the outgroup *Trifolium*). Branch lengths that are not significantly greater than zero are collapsed. This tree had a log likelihood value of –2561.75. Sectional and subsectional classification of *Medicago* is based on Small and Jomphe (1989); complete taxon names are presented in Table 1. The distribution of *rpo*C1 intron loss is indicated by the solid bars and is optimized on this tree using the criterion of Dollo parsimony. Based on this tree, the intron in chloroplast gene *rpo*C1 is inferred to have been lost independently a minimum of five times during the evolution of these taxa. When the putative hybrid *M. cancellata* is removed (indicated by the broken line) and the analysis rerun (log likelihood value = –2535.41), the same tree results and four such intron losses are inferred. Asterisks show those four taxa whose *rpo*C1 intron presence could not be determined.

Because of its hybrid nature and potential to disrupt relationships, *M. cancellata* was excluded from a second round of phylogenetic analyses. The results, however, were identical to those inferred from the first set of analyses but, of course, with the exclusion of this taxon (see Figs. 2–4). The MP analysis resulted in at least 5000 minimal length trees (tree length = 297 steps excluding indels; CIs with and without uninformative characters = 0.670 and 0.553, respectively; RI = 0.794). The best maximum likelihood tree had a log likelihood value of –2535.41. Based on these analyses of ITS data, and in the absence of *M. rupestris*, the inclusion of *M. cancellata* in our initial analyses appears to have had no effect on the resultant phylogenies.

Distribution of *rpo*C1 intron loss

Of the 17 species of *Medicago* lacking the plastid *rpo*C1 intron (Table 1), complete ITS sequence data are available for 15 taxa. Partial ITS data are available for *M. isthmocarpa*, and this species allies with *Medicago* section *Buceras* based on parsimony analysis of partial ITS sequence data. *Medicago coronata* was not sequenced; morphologically it is similar to *M. praecox* (Small and Jomphe 1989). The number of intron losses across all 5000 MP 301-step trees, calculated using MACCLADE and optimized using Dollo parsimony, ranged between 4 and 6 (the same range was obtained when *Melilotus* was used to root the trees). The evolutionary pattern of intron loss, as suggested by one of these 5000 MP trees, is illustrated in Fig. 3 (solid rectangles). Here four such losses are apparent, assuming that loss of a chloroplast intron is irreversible. However, if the putative hybrid *M. cancellata* is removed from these trees, only three intron losses are inferred. Mapping the distribution of intron loss parsimoniously onto the ML tree (Fig. 4) revealed five instances of intron loss (four if *M. cancellata* is excluded). Constraining the topology so that the 15 intronless taxa form a monophyletic group results in trees 11 steps longer than those most parsimonious (tree length = 312 steps; CIs with and without uninformative characters = 0.638 and 0.521, respectively; RI = 0.765). When *M. cancellata* is excluded, the remaining 14 intronless taxa form a clade in trees six steps longer than those most parsimonious (tree length = 303 steps; CIs with and without uninformative characters = 0.657 and 0.538, respectively; RI = 0.781). Because a completed heuristic search of the ITS data using parsimony was not possible, information on the phylogenetic placement of the intronless *M. coronata* unavailable, and not all species within the genus included in this investigation, a more precise estimate of the number of independent losses of the *rpo*C1 intron in the genus cannot be made. Nevertheless, this intron loss character is clearly homoplastic within *Medicago*, with a minimum of three independent losses inferred for the group.

Based on the relationships suggested by the phylogenetic analyses of ITS data, the following clades are characterized by absence of the *rpo*C1 intron: (*i*) *M. suffruticosa* ssp. *leiocarpa* and *M. hybrida*; (*ii*) *M. lupulina*, *M. lanigera*, *M. disciformis*, *M. minima*, and *M. tenoreana* in the ML and a subset of the MP trees; and (*iii*) *M. medicaginoides*, *M. polyceratia*, *M. fischeriana*, *M. monantha*, *M. orthoceras*, *M. monspeliaca*, and *M. isthmocarpa* (the latter based on partial ITS sequence data). In addition, both *M. cancellata* and *M. hypogaea* are also characterized by intron absence.

Phylogenetic resolutions

Small and Jomphe (1989) recognized 12 sections and eight subsections of *Medicago*, of which all but one subsection (*Buceras* subsection *Deflexae*) was not represented in our study. Five sections (*Carstiensae*, *Geocarpa*, *Heynianae*, *Orbicularis*, and *Hymenocarpos*) and one subsection (*Buceras* subsection *Reflexae*) are monotypic. Of the remaining six sections and five subsections where more than one species was examined for ITS sequence variation, only section *Buceras* (seven species including *M. isthmocarpa*) and *Spirocarpos* subsection *Intertextae* (four species) are monophyletic. Monophyly of section *Buceras* is supported by the shared absence of the *rpo*C1 intron in all examined members. In contrast, section *Spirocarpos* (and, in particular, its subsections *Leptospireae* and *Rotatae*) is highly unnatural with many independent derivations in the cladograms. Sections *Medicago* and *Platycarpae* are also clearly polyphyletic.

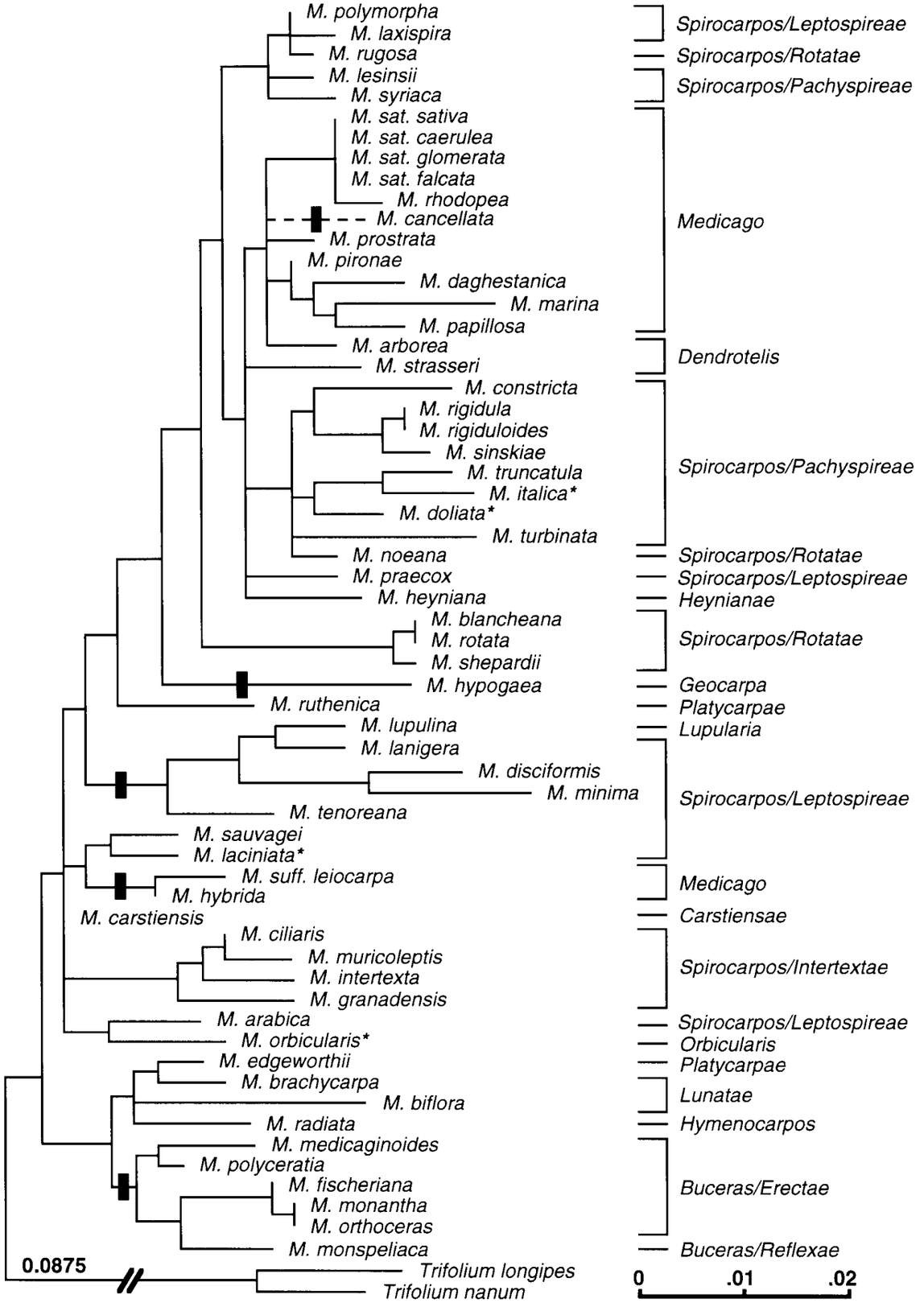
In addition to *Medicago* section *Buceras* and *Spirocarpos* subsection *Intertextae*, several additional major clades are recognized in all phylogenetic analyses. As examples, two of these include the group of *M. edgeworthii*, *M. brachycarpa*, *M. biflora*, and *M. radiata* and the group of *M. polymorpha*, *M. laxispira*, *M. rugosa*, *M. lesinsii*, and *M. syriaca*. Many of these major clades have yet to be recognized in any previous taxonomic study. However, the lack of resolution seen in several areas of the MP and ML trees, the weak to moderate bootstrap support prevalent for many of the clades in the former, and the topological differences among MP trees, indicate that we proceed with caution in suggesting relationships within *Medicago* based on these ITS sequence and *rpo*C1 intron distributional data.

Discussion

Taxonomic distribution of the *rpo*C1 intron

We have surveyed for the distribution of the intron in chloroplast gene *rpo*C1 in 65 of the 86 described species of *Medicago* (Small and Jomphe 1989; Small 1990*a*, 1990*b*; Small and Brookes 1991). Our results indicate that this intron, having

© 1998 NRC Canada



© 1998 NRC Canada

been precisely excised from the gene in each of the three species investigated, is absent from 17 species in 5 of the 12 sections recognized in the genus (Small and Jomphe 1989). The intron is absent from section *Geocarpa* and all examined members of section *Buceras*. Sections *Medicago*, *Spirocarpos*, and *Lupularia* are heterogeneous with regard to intron content, whereas the intron is present in all examined representatives of sections *Dendrotelis*, *Carstiensae*, *Heynianae*, *Hymenocarpus*, *Platycarpae*, and *Lunatae*. The intron's presence in section *Orbicularis* could not be determined.

Phylogenies derived from nuclear ribosomal DNA ITS sequences, estimated using MP and ML methods, reveal that many of the sections and subsections of *Medicago* are not monophyletic as currently circumscribed and are, in some cases, grossly polyphyletic. These trees, therefore, provide an important historical framework for estimating the number of evolutionary losses of the intron in chloroplast gene *rpo*C1. The phylogenies suggest that the intron has been lost independently a minimum of three times and, perhaps, as many as six times during the evolution of these plants, assuming that intron regains are impossible. While this assumption may be currently sound for angiosperm chloroplast genomes, the recent origin of introns has been suggested for the plastid genome of the algae *Euglena* (Thompson et al. 1995). As a consequence, regains of the *rpo*C1 intron in *Medicago* may be a possibility.

Given the relative rarity of documented intron loss in angiosperm cpDNAs at high taxonomic levels (Downie et al. 1991, 1996; Downie and Palmer 1992), the frequent loss of the *rpo*C1 intron in *Medicago* is indeed a surprise. Polymorphism for the presence of the *rpl*2 intron in *Bauhinia* (also Fabaceae), including three instances of intraspecific polymorphism, has also been reported (Lai et al. 1997). Evidently, in these two genera at least, intron presence–absence is a very labile character. Whether this pattern of frequent intron loss is unique to these taxa (because of some yet to be discovered molecular process) or extends to other legume genera as well remains to be seen.

In *Medicago*, the following groups of taxa are distinguished by absence of the *rpo*C1 intron: (*i*) *M. hypogaea*; (*ii*) Section *Buceras*; (*iii*) the clade of *M. suffruticosa* ssp. *leiocarpa* and *M. hybrida*; (*iv*) the clade of *M. lupulina*, *M. lanigera*, *M. disciformis*, *M. minima*, and *M. tenoreana*; and (*v*) *M. cancellata.* Two of these groups (i.e., the clades containing *M. suffruticosa* and *M. lupulina*) ally in a subset of the MP trees (Fig. 3). In other subsets of these MP trees, the *M. lupulina* clade comprises several, separate lineages.

Medicago hypogaea is unusual within the genus in producing its fruits underground. Because its affinities to other *Medicago* are obscure, it was placed in its own section (Small and Brookes 1984; Small and Jomphe 1989). The molecular data presented herein, however, suggest that it may be sister to a large, 32-taxon clade that includes *M. polymorpha, M. sativa*, and *M. shepardii* (Figs. 2–4).

All examined members of *Medicago* section *Buceras* (i.e., *M. medicaginoides, M. polyceratia, M. fischeriana, M. monantha, M. orthoceras, M. monspeliaca,* and *M. isthmocarpa*) lack the *rpo*C1 intron and are monophyletic based on the ITS sequencing results. The fruits of *Buceras* also suggest monophyly, unusual in the genus by being linear and mostly subterete. The presence of this section, along with the clade of *M. edgeworthii, M. brachycarpa, M. biflora,* and *M. radiata,* at the base of the ML tree supports, in part, Small's (1987, 1989) tentative phylogeny for the group, where it was suggested that sections *Buceras*, *Lunatae*, and *Platycarpae* represent basally branching lineages.

Medicago suffruticosa ssp. *leiocarpa* and *M. hybrida* lack the intron and are sister taxa in all phylogenetic analyses. While both Small and Jomphe (1989) and Lesins and Lesins (1979) considered these species in section *Medicago* (or subgenus *Medicago*), Lesins and Lesins placed them together in their own subsection, *Suffruticosae*. While the ITS results presented herein support a close relationship between *M. suffruticosa* and *M. hybrida*, it also shows that these two species are distantly related to other members of section *Medicago*. This isolation of *M. suffruticosa* and *M. hybrida* from the other members of this section is also supported by chemical data (Classen et al. 1982).

Medicago lupulina, *M. lanigera*, *M. disciformis*, *M. minima*, and *M. tenoreana* comprise a clade in the ML tree (Fig. 4) and in a subset of the MP trees (Fig. 3). In the MP strict consensus tree, however, their relationships are not resolved (Fig. 2). With the exception of *M. lupulina*, which has been treated alongside the *rpo*C1 intron-containing *M. secundiflora* in section *Lupularia*, all remaining taxa have been placed in *Spirocarpos* subsection *Leptospireae* (Small and Jomphe 1989).

Medicago cancellata, a putative allopolyploid between *M. sativa* and *M. rupestris* (Lesins and Lesins 1979; Small 1987), is also characterized by absence of the *rpo*C1 intron. The four accessions of *M. sativa* examined all possess the intron, whereas material of *M. rupestris* was unavailable for analysis. It would definitely be valuable to ascertain the presence of the intron in the latter.

Alternative scenarios

One alternative explanation to account for the distribution of *rpo*C1 intron loss in *Medicago*, a scenario proposed for the variable presence of the *rpl*2 intron in *Bauhinia* (Lai et al. 1997), is ancestral polymorphism for the loss and subsequent lineage sorting. Plastid polymorphism, relatively rare in angiosperm cpDNAs, is known from *Medicago* (Johnson and Palmer 1989). Moreover, species of *Medicago* have been reported to exhibit biparental inheritance, a mechanism leading to cpDNA heterogeneity (Smith et al. 1986; Johnson and Palmer 1989). In this lineage sorting hypothesis, only a single loss of the intron is required. With the exception of the branches leading to *M. cancellata*, all potentially polymorphic branches are found near the base of the trees; thus, polymorphism is not required to persist beyond a few speciation events. The absence of the intron, however, in *M. cancellata* (and *M. coronata*, not examined for ITS sequence variation but putatively allied to *M. praecox*; Small and Jomphe 1989) would require that the polymorphism be retained over what is likely to have been a long evolutionary period. Nevertheless, where plastid polymorphism occurs, the potential for lineage sorting exists.

Another explanation is that the ITS trees (and traditional taxonomy) are unreliable indicators of phylogenetic history and that *rpo*C1 intron loss should be considered an unambiguous marker of monophyly within the genus. It may be possible that lineage sorting of ancestral ITS polymorphisms has led to erroneous species tree inference from the ITS data. As gene

© 1998 NRC Canada

sequence data from the plastid genomes of these plants become available, this hypothesis can be tested.

Phylogenetic utility of intron loss characters

Surveys of angiosperm chloroplast genomes for structural rearrangements have shown that intron loss characters may often be homoplastic. Downie et al. (1991) first reported the loss of the *rpl*2 intron from a minimum of six different dicot lineages, whereas Doyle et al. (1995) hypothesized four additional losses of the *rpl*2 intron in Fabaceae. Multiple losses of the *rpl*2 intron in *Bauhinia*, including intron presence polymorphism for three species, have also been described (Lai et al. 1997). The intron in chloroplast gene *rpo*C1 has evidently been lost multiple times in the angiosperms, in such diverse families as Goodeniaceae, Passifloraceae, Cactaceae, Poaceae, Aizoaceae, and Fabaceae (Downie et al. 1996). Other homoplastic intron losses include those in the chloroplast genes *rpl*16 and *rps*12 (Downie and Palmer 1992; Hoot and Palmer 1994; Freyer et al. 1995; Campagna and Downie 1998). In this study (and that of Lai et al. 1997), it has been demonstrated that intron loss can occur repeatedly within a genus, severely reducing the utility of this character as an unambiguous phylogenetic marker at least at this level of phylogenetic analysis.

Conclusions

The primary objective of this study was to survey for the occurrence of the plastid *rpo*C1 intron in *Medicago*. To ascertain the historical pattern of intron loss, and in the absence of any existing rigorously constructed phylogenetic estimate for the group, a comparative analysis of nuclear ribosomal DNA ITS sequences was simultaneously undertaken. Our main goal was not to carry out a thorough molecular systematic revision of the group but to ascertain the utility of *rpo*C1 intron loss as a phylogenetic marker. On these bases, it is inferred that the *rpo*C1 intron has been lost independently a minimum of three times during the evolution of these plants. While each of these intron loss characters has potential to demarcate clades, the availability of an independent data set is paramount to ascertain the extent of homoplasy of this intron loss character. Because of the conservative nature of ITS sequence evolution in *Medicago*, further resolution of relationships and intron loss will require the incorporation of additional molecular and nonmolecular data. These studies are currently in progress (K. Steele, unpublished data). Once a robust phylogenetic hypothesis is at hand, trends in the evolution of morphological, cytological, ecological, and phytochemical characters, as well as a more precise estimate of the number of independent losses of the *rpo*C1 intron in the genus, can be assessed.

Acknowledgments

The authors thank U.S. Department of Agriculture, Agricultural Research Service, Western Regional Plant Introduction Station, Pullman, Wash., for seed material; E. Llanas for laboratory assistance; and J. Shore and two anonymous reviewers for very constructive criticism of the manuscript. This work was funded, in part, by grants from the National Science Foundation (DEB-9407712), the Campus Research Board of the University of Illinois, and the Howard Hughes Medical Institute's Undergraduate Research Program.

References

Bailey, C.D., Doyle, J.D., Kajita, T., Nemoto, T., and Ohashi, H. 1997. The chloroplast *rpl*2 intron and ORF184 as phylogenetic markers in the legume tribe Desmodieae. Syst. Bot. **22**: 133–138.

Baldwin, B.G., Sanderson, M.J., Porter, J.M., Wojciechowski, M.F., Campbell, C.S., and Donoghue, M.J. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence of angiosperm phylogeny. Ann. Mo. Bot. Gard. **82**: 247–277.

Campagna, M.L., and Downie, S.R. 1998. The intron in chloroplast gene *rpl*16 is missing from the flowering plant families Geraniaceae, Goodeniaceae, and Plumbaginaceae. Trans. Ill. State Acad. Sci. **91**: 1–11.

Classen, D., Nozzolillo, C., and Small, E. 1982. A phenolic–taxometric study of *Medicago* (Leguminosae). Can. J. Bot. **60**: 2477–2495.

Cronquist, A. 1987. A botanical critique of cladism. Bot. Rev. **53**: 1–52.

Downie, S.R., and Palmer, J.D. 1992. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. *In* Molecular systematics of plants. *Edited by* P.S. Soltis, D.E. Soltis, and J.J. Doyle. Chapman & Hall, New York. pp. 14–35.

Downie, S.R., and Katz-Downie, D.S. 1996. A molecular phylogeny of Apiaceae subfamily Apioideae: evidence from nuclear ribosomal DNA internal transcribed spacer sequences. Am. J. Bot. **83**: 234–251.

Downie, S.R., Olmstead, R.G., Zurawski, G., Soltis, D.E., Soltis, P.S., Watson, J.C., and Palmer, J.D. 1991. Six independent losses of the chloroplast DNA *rpl*2 intron in dicotyledons: molecular and phylogenetic implications. Evolution, **45**: 1245–1259.

Downie, S.R., Llanas, E., and Katz-Downie, D.S. 1996. Multiple independent losses of the *rpo*C1 intron in angiosperm chloroplast DNAs. Syst. Bot. **21**: 135–151.

Doyle, J.J., and Doyle, J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. **19**: 11–15.

Doyle, J.J., Doyle, J.L., and Palmer, J.D. 1995. Multiple independent losses of two genes and one intron from legume chloroplast genomes. Syst. Bot. **20**: 272–294.

Farris, J.S. 1977. Phylogenetic analysis under Dollo's Law. Syst. Zool. **26**: 77–88.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution, **39**: 783–791.

Freyer, R., Neckermann, K., Maier, R.A., and Kössel, H. 1995. Structural and functional analysis of plastid genomes from parasitic plants: loss of an intron within the genus *Cuscuta*. Curr. Genet. **27**: 580–586.

Funk, V.A. 1985. Phylogenetic patterns and hybridization. Ann. Mo. Bot. Gard. **72**: 681–715.

Holmgren, P.K., Holmgren, N.H., and Barnett, L.C. 1990. Index herbariorum. New York Botanical Garden, New York.

Hoot, S.B., and Palmer, J.D. 1994. Structural rearrangements, including parallel inversions, within the chloroplast genome of *Anemone* and related genera. J. Mol. Evol. **38**: 274–281.

Hull, D.L. 1979. The limits of cladism. Syst. Zool. **28**: 416–440.

Johnson, L.B., and Palmer, J.D. 1989. Heteroplasmy of chloroplast DNA in *Medicago*. Plant Mol. Biol. **12**: 3–11.

Lai, M., Sceppa, J., Ballenger, J.A., Doyle, J.L., Doyle, J.J., and Wunderlin, R. 1997. Polymorphism for the presence of the *rpl*2 intron in chloroplast genomes of *Bauhinia* (Leguminosae). Syst. Bot. **22**: 519–528.

Le Quesne, W. 1974. The uniquely evolved character concept and its cladistic application. Syst. Zool. **23**: 513–517.

Lesins, K.A., and Lesins, I. 1979. Genus *Medicago* (Leguminosae). A taxogenetic study. Dr. W. Junk, Publishers, The Hague, The Netherlands.

Maddison, W.P., and Maddison, D.R. 1992. MACCLADE: analysis

© 1998 NRC Canada

of phylogeny and character evolution, version 3.0 edition. Sinauer Associates, Sunderland, Mass.

McDade, L.A. 1992. Hybrids and phylogenetic systematics II. The impact of hybrids on cladistic analysis. Evolution, **46**: 1329–1346.

McDade, L.A. 1995. Hybridization and phylogenetics. *In* Experimental and molecular approaches to plant biosystematics. *Edited by* P.C. Hoch and A.G. Stephenson. Monogr. Syst. Bot. Mo. Bot. Gard. No. 53. pp. 305–331.

Olsen, G. J., Matsuda, H., Hagstrom, R., and Overbeek, R. 1994. fastDNAml: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. Comput. Appl. Biosci. **10**: 41–48.

Palmer, J.D. 1991. Plastid chromosomes: structure and evolution. *In* Cell culture and somatic cell genetics of plants. Vol. 7A. The molecular biology of plastids. *Edited by* L. Bogorad and I.K. Vasil. Academic Press, San Diego, Calif. pp. 5–53.

Sanderson, M.J., and Wojciechowski, M.F. 1996. Diversification rates in a temperate legume clade: are there "so many species" of *Astragalus* (Fabaceae)? Am. J. Bot. **83**: 1488–1502.

Small, E. 1987. Generic changes in Trifolieae subtribe Trigonellinae. *In* Advances in legume systematics, Part 3. *Edited by* C.H. Stirton. Royal Botanic Gardens, Kew, England. pp. 169–181.

Small, E. 1989. The evolution of genera in the Leguminosae. *In* Advances in legume biology. *Edited by* C.H. Stirton, and J.L. Zarucchi. Monogr. Syst. Bot. Mo. Bot. Gard. No. 29. pp. 467–486.

Small, E. 1990*a*. *Medicago syriaca*, a new species. Can. J. Bot. **68**: 1473–1478.

Small, E. 1990*b*. *Medicago rigiduloides*, a new species segregated from *M. rigidula*. Can. J. Bot. **68**: 2614–2617.

Small, E., and Brookes, B.S. 1984. Reduction of the geocarpic *Factorovskya* to *Medicago*. Taxon, **33**: 622–635.

Small, E., and Brookes, B. 1991. A clarification of *Medicago sinskiae*. Can. J. Bot. **69**: 100–106.

Small, E, and Jomphe, M. 1989. A synopsis of the genus *Medicago* (Leguminosae). Can. J. Bot. **67**: 3260–3294.

Smith, S.E., Bingham, E.T., and Fulton, R.W. 1986. Transmission of chlorophyll deficiencies in *Medicago sativa*. Evidence for biparental inheritance of plastids. J. Hered. **77**: 35–38.

Swofford, D.L. 1993. PAUP: phylogenetic analysis using parsimony, version 3.1. Computer program distributed by the Illinois Natural History Survey, Champaign, Ill.

Thompson, M.D., Copertino, D.W., Thompson, E., Favreau, M.R., and Hallick, R.B. 1995. Evidence for the late origin of introns in chloroplast genes from an evolutionary analysis of the genus *Euglena*. Nucleic Acids Res. **23**: 4745–4752.

White, T.J., Bruns, T., Lee, S., and Taylor, J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *In* PCR protocols: a guide to methods and applications. *Edited by* M.A. Innis, D.H., Gelfand, J.J. Sninsky, and T.J. White. Academic Press, San Diego, Calif. pp. 315–322.

© 1998 NRC Canada