



9-2008

Complete Plastid Genome Sequence of the Chickpea (*Cicer Arietinum*) and the Phylogenetic Distribution of *rps12* and *clpP* Intron Losses Among Legumes (Leguminosae)

Robert K. Jansen

Martin F. Wojciechowski

Elumalai Sanniyasi

Seung-Bum Lee

Henry Daniell
University of Pennsylvania

Follow this and additional works at: https://repository.upenn.edu/dental_papers

 Part of the [Dentistry Commons](#)

Recommended Citation

Jansen, R. K., Wojciechowski, M. F., Sanniyasi, E., Lee, S., & Daniell, H. (2008). Complete Plastid Genome Sequence of the Chickpea (*Cicer Arietinum*) and the Phylogenetic Distribution of *rps12* and *clpP* Intron Losses Among Legumes (Leguminosae). *Molecular Phylogenetics and Evolution*, 48 (3), 1204-1217.
<http://dx.doi.org/10.1016/j.ympev.2008.06.013>

At the time of publication, author Henry Daniell was affiliated with the University of Central Florida. Currently, he is a faculty member at the School of Dental Medicine at the University of Pennsylvania.

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/dental_papers/277
For more information, please contact repository@pobox.upenn.edu.

Complete Plastid Genome Sequence of the Chickpea (*Cicer Arietinum*) and the Phylogenetic Distribution of *rps12* and *clpP* Intron Losses Among Legumes (Leguminosae)

Abstract

Chickpea (*Cicer arietinum*, Leguminosae), an important grain legume, is widely used for food and fodder throughout the world. We sequenced the complete plastid genome of chickpea, which is 125,319 bp in size, and contains only one copy of the inverted repeat (IR). The genome encodes 108 genes, including 4 rRNAs, 29 tRNAs, and 75 proteins. The genes *rps16*, *infA*, and *ycf4* are absent in the chickpea plastid genome, and *ndhB* has an internal stop codon in the 5'exon, similar to other legumes. Two genes have lost their introns, one in the 3'exon of the transpliced gene *rps12*, and the one between exons 1 and 2 of *clpP*; this represents the first documented case of the loss of introns from both of these genes in the same plastid genome. An extensive phylogenetic survey of these intron losses was performed on 302 taxa across legumes and the related family Polygalaceae. The *clpP* intron has been lost exclusively in taxa from the temperate "IR-lacking clade" (IRLC), whereas the *rps12* intron has been lost in most members of the IRLC (with the exception of *Wisteria*, *Callerya*, *Afgekia*, and certain species of *Millettia*, which represent the earliest diverging lineages of this clade), and in the tribe Desmodieae, which is closely related to the tribes Phaseoleae and Psoraleae. Data provided here suggest that the loss of the *rps12* intron occurred after the loss of the IR. The two new genomic changes identified in the present study provide additional support of the monophyly of the IR-loss clade, and resolution of the pattern of the earliest-branching lineages in this clade. The availability of the complete chickpea plastid genome sequence also provides valuable information on intergenic spacer regions among legumes and endogenous regulatory sequences for plastid genetic engineering.

Keywords

Plastid genetic engineering, Genome evolution, Phylogeny of legumes, Leguminosae, Intron loss, Chickpea, *Cicer*

Disciplines

Dentistry

Comments

At the time of publication, author Henry Daniell was affiliated with the University of Central Florida. Currently, he is a faculty member at the School of Dental Medicine at the University of Pennsylvania.



Published in final edited form as:

Mol Phylogenet Evol. 2008 September ; 48(3): 1204–1217. doi:10.1016/j.ympev.2008.06.013.

Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of *rps12* and *clpP* intron losses among legumes (Leguminosae)

Robert K. Jansen^a, Martin F. Wojciechowski^b, Elumalai Sanniyasi^c, Seung-Bum Lee^c, and Henry Daniell^{c,*}

^a Section of Integrative Biology and Institute of Cellular and Molecular Biology, Biological Laboratories 404, University of Texas, Austin, TX 78712, USA

^b School of Life Sciences, Genomics, Evolution and Bioinformatics Group, Arizona State University, Tempe, AZ 85287-4501, USA

^c University of Central Florida, College of Medicine, Department of Molecular Biology & Microbiology, Biomolecular Science, 4000 Central Florida Boulevard, Building #20, Room 336, Orlando, FL 32816-2364, USA

Abstract

Chickpea (*Cicer arietinum*, Leguminosae), an important grain legume, is widely used for food and fodder throughout the world. We sequenced the complete plastid genome of chickpea, which is 125,319 bp in size, and contains only one copy of the inverted repeat (IR). The genome encodes 108 genes, including 4 rRNAs, 29 tRNAs, and 75 proteins. The genes *rps16*, *infA*, and *ycf4* are absent in the chickpea plastid genome, and *ndhB* has an internal stop codon in the 5'exon, similar to other legumes. Two genes have lost their introns, one in the 3'exon of the transpliced gene *rps12*, and the one between exons 1 and 2 of *clpP*; this represents the first documented case of the loss of introns from both of these genes in the same plastid genome. An extensive phylogenetic survey of these intron losses was performed on 302 taxa across legumes and the related family Polygalaceae. The *clpP* intron has been lost exclusively in taxa from the temperate "IR-lacking clade" (IRLC), whereas the *rps12* intron has been lost in most members of the IRLC (with the exception of *Wisteria*, *Callerya*, *Afgekia*, and certain species of *Millettia*, which represent the earliest diverging lineages of this clade), and in the tribe Desmodieae, which is closely related to the tribes Phaseoleae and Psoraleae. Data provided here suggest that the loss of the *rps12* intron occurred after the loss of the IR. The two new genomic changes identified in the present study provide additional support of the monophyly of the IR-loss clade, and resolution of the pattern of the earliest-branching lineages in this clade. The availability of the complete chickpea plastid genome sequence also provides valuable information on intergenic spacer regions among legumes and endogenous regulatory sequences for plastid genetic engineering.

Keywords

Plastid genetic engineering; Genome evolution; Phylogeny of legumes; Leguminosae; Intron loss; Chickpea; *Cicer*

* Corresponding author. Fax: +1 407 823 0956. E-mail address: daniell@mail.ucf.edu (H. Daniell).

1. Introduction

Gene mapping and genomic sequencing have demonstrated that plastid genome organization is generally highly conserved among angiosperms (Palmer, 1991; Raubeson and Jansen, 2005). Most genomes are characterized by a quadripartite structure, with two copies of an inverted repeat (IR) separating the large (LSC) and small (SSC) single copy regions. The genomes usually include 120–130 genes and range in size from 120 to 170 kilobases (kb). Gene content and gene order are conserved throughout angiosperms with the ancestral configuration depicted by the earliest-branching angiosperm clades *Amborella* and Nymphaeales (Goremykin et al., 2003; Raubeson et al., 2007; Jansen et al., 2007). In a typical angiosperm chloroplast genome, there are 18 genes containing introns, six in tRNA genes and the remaining twelve in protein-coding genes. Fifteen of the intron-containing genes have only two exons, and the remaining three have three exons.

Changes in this highly conserved organization of plastid genomes have been utilized to resolve phylogenetic relationships among major clades in a number of angiosperm families, including Asteraceae (Jansen and Palmer, 1987; Kim et al., 2005), Berberidaceae (Kim and Jansen, 1995), Cactaceae (Wallace and Cota, 1996), Campanulaceae (Cosner et al., 2004), Leguminosae (Bruneau et al., 1990; Lavin et al., 1990; Doyle et al., 1995; Doyle et al., 1996), Geraniaceae (Chumley et al., 2006), Lobeliaceae (Knox et al., 1993), Oleaceae (Lee et al., 2007), Onagraceae (Hachtel et al., 1991; Greiner et al., 2008), Poaceae (Doyle et al., 1992), and Ranunculaceae (Hoot and Palmer, 1994; Johansson, 1999). The types of changes that have been used include inversions, loss of the 22–25 kb IR which contains a duplicated set of rRNA and tRNA genes, expansion/contraction of the IR, and gene/intron loss (Downie and Palmer, 1992). Although all of these genomic changes have exhibited some homoplasy, they have served as powerful phylogenetic markers for several reasons: (1) these types of changes are generally rare, resulting in lower homoplasy relative to nucleotide substitutions; (2) assessing homology of these events is generally straightforward; (3) the direction of evolutionary change is easily discerned; and (4) once a rearrangement is detected it is relatively easy to survey numerous taxa for each event. The relative phylogenetic utility of the different types of plastid genomic rearrangements varies considerably with inversions exhibiting the least amount of homoplasy. However, even genomic changes that are homoplasious provide valuable phylogenetic characters within each major lineage in which they occur. One example of this phenomenon concerns the *rpoCI* intron, which has been lost independently six times among angiosperms (Downie et al., 1996) but its absence is still a valuable phylogenetic marker for resolving relationships within each of these six lineages.

The Leguminosae (also Fabaceae; the legumes) is one angiosperm family that has experienced considerable numbers of plastid genomic rearrangements. Legumes are the third largest family of angiosperms with 730 genera and more than 19,000 species distributed throughout the world (Lewis et al., 2005). Legumes are second only to grasses in their agricultural and economic value, and include many important species grown for food, fodder, wood, ornamentals, and raw materials for industry and also for their ecologically important role in biological nitrogen fixation. A number of previous studies have examined the phylogenetic distribution of different plastid genome rearrangements among legumes, including the loss of one copy of the IR (Palmer and Thompson, 1982; Lavin et al., 1990), inversions of 50 kb (Palmer and Thompson, 1981; Doyle et al., 1996), and 78 kb (Bruneau et al., 1990), loss of the *rpl22* and *rps16* genes (Doyle et al., 1995), and loss of the *rpl2* intron (Doyle et al., 1995). These genomic rearrangements, combined with DNA sequence data (Doyle, 1995; Käss and Wink, 1995, 1996; Doyle et al., 1997, 2000; Kajita et al., 2001; Wojciechowski et al., 2004), have provided valuable phylogenetic data for resolving relationships among several deep nodes of legumes. The first, and probably most dramatic example of the phylogenetic utility of a plastid genomic rearrangement among legumes is the loss of one copy of the IR by all members sampled from

the tribes Carmichaelieae, Cicereae, Hedysareae, Trifolieae, Fabeae (Vicieae), Galegeae, and three genera of Millettieae (Lavin et al., 1990; Liston, 1995). The monophyly of this clade, known as the “IR-lacking clade” or IRLC (Wojciechowski et al., 2000), was later confirmed by phylogenetic analyses of DNA sequences of the plastid genes *rbcl* (Doyle et al., 1997; Käss and Wink, 1997) and *matK* (Wojciechowski et al., 2004), the plastid *trnL* intron (Pennington et al., 2001), and the internal transcribed spacer (ITS) regions of the nuclear ribosomal DNA (Hu et al., 2002).

In addition to utilization of plastid genome sequences for phylogenetic studies, they are very useful in engineering foreign genes. However, complete chloroplast genome sequence of only six species of crop plants were determined until 2004. Therefore, complete plastid genome sequences of several major crop species including fiber crops (Lee et al., 2006), tubers (Daniell et al., 2006, 2008), cereals (Saski et al., 2007), trees (Steane, 2005; Bausher et al., 2006; Ravi et al., 2006; Samson et al., 2007), vegetables (Ruhlman et al., 2006), fruits (Jansen et al., 2006; Daniell et al., 2006) and legumes (Saski et al., 2005; Guo et al., 2007) have been determined recently. Plastid genetic engineering offers a number of unique advantages including high level of transgene expression (DeCosa et al., 2001), multi-gene engineering in a single transformation event (Quesda-Vargas et al., 2005), transgene containment via maternal inheritance (Daniell, 2002; Daniell, 2007) or cytoplasmic male sterility (Ruiz and Daniell, 2005). Plastid transgenic lines also lack gene silencing (DeCosa et al., 2001; Lee et al., 2003), position effect due to site specific transgene integration (Daniell et al., 2002) and pleiotropic effects due to subcellular compartmentalization of transgene products (Lee et al., 2003; Daniell et al., 2001; Leelavathi et al., 2003); concerns of transgene silencing, position effect and pleiotropic effects are often encountered in nuclear genetic engineering. Therefore, transgenes have been integrated into plastid genomes to confer valuable agronomic traits, including herbicide resistance (Daniell et al., 1998), insect resistance (McBride et al., 1995; DeCosa et al., 2001), disease resistance (DeGray et al., 2001), drought tolerance (Lee et al., 2003), salt tolerance (Kumar et al., 2004), phytoremediation (Ruiz et al., 2003; Hussein et al., 2007) or expression of various therapeutic proteins or biomaterials (Verma and Daniell, 2007; Kamarajugadda and Daniell, 2006; Daniell et al., 2005). However, soybean is the only legume that has been transformed via the plastid genome so far (Dufoumantel et al., 2004, 2005) and more genome sequence information is needed to facilitate plastid genetic engineering in other economically important legumes.

During the past eight years plastid genome sequences have been completed for four legumes, including *Lotus japonicus* (Regel) K. Larson, *Medicago truncatula* Gaertn., *Glycine max* Merr., and *Phaseolus vulgaris* L. In this paper, we report on the complete genome sequence of *Cicer arietinum* L. (chickpea). Sequences of these five legume plastid genomes, all from taxa belonging to the subfamily Papilionoideae and two of which are from members of the IRLC (*Cicer*, *Medicago*), will enable more detailed comparisons of the organization and evolution of the plastid genomes of legumes. Our comparisons have identified two additional genomic rearrangements, the losses of introns in the *clpP* and *rps12* (3'-end) genes, and we survey the phylogenetic distribution of these changes in 302 taxa of legumes and the related family Polygalaceae.

2. Materials and methods

2.1. Plant material and plastid isolation

Chickpea (*C. arietinum* L.) seeds were obtained from IARI (Indian Agricultural Research Institute) New Delhi, India. Fresh leaves were harvested from greenhouse grown chickpea seedlings. Prior to plastid isolation, plants were kept in the dark for 48 h to reduce the levels of starch. Plastid isolation was performed as described by Jansen et al. (2005) and Samson et al. (2007).

Purified plastids were used to amplify the entire plastid genome by rolling circular amplification (RCA) using the Repli-g RCA-KIT (Qiagen GmbH, Hilden, Germany) following the methods described in Jansen et al. (2005). The success of genome amplification and the quality of the DNA was verified by digesting with restriction enzymes BstXI, EcoRI, and HindIII, and visualization of the resulting fragments on ethidium bromide stained, 1% agarose gels.

2.2. Plastid genome sequencing and assembly

Purified RCA products were subjected to nebulization, followed by end repair, and size-fractionated by agarose gel electrophoresis to obtain fragment lengths ranging from 2.0 to 3.5 kb. Repaired products were blunt-end cloned into the 4blunt-TOPO vector, followed by transformation into *Escherichia coli* ElectroMax TM-DH5 α cells by electroporation (TOPO shotgun Cloning Kit; Invitrogen, Carlsbad, CA, USA). Transformed cells were selected on Luria-Bertani (LB) agar containing 100 μ g/ml ampicillin and arrayed into 30 \times 96-well microtitre plates. Sequencing reactions were carried out in both the forward and reverse directions using the BigDye Terminator v3.1 Cycle Sequencing Kit and separated by a 3730 \times L DNA Sequence Analyzer (Applied Biosystems, Foster City, CA, USA). Sequence data were assembled using Sequencher version 4.5 (Gene Codes, Ann Arbor, MI, USA) following quality and vector trimming. Gap regions were filled by sequencing PCR fragments generated using primers that flank the gaps. The assembly was considered complete when a quality score of ≥ 20 was obtained at every base position with at least 6 \times coverage.

2.3. Genome annotation

The annotation program Dual Organellar Genome Annotator (DOGMA; Wyman et al., 2004) was used to annotate the *C. arietinum* plastid genome, after uploading a FASTA-formatted file of complete nucleotide sequence to the program's server. BLASTX and BLASTN searches, against a custom database of previously published plastid genomes, identified putative protein-coding genes, tRNAs, and rRNAs. For genes with low sequence identity, manual annotation was performed, after identifying the position of the start and stop codons, as well as the translated amino acid sequence, using the plastid/bacterial genetic code.

2.4. Whole genome sequence alignment

MultiPipMaker (Schwartz et al., 2003; <http://bio.cse.psu.edu>) was used for multiple genome alignment of chickpea with four published legume plastid genomes from the subfamily Papilionoideae; *Lotus japonicus* (NC_002694, Kato et al., 2000), *Medicago truncatula* (NC_003119), *Glycine max* (NC_007942, Sasaki et al., 2005), and *Phaseolus vulgaris* (NC_009259, Guo et al., 2007). We generated the alignments of whole genomes using chickpea as the reference genome.

2.5. Survey for loss of *clpP* and *rps12* introns

We surveyed for the presence/absence of two introns from 318 accessions of 301 legume species representing all 3 subfamilies and 198 genera, and 1 member of the related family Polygalaceae (Table 1) using primers designed that span the intron in each gene: *clpPF3* and *clpPR3* (5'-ATGCCMATTGGTGTTCCTCAAAGTRCC and 5'-GCGTGAGGGAATGCTAGACGTTTGGT) for the *clpP* gene, and *rps12F* and *rps12R* (5'-CCYAAAAACCAAACCTCTGCYTTACGTAAA and 5'-TTATTTTGGCTTTTTBGCMMCCATATT) for the *rps12* gene. PCR amplification products were resolved on 1% agarose gels and fragment sizes were determined by comparison to DNA size markers.

3. Results

3.1. Size, gene content, order, and organization of chickpea plastid genome

Chickpea has a circular plastid genome 125,319 bp in length with only one copy of the IR region (Fig. 1, GenBank accession number EU 835853). Gene order in chickpea is similar to the ancestral angiosperm gene order (Raubeson et al., 2007) except for the loss of one copy of the IR and by the presence of a single, large inversion of approximately 50 kb that reverses the order of the genes between *rbcL* and *rps16*. The same inversion is present in the four other completely sequenced legume plastid genomes *Glycine max* (Sasaki et al., 2005), *Lotus japonicus* (Kato et al., 2000), *Medicago truncatula*, and *Phaseolus vulgaris* (Guo et al., 2007), and is apparently shared by the majority of papilionoid legumes (Doyle et al., 1996). The AT content of the chickpea plastid genome is 66.1%, similar to other legumes including *Glycine max* (64.63%), *Lotus japonicus* (64.0%), *Phaseolus vulgaris* (64.56%), and *Medicago truncatula* (66.03%).

The chickpea plastid genome has 108 total genes, including 4 rRNA genes, 29 tRNA genes, and 75 protein-coding genes. Three genes, *rps16*, *infA*, and *ycf4*, found in most angiosperm plastid genomes, including representatives of the early-branching lineages (Goremykin et al., 2003; Hansen et al., 2007; Raubeson et al., 2007) are not present in the chickpea plastid genome. In *ndhB*, there is an internal stop codon, similar to other legume plastid genomes. There is no stop codon (Met...Val) in the *rps8* gene, a characteristic feature of the *Medicago truncatula* plastid genome.

Fifteen genes contain one or two introns, nine of which are in protein-coding genes while six are in the tRNA genes. The protein-coding gene *rpl2* in chickpea, which contains a single intron of 669 bp has 16 amino acids missing at the 5'-end relative to most other legumes (Fig. 2), while *Medicago* has 11 amino acids missing in the same portion of the *rpl2* gene. Among intron-containing genes, *trnK-UUU* has the largest intron (2491 bp), and it includes the *matK* gene. The smallest intron is in *trnL-UAA* (555 bp). The *ycf3* gene has two introns of 733 and 737 bp.

3.2. Phylogenetic distribution of intron losses in *clpP* and *rps12* genes among legumes

In the *Cicer arietenum* chloroplast genome we observed the absence of the *clpP* and *rps12* 3'-end introns. Observation of loss of both of these introns represents the first documented case of such loss within the same plastid genome. Therefore, 301 legume taxa from 198 genera and one member of the related family Polygalaceae (Table 1) were subjected to a PCR-based survey for the presence/absence of the *clpP* and *rps12*-3'-end introns. These taxa represent all major groups of legumes, with 23 caesalpinoid species (18 genera), 18 mimosoid species (15 genera), 260 species of papilionoids (165 genera), and the genus *Polygala* L. in the family Polygalaceae (potential sister group to Leguminosae in Fabales; sensu APG II, 2003). For *clpP*, expected fragments sizes should be 1100–1300 bp if the intron is present and 300–350 bp if the intron is lost (Fig. 3). For *rps12*, expected fragments sizes should be 800–850 bp if the intron is present and 250–280 bp if the intron is lost (Fig. 4). Comparison of the sizes of both the *clpP* and *rps12* PCR products from a diversity of legumes, represented in Figs. 3 and 4, reveals fragments very similar in length, a result consistent with a process by which plastid introns are excised precisely and entirely, as observed earlier by Doyle et al. (1995) and in other taxa by Downie et al. (1991). Indeed, sequence analysis of ten taxa selected from our survey that included those with or without the *clpP* and *rps12* introns confirms that intron excision has occurred at precisely the same points in the gene sequence in each taxon lacking the intron (MFW, unpublished data). Furthermore, the minor variation in length of the *clpP* PCR products we observed (Fig. 3A) was due to the length of the intron in each taxon, which ranged from 702 bp (*Lespedeza cuneata*) to 799 bp (*Lotus corniculatus*) in the taxa we sequenced, whereas

the length of the *rps12*-3' intron was 529–532 bp in the taxa analyzed (*Glycine max*, *Phaseolus vulgaris*, *Callerya reticulata*; MFW, unpublished data).

The presence of extra minor PCR products, reactions that were weak or inconclusive, and polymorphic for intron presence/absence have been identified in our survey for these two introns (Table 1). Virtually all of these examples were found in reactions surveying for the first (of two) *clpP* intron in taxa that presumably have the intron present. Possible explanations include partial excision reactions, poor quality DNA (many taxa have been sampled from herbarium specimens), sequence variation in primer binding sites or rearrangements in/near the gene containing the intron.

Our survey (Table 1) reveals that the loss of the *clpP* intron is, with a few exceptions, limited to the large IR-lacking clade (Fig. 5). Loss of the *clpP* intron was detected in individual accessions of the mimosoid *Inga* Mill. and the papilionoids *Aotus* Sm., and *Hypocalyptus* Thunb., *Platymiscium* Vogel, although the latter three results are ambiguous and need to be confirmed with sampling of additional specimens. The *rps12* intron is also lost in all members of the IRLC surveyed, with notable exceptions (all accessions of *Callerya* Endl., *Wisteria* Nutt., *Afgekia* Craib, and *Millettia japonica* A. Gray). This intron is also lost independently in the papilionoid tribe Desmodieae (Fig. 5), a monophyletic group nested in the “Millettioids-Phaseoloids” clade (Kajita et al., 2001; Wojciechowski et al., 2004). Desmodieae, which consists of 30 genera and ca. 530 species distributed in tropical to warm temperate regions of the world (Lewis et al., 2005), is represented in this analysis by the genera *Alysicarpus* Desv., *Campylotropis* Bunge, *Desmodium* Desv., *Kummerowia* Schindl., and *Lespedeza* Michx. (Table 1).

4. Discussion

4.1. Comparison of gene/intron content and genome organization among legumes

A number of previous studies indicated that legume plastid genomes have experienced substantial numbers of rearrangements (Palmer and Thompson, 1981, 1982; Palmer et al., 1988; Lavin et al., 1990; Bruneau et al., 1990; Doyle et al., 1995, 1996). Complete sequencing of plastid genomes of five legumes (*Cicer*, *Glycine*, *Lotus*, *Medicago*, and *Phaseolus*), combined with earlier gene mapping studies, have revealed nine genomic rearrangements, including two large inversions, the loss of the IR, three gene losses and three intron losses, two of which are reported here (Table 2). Thus, the legumes represent one of only a few angiosperm families that have experienced multiple, plastid genomic rearrangements and gene/intron losses (Jansen et al., 2007), and serve as an excellent choice in which to investigate contrasting patterns of plastid DNA evolution. Other families known to have comparable plastid genomic rearrangements include Asteraceae (two inversions; Jansen and Palmer, 1987; Kim et al., 2005), Campanulaceae (up to 42 inversions, two gene losses, 8 putative transpositions; Cosner et al., 1997, 2004; Haberle et al., 2008), Geraniaceae (12 inversions, 8 IR boundary changes; Chumley et al., 2006), Lobeliaceae (11 inversions; Knox et al., 1993), Oleaceae (4 inversions, 1 gene loss, 1 intron loss, 5 gene duplications; Lee et al., 2007), Poaceae (3 inversions, 3 gene losses, 2 intron losses; Doyle et al., 1992), and Ranunculaceae (9 inversions, 1 intron loss; Hoot and Palmer, 1994; Johansson, 1999).

The causes for the propensity of plastid genomic changes in these lineages are unknown but several explanations have been proposed. For legumes, it was suggested that the loss of the IR has a destabilizing effect on genome organization (Palmer, 1991; Palmer and Thompson, 1982). However, given that most of the angiosperms with highly rearranged plastid genomes still retain two copies of the IR (i.e., Campanulaceae, Lobeliaceae, Oleaceae, Poaceae, Ranunculaceae, and most Geraniaceae) IR-loss does not provide a general explanation for the extensive rearrangements in plastid genomes. Moreover, the fact that the majority of the known

plastid genomic rearrangements in legumes are also found in papilionoid taxa with two copies of the IR in their genomes (Table 2; Fig. 5) argues against this explanation as well. Another possible explanation for the higher incidence of rearrangements in some lineages is the presence of dispersed sequence repeats, which could facilitate rearrangements by intramolecular recombination (Palmer, 1991). In wheat (Ogihara et al., 1988) and *Oenothera* (Hupfer et al., 2000), such repeats have been directly implicated in inversions, and the strong correlation detected between the number of dispersed repeats and the extent of genomic rearrangements in several lineages is consistent with this explanation (Pombert et al., 2005, 2006; Haberle et al., 2008). Recent comparisons of the number and distribution of repeated sequences in completely sequenced legume plastid genomes have demonstrated the presence of numerous dispersed repeats, many more than in related rosoid genomes that are not rearranged (Saski et al., 2005).

4.2. Phylogenetic distribution of chloroplast genomic rearrangements among legumes

We have plotted the distribution of several of the less ambiguous of the nine legume plastid genomic rearrangements (Table 2) on a phylogenetic tree based on cladistic analyses of complete nucleotide sequences of the plastid *matK* gene to assess the phylogenetic implications of these rare genomic changes (Fig. 5; tree summarized from Wojciechowski et al., 2004). Clearly, most plastid genomic rearrangements among legumes are restricted to the papilionoids with the exception of the loss of the *rpl22* gene (Downie and Palmer, 1992; Doyle et al., 1995), which characterizes all taxa sampled from all three subfamilies of legumes, and the loss of the *rpl2* intron in numerous species of the caesalpinoid genus *Bauhinia* (Lai et al., 1997). Interestingly, the *rpl22* gene has not been lost from any other land plants (Downie and Palmer, 1992), and a functional copy has been isolated from the nuclear genome in *Pisum sativum* (Gantt et al., 1991).

The loss of one copy of the IR, as originally suggested by Lavin et al. (1990), has occurred only once among legumes and is restricted to a large clade of papilionoid legumes that includes the traditional tribes Carmichaelieae, Cicereae (chickpea), Galegeae, Hedysareae, Trifolieae, and Fabae (Viciae) and several genera formerly treated in the tribe Millettieae (so-called “IRLC millettoids”), including *Callerya*, *Wisteria*, *Afgekia*, *Endosamara* R. Geesink, and probably *Antheroporum* Gagnep. and *Sarcodum* Lour. (Hu et al., 2000, 2002; Hu and Chang, 2003). Taxa lacking the IR have been shown to form a monophyletic group informally known as the IRLC, which is well supported in phylogenetic trees based plastid *matK* and nuclear rDNA sequence analyses (Hu et al., 2000, 2002; Wojciechowski et al. 2000, 2004). From its taxonomic distribution, this mutational event in the plastid genome occurred relatively later in the evolution and diversification of the Papilionoideae, a molecular synapomorphy for a large (ca. 4400 species), derived group of primarily herbaceous taxa with a temperate distribution (Wojciechowski et al., 2000, 2004) and an estimated age of 39 Ma (Lavin et al., 2005). Like the now established monophyly of the taxa marked by loss of the IR (Palmer et al., 1988), results from molecular phylogenetic studies have provided both greater resolution and corroborating evidence for the relationships of many of the temperate and tropical groups long considered “derived” within papilionoids based upon the presence of morphological and/or chemical characters that have served as important taxonomic markers (Polhill, 1994). For example, the hypothesis for a single origin of canavanine biosynthesis (production of a non-protein amino acids such as L-canavanine, a close analog of arginine) in Papilionoideae (Bell, 1981), which occurs in all the tribes that comprise the IRLC plus the related tropical tribes that comprise their sister group (i.e., Millettieae, Phaseoleae, and allies), has been supported by recent analyses of plastid *rbcL* and *matK* gene sequences in legumes (Kajita et al., 2001; Wojciechowski et al., 2004). Distribution of the two legume plastid genome inversions provided additional support for clades identified in phylogenetic trees based on analyses of gene sequences. The 50-kb inversion defines an early evolutionary split in the diversification

of the papilionoid clade with all members of this clade having the inversion except for taxa from the tribes Sophoreae, Swartzieae, and Dipterygeae (Doyle et al., 1996; Pennington et al., 2001; Wojciechowski et al., 2004), although the exact membership of this clade remains unresolved (due to lack of sampling all relevant taxa). The 78-kb inversion is much more limited in its distribution, being restricted to a majority of the genera that traditionally comprise subtribe Phaseolinae of the tribe Phaseoleae (Bruneau et al., 1990), which is also supported as a monophyletic group in trees based on plastid gene sequences (Thulin et al., 2004; Wojciechowski et al., 2004). Recent evidence indicates this inversion may be a synapomorphy for this clade, defined by the most recent common ancestor of the genera *Wajira* Thulin and *Phaseolus* L. (M. Moore, M.F. Wojciechowski, A. Delgado, and P.S. Soltis, unpublished data).

Two other gene losses in legumes have been detected in at least one genus in 15 (*rps16* and *ycf4*) of the 28 papilionoid tribes (sensu Lewis et al., 2005). The taxonomic distribution of the *rps16* loss based on filter hybridizations (Doyle et al., 1995) suggested multiple, independent losses within papilionoids but more rigorous PCR and sequencing strategies are needed to confirm these events. Among 64 completely sequenced seed plant plastid genomes there have been four independent losses of *rps16*: in *Pinus*, legumes, two members of the Malpighiales (*Passiflora* and *Populus*) and the monocot *Dioscorea* (Jansen et al., 2007). This gene has also been lost in the genus *Adonis* in the Ranunculaceae based on filter hybridization data (Johansson, 1999). The loss of *ycf4* (formerly called ORF184) has been documented in *Pisum* (Nagano et al., 1991; Smith et al., 1991) and it is lacking in three (*Cicer*, *Glycine*, and *Medicago*) of the five completely sequenced legume plastid genomes. An earlier survey of the phylogenetic distribution of this loss among 392 legume genera based on filter hybridization screens with *ycf4* gene-specific probes indicated at least 15 independent losses within tribe Phaseoleae alone (Doyle et al., 1995). Both the genome sequences and the filter hybridization data suggest that considerable homoplasy will limit the phylogenetic utility of this gene loss within legumes.

The two plastid genomic changes identified by sequencing the chickpea genome provide valuable information for resolving relationships among the IRLC papilionoid legumes (Table 2 and Fig. 5). Intron losses for both the *clpP* and *rps12* genes have been identified in other angiosperm lineages as well. For example, both of the *clpP* gene introns have been lost in Poaceae, Onagraceae, Oleaceae, and *Pinus* (reviewed in Jansen et al., 2007), and the intron in the 3'-end of *rps12* has been lost independently twice in the monocot order Asparagales (McPherson et al., 2004). However, the losses in *Cicer* represent the first documented case of the loss of introns from both of these genes in the same plastid genome. The *clpP* intron loss, which appears to have occurred only once within Leguminosae, provides additional support for the monophyly of the IRLC.

While the data suggest that loss of the *rps12* intron generally coincides phylogenetically with the loss of the IR (Table 1), the distribution of the *rps12* intron loss is more informative because it marks a slightly less-inclusive clade within the IRLC that provides additional data to resolve relationships among the early-branching lineages of the IRLC. That *Callerya*, *Wisteria*, *Afgekia*, and *Millettia japonica* unambiguously possess the intron in the *rps12* gene is interesting because trees based on nucleotide sequences (Wojciechowski et al., 2000,2004;Hu et al., 2002) have not been able to resolve the relationships of these and other lineages at the base of this clade. Indeed, results from molecular phylogenetic analyses are not in agreement on this point, with some data suggesting *Glycyrrhiza* L. is the sister group to the rest of the IRLC or *Glycyrrhiza* + *Callerya* and/or *Wisteria* s.l. (i.e., including other IRLC millettoids) are sister to the rest of the IRLC. While the consensus seems to be that *Callerya* + *Wisteria* s.l. form a clade (e.g., Hu and Chang, 2003), the presence of the *rps12* intron in *Callerya*, *Wisteria*, and other members of the IRLC millettoids suggests they comprise the earliest-branching lineages that form the sister group to the rest of the IRLC, which are characterized

by the loss of the intron from *rps12*. Furthermore, this result indicates that the loss of this intron occurred subsequent to the loss of one copy of the IR in these taxa. A second, independent loss of the *rps12* intron has occurred in the more distantly related tribe Desmodieae (Fig. 5), which retains both copies of the IR, a group that is also marked by loss of the *rpl2* intron (Doyle et al. 1995; Bailey et al. 1997).

5. Conclusions

Legume plastid genomes have undergone considerable diversification in gene/intron content and gene order, and these changes provide valuable information for resolving phylogenetic relationships among and within some major clades identified on the basis of analyses of DNA sequences. The two new genomic changes identified in the present study provide additional support of the monophyly of the IR-loss clade, and resolution of the early-branching pattern in this clade. In addition to providing insight into plastid genome evolution and phylogenetic relationships of legumes, the availability of complete plastid genome sequences facilitates plastid genetic engineering for improvement of agronomic traits and production of vaccines, biopharmaceuticals, biomaterials and industrial enzymes. Complete plastid genome sequences provide valuable information on spacer regions for integration of transgenes at optimal sites via homologous recombination, as well as endogenous regulatory sequences for optimal expression of transgenes, and should help in expanding plastid technology to other economically important crops.

Acknowledgments

Investigations reported in this article were supported in part by Grants from USDA 3611-21000-017-00D and NIH R01 GM 63879 to H.D. Research by R.K.J. was supported, in part, by NSF ATOL Grant DEB0120709. M.F.W. was supported, in part, by NSF Grant DEB0542958. The authors thank Matt Lavin, Alfonso Delgado-Salinas, Jennifer Trusty, Kelly Steele, Jer-Ming Hu, R. Toby Pennington, Mats Thulin, Aaron Listen, herbaria of The Royal Botanic Gardens at Kew (K), Edinburgh (E), Munich (M/MSB), and the Arizona State University Vascular Plant Herbarium (ASU) for tissue or DNA samples, and Dr. Michael Bausher (USDA) for help in early stages of genome sequencing. We also thank Anne Bruneau and an anonymous reviewer for their helpful comments on the manuscript.

References

- Angiosperm Phylogeny Group. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Bot J Linn Soc* 2003;141:399–436.
- Bailey CD, Doyle JJ, Kajita T, Nemoto T, Ohashi H. The chloroplast *rpl2* intron and ORF184 as phylogenetic markers in the legume tribe Desmodieae. *Syst Bot* 1997;22:13–138.
- Bausher MG, Singh ND, Lee SB, Jansen RK, Daniell H. The complete chloroplast genome sequence of *Citrus sinensis* (L.) Osbeck var ‘Ridge Pineapple’: organization and phylogenetic relationships to other angiosperms. *BMC Plt Biol* 2006;6:21.
- Bell, EA. Non-protein amino acids in the Leguminosae. In: Polhill, RM.; Raven, PH., editors. *Advances in Legume Systematics, Part 2*. Royal Botanic Gardens; Kew, UK: 1981. p. 489-499.
- Bruneau A, Doyle JJ, Palmer JD. A chloroplast DNA inversion as a subtribal character in the Phaseoleae (Leguminosae). *Syst Bot* 1990;15:378–386.
- Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK. The complete chloroplast genome sequence of *Pelargonium × hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol Biol Evol* 2006;23:2175–2190. [PubMed: 16916942]
- Cosner ME, Jansen RK, Palmer JD, Downie SR. The highly rearranged chloroplast genome of *Trachelium caeruleum* Campanulaceae: multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families. *Curr Genet* 1997;31:419–429. [PubMed: 9162114]
- Cosner ME, Raubeson LA, Jansen RK. Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. *BMC Evol Biol* 2004;4:27. [PubMed: 15324459]

- Daniell H. Molecular strategies for gene containment in transgenic crops. *Nat Biotechnol* 2002;20:581–586. [PubMed: 12042861]
- Daniell H. Transgene containment by maternal inheritance: Effective or elusive? *Proc Natl Acad Sci USA* 2007;104:6879–6880. [PubMed: 17440039]
- Daniell H, Datta R, Varma S, Gray S, Lee SB. Containment of herbicide resistance through genetic engineering of the chloroplast genome. *Nat Biotechnol* 1998;16:345–348. [PubMed: 9555724]
- Daniell H, Lee SB, Panchal T, Wiebe PO. Expression of the native cholera toxin B subunit gene and assembly as functional oligomers in transgenic tobacco chloroplasts. *J Mol Biol* 2001;311:1001–1009. [PubMed: 11531335]
- Daniell H, Khan MS, Allison L. Milestones in chloroplast genetic engineering: an environmentally friendly era in biotechnology. *Trends Plant Sci* 2002;7:84–91. [PubMed: 11832280]
- Daniell H, Kumar S, Dufourmantel N. Breakthrough in chloroplast genetic engineering of agronomically important crops. *Trends Biotechnol* 2005;23:238–245. [PubMed: 15866001]
- Daniell H, Lee SB, Grevich J, Saksi C, Quesada-Vargas T, Guda C, Tomkins J, Jansen RK. Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analyses with other Solanaceae genomes. *Theor Appl Genet* 2006;112:1503–1518. [PubMed: 16575560]
- Daniell H, Wurdack KJ, Kanagaraj A, Lee SB, Saksi C, Jansen RK. The complete nucleotide sequence of the cassava (*Manihot esculenta*) chloroplast genome and multiple losses of the *atpF* intron in Malpighiales. *Theor Appl Genet* 2008;116:723–737. [PubMed: 18214421]
- DeCosa B, Moar W, Lee SB, Milller M, Daniell H. Overexpression of the Bt Cry2Aa2 operon in chloroplasts leads to formation of insecticidal crystals. *Nat Biotechnol* 2001;19:71–74. [PubMed: 11135556]
- DeGray G, Rajasekaran K, Smith F, Sanford J, Daniell H. Expression of an antimicrobial peptide via the chloroplast genome to control phytopathogenic bacteria and fungi. *Plant Physiol* 2001;127:852–862. [PubMed: 11706168]
- Downie, SR.; Palmer, JD. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. In: Soltis, PS.; Soltis, DE.; Doyle, JJ., editors. *Molecular Systematics of Plants*. Chapman and Hall; New York: 1992. p. 14-35.
- Downie SR, Olmstead RG, Zurawski G, Soltis DE, Soltis PS, Watson JC, Palmer JD. Six independent losses of the chloroplast DNA *rpl2* intron in dicotyledons: molecular and phylogenetic implications. *Evolution* 1991;45:1245–1259.
- Downie SR, Llanas E, Katz-Downie DS. Multiple independent losses of the *rpoC1* intron in angiosperm chloroplast DNAs. *Syst Bot* 1996;21:135–151.
- Doyle, JJ. DNA data and legume phylogeny: a progress report. In: Crisp, MD.; Doyle, JJ., editors. *Advances in Legume Systematics, Part 7: Phylogeny*. Royal Botanic Gardens; Kew, UK: 1995. p. 11-30.
- Doyle JJ, Davis JI, Soreng RJ, Garvin D, Anderson MJ. Chloroplast DNA inversions and the origin of the grass family (Poaceae). *Proc Natl Acad Sci USA* 1992;89:7723–7726.
- Doyle JJ, Doyle JL, Palmer JD. Multiple independent losses of two genes and one intron from legume chloroplast genomes. *Syst Bot* 1995;20:272–294.
- Doyle JJ, Doyle JL, Palmer JD. The distribution and phylogenetic significance of a 50-kb chloroplast DNA inversion in the flowering plant family Leguminosae. *Mol Phylogenet Evol* 1996;5:429–438. [PubMed: 8728401]
- Doyle JJ, Doyle JL, Ballenger JA, Dickson EE, Kajita T, Ohashi H. A phylogeny of the chloroplast gene *rbcL* in the Leguminosae: taxonomic correlations and insights into the evolution of nodulation. *Am J Bot* 1997;84:541–554.
- Doyle, JJ.; Chappill, JA.; Bailey, CD.; Kajita, T. Towards a comprehensive phylogeny of legumes: evidence from *rbcL* sequences and non-molecular data. In: Herendeen, PS.; Bruneau, A., editors. *Advances in Legume Systematics, Part 9*. Royal Botanic Gardens; Kew, UK: 2000. p. 1-20.
- Dufoumantel N, Pelissier B, Gracon F, Peltier G, Ferullo JM, Tissot G. Generation of fertile transplastomic soybean. *Plant Mol Biol* 2004;55:479–489. [PubMed: 15604694]

- Dufourmantel N, Tissot G, Goutorbe F, Garcon F, Muhr C, Jansens S, Pelissier B, Peltier G, Dubald M. Generation and analysis of soybean plastid transformants expressing *Bacillus thuringiensis* Cry1Ab protoxin. *Plant Mol Biol* 2005;58:658–659.
- Gantt JS, Baldauf SL, Calie PJ, Weeden NF, Palmer JD. Transfer of *rpl22* to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. *EMBO J* 1991;10:3073–3078. [PubMed: 1915281]
- Goremykin VV, Hirsch-Ernst KI, Wolf S, Hellwig FH. Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. *Mol Biol Evol* 2003;20:1499–1505. [PubMed: 12832641]
- Greiner S, Wang X, Rauwolf U, Silber MV, Mayer K, Meurer J, Haberer G, Herrmann RG. The complete nucleotide sequences of the five genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: I. Sequence evaluation and plastome evolution. *Nucl Acids Res*. 2008;10.1093/nar/gkn081
- Guo X, Castillo-Ramírez S, González V, Bustos P, Luís Fernández-Vázquez J, Santamaría RI, Arellano J, Cevallos MA, Dávila G. Rapid evolutionary change of common bean (*Phaseolus vulgaris* L.) plastome, and the genomic diversification of legume chloroplasts. *BMC Genomics* 2007;8:228. [PubMed: 17623083]
- Haberle RC, Fourcade HM, Boore JL, Jansen RK. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *J Mol Evol*. 2008;10.1007/s00239-008-9086-4
- Hachtel W, Neuss A, vom Steim J. A chloroplast DNA inversion marks an evolutionary split in the genus *Oenothera*. *Evolution* 1991;45:1050–1052.
- Hansen DR, Dastidar SG, Cai Z, Penafior C, Kuehl JV, Boore JL, Jansen RK. Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae). *Mol Phylogenet Evol* 2007;45:547–563. [PubMed: 17644003]
- Hoot SB, Palmer JD. Structural rearrangements, including parallel inversions, within the chloroplast genome of *Anemone* and related genera. *J Mol Evol* 1994;38:274–281. [PubMed: 8006994]
- Hu JM, Chang SP. Two new members of the *Callerya* group (Fabaceae) based on phylogenetic analysis of *rbcL* sequences: *Endosamara racemosa* (Roxb.) Geesink and *Callerya vasta* (Kosterm.). *Schot Taiwan* 2003;48:118–128.
- Hu JM, Lavin M, Wojciechowski MF, Sanderson MJ. Phylogenetic systematics of the tribe Millettieae (Leguminosae) based on *matK* sequences, and implications for evolutionary patterns in Papilionoideae. *Am J Bot* 2000;87:418–430. [PubMed: 10719003]
- Hu JM, Lavin M, Wojciechowski MF, Sanderson MJ. Phylogenetic analysis of nuclear ribosomal ITS/5.8 S sequences in the tribe Millettieae (Fabaceae): *Poecilanthus-Cyclolobium*, the core Millettieae, and the *Callerya* group. *Syst Bot* 2002;27:722–733.
- Hupfer H, Swiatek M, Hornung S, Herrmann RG, Maier RM, Chiu WL, Sears B. Complete nucleotide sequence of the *Oenothera elata* plastid chromosome, representing plastome I of the five distinguishable *Euoenothera* plastomes. *Mol Gen Genet* 2000;263:581–585. [PubMed: 10852478]
- Hussein H, Ruiz ON, Terry N, Daniell H. Phytoremediation of mercury and organomercurials in chloroplast transgenic plants: enhanced root uptake, translocation to shoots and volatilization. *Environ Sci Technol* 2007;41:8439–8446. [PubMed: 18200876]
- Jansen RK, Palmer JD. A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). *Proc Natl Acad Sci USA* 1987;84:5818–5822. [PubMed: 16593871]
- Jansen RK, Raubeson LA, Boore JL, de Pamphilis CW, Chumley TW, Haberle RC, Wyman SK, Alverson AJ, Peery R, Herman SJ, Fourcade HM, Kuehl JV, McNeal JR, Leebens-Mack J, Cui L. Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods Enzymol* 2005;395:348–384. [PubMed: 15865976]
- Jansen RK, Kaittanis C, Saski C, Lee SB, Tomkins J, Alverson AJ, Daniell H. Phylogenetic analyses of *Vitis* (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among rosids. *BMC Evol Biol* 2006;6:32. [PubMed: 16603088]

- Jansen RK, Cai Z, Raubeson LA, Daniell H, dePamphilis CW, Leebens-Mack J, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW, Lee SB, Peery R, McNeal J, Kuehl JV, Boore JL. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc Natl Acad Sci USA* 2007;104:19369–19374. [PubMed: 18048330]
- Johansson JT. Three large inversions in the chloroplast genomes and one loss of the chloroplast gene *rps16* suggest an early evolutionary split in the genus *Adonis* (Ranunculaceae). *Plant Syst Evol* 1999;218:133–143.
- Kajita TH, Ohashi Y, Tateishi Y, Bailey CD, Doyle JJ. *rbcL* and legume phylogeny, with particular reference to Phaseoleae, Millettieae, and Allies. *Syst Bot* 2001;26:515–536.
- Kamarajugadda S, Daniell H. Chloroplast-derived anthrax vaccine and other vaccine antigens: their immunogenic and immunoprotective properties. *Expert Rev Vaccines* 2006;5:839–849. [PubMed: 17184221]
- Käss E, Wink M. Molecular phylogeny of the Papilionoideae (Family Leguminosae): *rbcL* gene sequences versus chemical taxonomy. *Bot Acta* 1995;108:149–162.
- Käss E, Wink M. Molecular evolution of the Leguminosae: phylogeny of the three subfamilies based on *rbcL* sequences. *Biochem Syst Ecol* 1996;24:365–378.
- Käss E, Wink M. Phylogenetic relationships in the Papilionoideae (Family Leguminosae) based on nucleotide sequences of cpDNA (*rbcL*) and ncDNA (ITS1 and 2). *Mol Phylogenet Evol* 1997;8:65–88. [PubMed: 9242596]
- Kato T, Kaneko T, Sato S, Nakamura Y, Tabata S. Complete structure of the chloroplast genome of a legume, *Lotus japonicus*. *DNA Res* 2000;7:323–330. [PubMed: 11214967]
- Kim YD, Jansen RK. Phylogenetic relationships of chloroplast DNA variation in the Berberidaceae. *Plt Syst Evol* 1995;9:341–349.
- Kim HG, Choi KS, Jansen RK. Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). *Mol Biol Evol* 2005;22:1–10. [PubMed: 15356277]
- Knox EB, Downie SR, Palmer JD. Chloroplast genome rearrangements and the evolution of giant lobelias from herbaceous ancestors. *Mol Biol Evol* 1993;10:414–430.
- Kumar S, Dhingra A, Daniell H. Chloroplast-expressed betaine aldehyde dehydrogenase gene in carrot cultured cells, roots and leaves confers enhanced salt tolerance. *Plant Physiol* 2004;136:2843–2854. [PubMed: 15347789]
- Lai M, Sceppa J, Ballenger JA, Doyle JJ, Wunderlin RP. Polymorphism for the presence of the *rpl2* intron in chloroplast genomes of *Bauhinia* (Leguminosae). *Syst Bot* 1997;22:519–528.
- Lavin M, Doyle JJ, Palmer JD. Evolutionary significance of the loss of the chloroplast–DNA inverted repeat in the Leguminosae subfamily Papilionoidae. *Evolution* 1990;44:390–402.
- Lavin M, Herendeen PS, Wojciechowski MF. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Syst Biol* 2005;54:530–549. [PubMed: 16085573]
- Lee SB, Kwon H, Kwon S, Park S, Jeong M, Han S, Daniell H. Accumulation of trehalose within transgenic chloroplast confers drought tolerance. *Mol Breed* 2003;11:1–13.
- Lee SB, Kaittanis C, Jansen RK, Hostetler JB, Tallon LJ, Town CD, Daniell H. The complete chloroplast genome sequence of *Gossypium hirsutum*: organization and phylogenetic relationships to other angiosperms. *BMC Genomics* 2006;7:61. [PubMed: 16553962]
- Lee HL, Jansen RK, Chumley TW, Kim KJ. Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions. *Mol Biol Evol* 2007;24:1161–1180. [PubMed: 17329229]
- Leelavathi S, Gupta N, Maiti S, Ghosh A, Reddy VS. Overproduction of an alkali- and thermo-stable xylanase in tobacco chloroplasts and efficient recovery of the enzyme. *Mol Breed* 2003;11:59–67.
- Lewis, G.; Schrire, B.; Mackinder, B.; Lock, M., editors. *Legumes of the World*. The Royal Botanic Gardens; Kew, UK: 2005.
- Liston, A. Use of the polymerase chain reaction to survey for the loss of the inverted repeat in the legume chloroplast genome. In: Crisp, M.; Doyle, JJ., editors. *Advances in Legume Systematics 7: Phylogeny*. Royal Botanic Gardens; Kew: 1995. p. 31–40.

- McBride KE, Svab Z, Schaaf DJ, Hogan PS, Stalker DM, Maliga P. Amplification of a chimeric *Bacillus* gene in chloroplasts leads to an extraordinary level of an insecticidal protein in tobacco. *BioTechnology* 1995;13:362–365. [PubMed: 9634777]
- McPherson MA, Fay MF, Chase MW, Graham SW. Parallel loss of a slowly evolving intron from two closely related families in Asparagales. *Syst Bot* 2004;29:296–307.
- Nagano Y, Matsuno R, Sasaki Y. Sequence and transcriptional analysis of the gene cluster *trnQ-zfpA-psal*-ORF231-*petA* in pea chloroplasts. *Curr Genet* 1991;20:431–436. [PubMed: 1807835]
- Ogihara Y, Terachi T, Sasakuma T. Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. *Proc Natl Acad Sci USA* 1988;85:8573–8577. [PubMed: 3186748]
- Palmer, JD. Plastid chromosomes: structure and evolution. In: Bogorad, L.; Vasil, I., editors. *Cell Culture and Somatic Cell Genetics of Plants*. Academic Press; San Diego, California, USA: 1991. p. 5-53.
- Palmer JD, Thompson WF. Rearrangements in the chloroplast genomes of mung bean and pea. *Proc Natl Acad Sci USA* 1981;78:5533–5537. [PubMed: 16593087]
- Palmer JD, Thompson WF. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell* 1982;29:537–550. [PubMed: 6288261]
- Palmer JD, Osorio B, Thompson W. Evolutionary significance of inversions in legume chloroplast DNAs. *Curr Genet* 1988;14:65–74.
- Pennington RT, Lavin M, Ireland H, Klitgaard BB, Preston J, Hu JM. Phylogenetic relationships of basal papilionoid legumes based upon sequences of the chloroplast *trnL* intron. *Syst Bot* 2001;26:537–556.
- Polhill, RM. Classification of the Leguminosae. In: Bisby, FA.; Buckingham, J.; Harborne, JB., editors. *Phytochemical Dictionary of the Leguminosae*. Chapman and Hall; New York, USA: 1994. p. xxxv-xlviii.
- Pombert JF, Otis C, Lemieux C, Turmel M. The chloroplast genome sequence of the green alga *Pseudendoclonium akinetum* Ulvophyceae reveals unusual structural features and new insights into the branching order of chlorophyte lineages. *Mol Biol Evol* 2005;22:1903–1918. [PubMed: 15930151]
- Pombert JF, Lemieux C, Turmel M. The complete chloroplast DNA sequence of the green alga *Oltmannsiellopsiviridis* reveals a distinctive quadripartite architecture in the chloroplast genome of early diverging ulvophytes. *BMC Biol* 2006;4:3. [PubMed: 16472375]
- Quesda-Vargas T, Ruiz ON, Daniell H. Over expression of the Bt *cry2Aa2* operon in transgenic chloroplasts: transcription, processing, translation. *Plant Physiol* 2005;138:1746–1762. [PubMed: 15980187]
- Raubeson, LA.; Jansen, RK. Chloroplast genomes of plants. In: Henry, H., editor. *Diversity and Evolution of Plants—Genotypic and Phenotypic Variation in Higher Plants*. CABI Publishing; UK: 2005. p. 45-68.
- Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, Jansen RK. Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics* 2007;8:174. [PubMed: 17573971]
- Ravi V, Khurana JP, Tyagi AK, Khurana P. The chloroplast genome of mulberry: complete nucleotide sequence, gene organization and comparative analysis. *Tree Genet Genom* 2006;3:49–59.
- Ruhlman T, Lee SB, Jansen RK, Hostetler JB, Tallon LJ, Town CD, Daniell H. Complete chloroplast genome sequence of *Daucus carota*: implications for biotechnology and phylogeny of angiosperms. *BMC Genomics* 2006;7:222. [PubMed: 16945140]
- Ruiz ON, Daniell H. Engineering cytoplasmic male sterility via chloroplast genome by expression of beta-ketothiolase. *Plant Physiol* 2005;138:1232–1246. [PubMed: 16009998]
- Ruiz ON, Hussein H, Terry N, Daniell H. Phytoremediation of organomercurial compounds via chloroplast genetic engineering. *Plant Physiol* 2003;132:1344–1352. [PubMed: 12857816]
- Samson N, Bausher MG, Lee SB, Jansen RK, Daniell H. The complete nucleotide sequence of the coffee (*Coffea arabica* L.) chloroplast genome: organization and implications for biotechnology and phylogenetic relationships amongst angiosperms. *Plt Biotechnol J* 2007;5:339–353.

- Saski C, Lee SB, Daniell H, Wood T, Tomkins J, Kim HG, Jansen RK. Complete chloroplast genome sequence of *Glycine Max* and comparative analyses with other legume genomics. *Plant Mol Biol* 2005;59:309–322. [PubMed: 16247559]
- Saski C, Lee SB, Fjellheim S, Guda C, Jansen RK, Luo H, Tomkins J, Rognli OA, Daniell H, LiuClarke J. Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes. *Theor Appl Genet* 2007;115:571–590. [PubMed: 17534593]
- Schwartz S, Elnitski L, Li M, Weirauch M, Riemer C, Smit A, Program NCS, Green ED, Hardison RC, Miller W. MultiPipMaker and supporting tools: alignments and analysis of multiple genomic DNA sequences. *Nucleic Acids Res* 2003;31:3518–3524. [PubMed: 12824357]
- Smith AG, Wilson RM, Kaethner TM, Willey DL, Gray JC. Pea chloroplast genes encoding a 4 kDa polynucleotide of photosystem I and a putative enzyme of C₁ metabolism. *Curr Genet* 1991;19:403–410. [PubMed: 1913879]
- Steane DA. Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum, *Eucalyptus globulus* (Myrtaceae). *DNA Res* 2005;12:215–220. [PubMed: 16303753]
- Thulin M, Lavin M, Pasquet R, Delgado-Salinas A. Phylogeny and biogeography of *Wajira* (Leguminosae): a monophyletic segregate of *Vigna* centered in the Horn of Africa region. *Syst Bot* 2004;29:903–920.
- Verma D, Daniell H. Chloroplast vector systems for biotechnology applications. *Plant Physiol* 2007;145:1129–1143. [PubMed: 18056863]
- Wallace RS, Cota JH. An intron loss in the chloroplast gene *rpoC1* supports a monophyletic origin for the subfamily Cactoideae of the Cactaceae. *Curr Genet* 1996;29:275–281. [PubMed: 8595674]
- Wojciechowski, MF.; Sanderson, MJ.; Steele, KP.; Liston, A. Molecular phylogeny of the “temperate herbaceous tribes” of papilionoid legumes: a supertree approach. In: Herendeen, P.; Bruneau, A., editors. *Advances in Legume Systematics, Part 9*. Royal Botanic Garden; Kew: 2000. p. 277-298.
- Wojciechowski MF, Lavin M, Sanderson MJ. A phylogeny of legumes (Leguminosae) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family. *Am J Bot* 2004;91:1846–1862.
- Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 2004;20:3252–3255. [PubMed: 15180927]

1 MAIHLYKTS-----S-----RNRLIYGQHHCS^TGRNARGIITAK^KHRGGG chickpea
 1 MAIHLYKTSIPSTR-----RNRLIYGQHHCGKGRNARGIITAGHRGGG alfalfa
 1 MAIHLYKTSTPSTRNGAVDSQVKS^NPNR^NHLIYGQHR^CGKGRNARGIITAGHRGGG soybean
 1 MAIHLYKTSTPSTRNGAVDSQVKS^NPNR^NHLIYGQHR^CGKGRNARGIITAGHRGGG frenchbean
 1 MAIHLYKTSTPSTRN^RAVDSQVKS^NPNR^NLIYGI^HCS^KGRNARGIITAGHRGGG lotus

40 HKRLYRKIDFRRNEK^DIYGRIVTIEYDPNRNAHICLIHYGDGEKRYILHP^KGAI^I chickpea
 45 HKRLYRKIDFRRNEK^DIYGRIVTIEYDPNRNAHICLIHYGDGEKRYILHPRGAI^I alfalfa
 56 HKRLYRKIDFRRNEKNIYGRIVTIEYDPNRN^AYICLIHYGDGEK^KYILHPRGAI^I soybean
 56 HKRLYR^CIDFRRNEKNIYGRIVTIEYDPNRN^ASICLIHYGDGEK^KYILHPRGAI^I frenchbean
 56 HKRLYRKIDFRRNEKNIYGR^IIVTIEYDPNRN^AYICLIHYGDGEKRYILHPRGAI^I lotus

95 GDT^VVY^GTEVPI^KKGNALPLTDMPLGTAIHNIET^EGR^GGGQLARAAGAVAKLIAK chickpea
 100 GDTIV^YGTEVPI^KMGNALPLTDMPLGTAIHNIET^LGKGGQLARAAGAVAKLIAK alfalfa
 111 GDTIVSGTEVPI^KMGNALPLTDMPLGTAIHNIET^LGKGGQLARAAGAVAKLIAK soybean
 111 GDTIVSGTEVPI^KMGNALPLTDMPLGTAIHNIET^LGKGGQLARAAGAVAKLIAK frenchbean
 111 GDTIVSGTEVPI^KMGNALPLTDMPLGTAIHNIET^EGKGG^KLARAAGAVAKLIAK lotus

150 EGKSATLKLPSGEVRLISKNCSATVGQVGNV^EVNQK^LLG^KAGSKRWLGKRPVVRG chickpea
 155 EGKSATLKLPSGEVRLISKNCSATVGQVGNVGNV^NQKSLGRAG^AKRWLGKRPVVRG alfalfa
 166 EGKSATLKLPSGEVRLISKNCSATVGQVGNVGNV^NQK^LGRAGSKCWLKRPVVRG soybean
 166 EGKSATLKLPSGEVRLISKNCSATVGQVGNVGNV^NQK^LGRAGSKCWLKRP^IVRG frenchbean
 166 EGKSATLKLPSGEVRLISKNCSATVGQVGNVGNV^NQKSLGRAGSKCWLKRPVVRG lotus

205 VVMNPVDH^PHGGGEGRAPIGRKK^PSTPWGYPALGRRSRK^SNKYSDNLILRRRTK* chickpea
 210 VVMNPVDH^PHGGGEGRAPIGRKK^PSTPWGYPALGRRSRK^KNKYSDNLILRRRSK* alfalfa
 221 VVMNPVDH^PHGGGEGRAPIGRKK^PATPWG^FPALGRRSRK^RKKYSDNLILRRRTK* soybean
 221 VVMNPVDH^PHGGGEGRAPIGRKK^PATPWG^FPALGRRSRK^RKKYSDNLILRRRTK* frenchbean
 221 VVMNPVDH^PHGGGEGRAPIGRKK^PTPWGYPALGRRSRK^RKKYSDNLILRRRTK* lotus

Fig. 2.

Amino acid alignment of *rps12* protein sequences from chickpea (*Cicer arietinum*) and other representative legumes; alfalfa, (*Medicago sativa*), soybean (*Glycine max*), frenchbean (*Phaseolus vulgaris*), and lotus (*Lotus japonicus*).

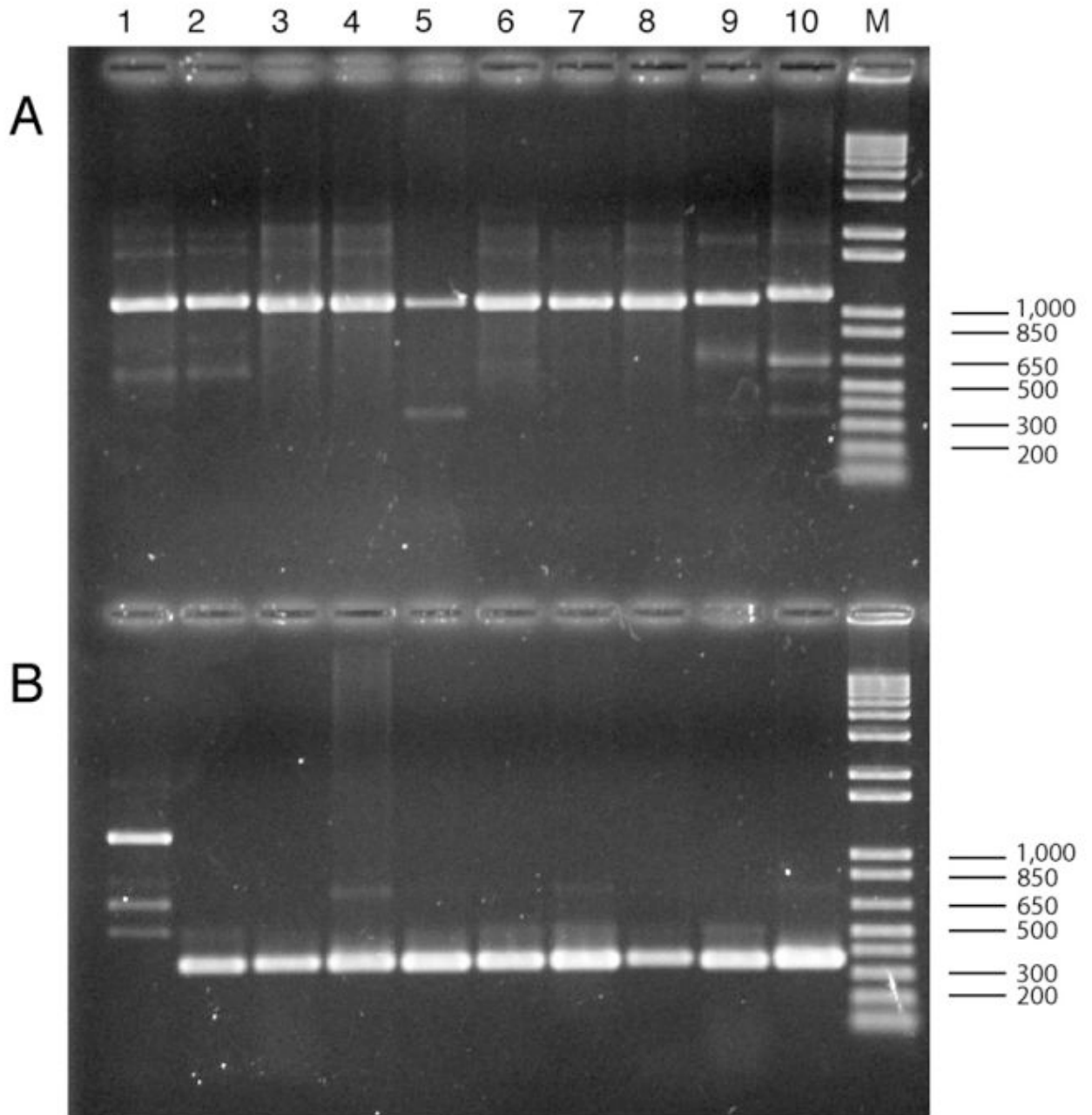


Fig. 3. Ethidium bromide stained 1.0% agarose gel showing PCR amplified products for selected legume taxa screened for the presence/absence of the *clpP* intron. (A) Lanes 1–10: (1) *Lupinus tegeticulatus*, present; (2) *Genista monspessulana*, present; (3) *Glycine max*, present; (4) *Phaseolus vulgaris*, present; (5) *Millettia thonningii*, present; (6) *Desmodium psilocarpum*, present; (7) *Lespedeza cuneata*, present; (8) *Wajira grahamiana*, present; (9) *Lotus purshianus*, present; (10) *Robinia neomexicana*, present. (B) Lanes 1–10: (1) *Sesbania punicea*, present; (2) *Callerya atropurpurea*, absent; (3) *Millettia japonica*, absent; (4) *Wisteria floribunda*, absent; (5) *Glycyrrhiza lepidota*, absent; (6) *Cicer arietinum*, absent; (7) *Medicago*

sativa, absent; (8) *Pisum sativum*, absent; (9) *Caragana arborescens*, absent; (10) *Astragalus canadensis*, absent. M = 1 kb + DNA marker ladder.

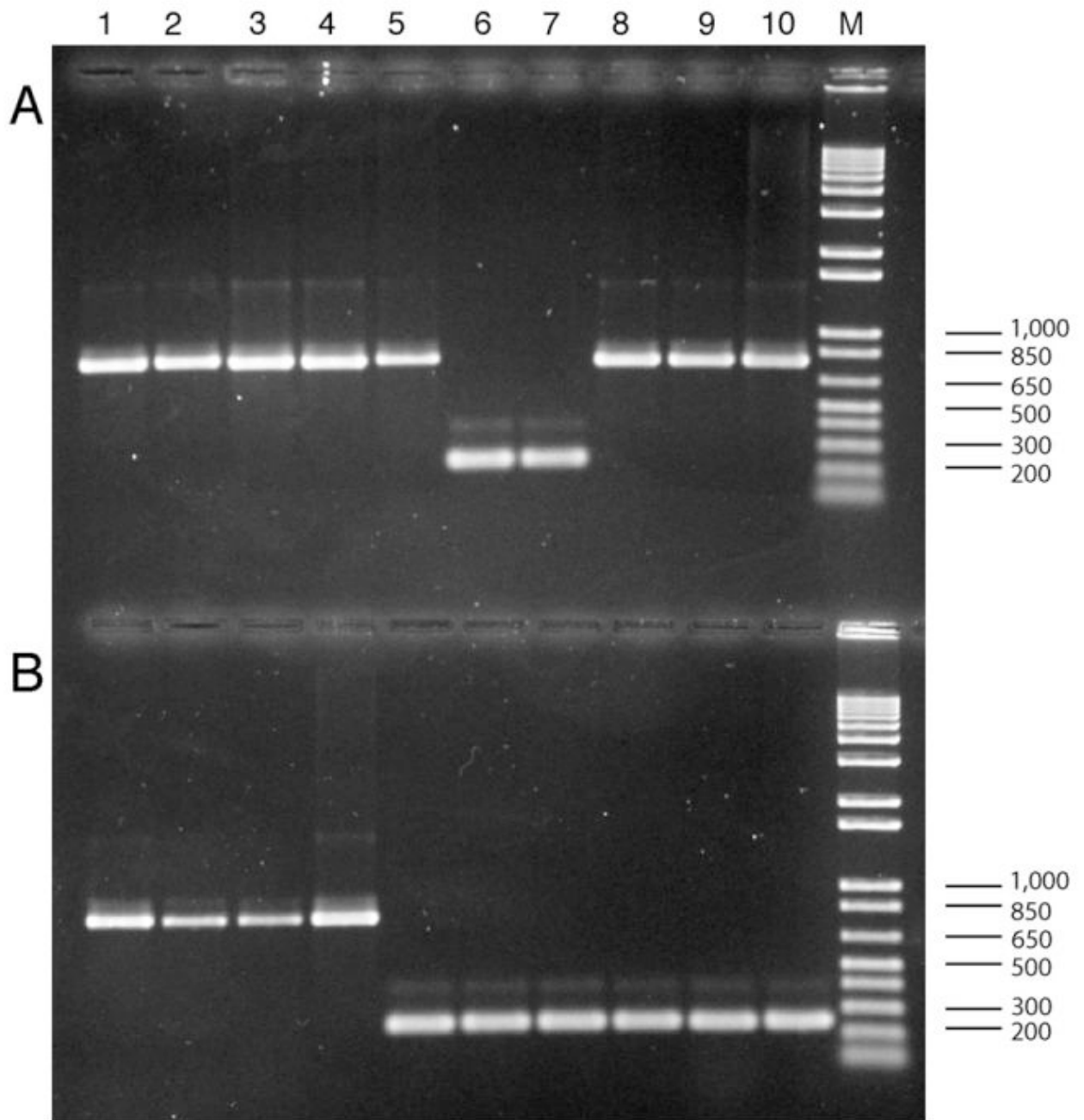


Fig. 4. Ethidium bromide stained 1.0% agarose gel showing PCR amplified products for selected legume taxa screened for the presence/absence of the *rps12* intron. (A) Lanes 1–10: (1) *Lupinus tegeticulatus*, present; (2) *Genista monspessulana*, present; (3) *Glycine max*, present; (4) *Phaseolus vulgaris*, present; (5) *Millettia thonningii*, present; (6) *Desmodium psilocarpum*, absent; (7) *Lespedeza cuneata*, absent; (8) *Wajira grahamiana*, present; (9) *Lotus purshianus*, present; (10) *Robinia neomexicana*, present. (B) Lanes 1–10: (1) *Sesbania punicea*, present; (2) *Callerya atropurpurea*, present; (3) *Millettia japonica*, present; (4) *Wisteria floribunda*, present; (5) *Glycyrrhiza lepidota*, absent; (6) *Cicer arietinum*, absent; (7)

Medicago sativa, absent; (8) *Pisum sativum*, absent; (9) *Caragana arborescens*, absent; (10) *Astragalus canadensis*, absent. M = 1 kb + DNA marker ladder.

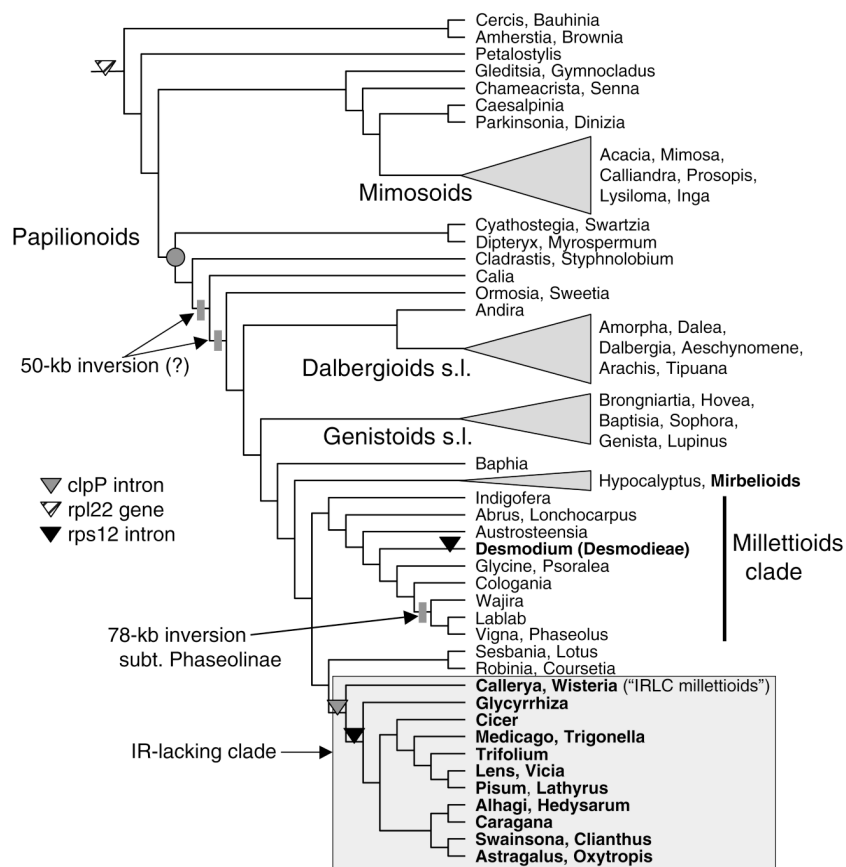


Fig. 5. Simplified representation of phylogenetic relationships among major groups of legumes, summarized from Wojciechowski et al. (2004). Names of informal groups and clades are shown. The phylogenetic distribution of six of nine plastid genomic rearrangements (from Table 2) is depicted on the tree.

Table 1

Survey of *clpP* and *rps12* intron losses among 302 taxa of Fabaceae and Polygalaceae based on PCR analyses of genomic DNA samples from representative legume taxa

TAXON	Collection/accession	<i>clpP</i> intron presence	<i>rps12</i> intron presence
Polygalaceae			
<i>Polygala californica</i> Nutt. ex Torr. & A. Gray	Wojciechowski & Steele 887 (ASU)	Y	Y
Fabaceae			
Caesalpinioideae			
<i>Arcoa gonavensis</i> Urb.	Lavin s.n. (MONT)	Y	Y
<i>Bauhinia tomentosa</i> L.	Wojciechowski 946 (ASU) (DES 199501431001)	Y	Y
<i>Caesalpinia gilliesii</i> Wall. ex Hook.	Wojciechowski 882 (ASU)	Y	Y
<i>Cassia javanica</i> L.	dePompert 48 (ASU)	Y	Y
<i>Ceratonia siliqua</i> L.	Wojciechowski 872 (ASU)	Y	Y
<i>Cercis canadensis</i> var. <i>texensis</i> (S. Watson) M. Hopkins	Fritsch 1465 (CAS)	Y	Y
<i>Chamaecrista fasciculata</i> (Michx.) Greene	Wojciechowski 873 (ASU)	Y	Y
<i>Chamaecrista nictitans</i> var. <i>mensalis</i> (Greenman) H.S. Irwin & Barneby	Wojciechowski 1465 (ASU)	Y	Y
<i>Colophospermum mopane</i> (J. Kirk ex Benth.) J. Léonard	Wojciechowski 947 (ASU) (DES 198301901003)	Y	Y
<i>Conzattia multiflora</i> Standl.	Werling 399 (ASU)	Y	Y
<i>Gleditsia triacanthos</i> L.	Wojciechowski 881 (ASU)	Y	Y
<i>Gymnocladus dioicus</i> K. Koch	Mayer & Mazzio 14545 (ASU)	Y	Y
<i>Gymnocladus chinensis</i> Baill.	P Herendeen 8-V-2003-1 (US)	Y	Y
<i>Haematoxylon brasiletto</i> H. Karst	Wojciechowski 953 (ASU) (cult., DES 197300990102)	Y	Y
<i>Hoffmannseggia glauca</i> (Ortega) Eifert	Wojciechowski 1501 (ASU)	Y	Y
<i>Hymenaea courbaril</i> L.	A Salywon 1264 (ASU)	Y	Y
<i>Parkinsonia aculeata</i> L.	A Salywon 918 (ASU)	Y	Y
<i>Parkinsonia microphylla</i> Torr.	Wojciechowski 1279 (ASU)	Y	Y
<i>Petalostylis labicheoides</i> R. Br.	Wojciechowski 945 (ASU) (cult., DES 1979971414)	Y	Y
<i>Poeppegia procera</i> C. Presl.	EJ Lott 4099 (ASU)	Y	Y
<i>Senna artemisioides</i> (Gaudich. ex DC.) Randell	Wojciechowski 1500 (ASU; cultivated)	Y	Y
<i>Senna covesii</i> (A. Gray) H.S. Irwin & Barneby	Wojciechowski 876 (ASU)	Y	Y
<i>Senna lindheimeriana</i> (Scheele) H.S. Irwin & Barneby	Wojciechowski 1275 (ASU)	Y	Y
Mimosoideae clade			
<i>Acacia greggii</i> A. Gray	Salywon 915 (ASU)	Y	Y
<i>Acacia willardiana</i> Rose	Wojciechowski 952 (ASU)	Y	Y
<i>Acacia hindsi</i> Benth.	Wojciechowski 854 (ASU)	Y	Y
<i>Albizia julibrissin</i> Durazz.	Wojciechowski 908 (ASU)	Y	Y
<i>Calliandra californica</i> Benth.	Wojciechowski 950 (ASU)	Y	Y

TAXON	Collection/accession	<i>clpP</i> intron presence	<i>rps12</i> intron presence
<i>Chloroleucon mangense</i> Britton & Rose	J Rebman 5862 (ASU)	Y	Y
<i>Desmanthus cooleyi</i> (Eaton) Branner & Coville	Wojciechowski 1018 (ASU)	Y	Y
<i>Desmanthus illinoensis</i> MacMill.	Wojciechowski 1171 (ASU)	Y	Y
<i>Dichrostachys cinerea</i> (L.) Wight & Arn.	Wojciechowski 1502 (ASU; cultivated)		
<i>Havardia mexicana</i> Britton & Rose	AL Reina 98–716 (ASU)	Y	Y
<i>Inga punctata</i> Willd.	Landrum 10430 (ASU)	N	Y
<i>Leucaena retusa</i> Benth.	Boke & Massey 419 (UC)	Y	Y
<i>Lysiloma watsonii</i> Rose	A Salywon 921 (ASU)	Y	Y
<i>Mimosa dysocarpa</i> Benth.	Damrel 107 (ASU)	Y	Y
<i>Parkia nitida</i> Miq.	UC 1589048	Y pm?	Y ?
<i>Pentaclethra macroloba</i> Kuntze.	Landrum 10317 (ASU)	Y	Y
<i>Prosopidastrum mexicanum</i> (Dressler) Burkart	J Rebman 4021 (DES)	Y	Y
<i>Prosopis glandulosa</i> Torr.	Wojciechowski 875 (ASU)	Y	Y
Papilionoideae clade (early-branching lineages)			
<i>Andira inermis</i> (Wright) DC.	SB 347 (E)	Y	Y
<i>Andira parviflora</i> Ducke	WR 11179 (INPA, K)	Y	Y
<i>Bobgunnia madagascariensis</i> (Desv.) J.H. Kirkbr. & Wiersema	Smith 1725 (K); DNA Data Bank 8220	Y	Y
<i>Bocoa viridiflora</i> (Ducke) R.S. Cowan	Nascimento (MO); DNA Data Bank 9454	Y	Y
<i>Cladrastis platycarpa</i> (Maxim.) Makino	P Herendeen 1-V-2003-11 (US)	Y†	Y
<i>Calia arizonica</i> (S. Watson) Yakovlev	Salywon 917 (ASU)	Y	Y
<i>Calia secundiflora</i> (Ortega) Yakovlev	Wojciechowski 951 (ASU); cult.	Y	Y
<i>Hymenolobium flavum</i> Kleinh.	RT Pennington 451 (E)	Y	Y
<i>Hymenolobium nitidum</i> Benth.	WR 11177 (K)	Y	Y
<i>Myrocarpus frondosus</i> Allem.	Tressens et al. 3443 (ASU)	Y	Y
<i>Myroxylon balsamum</i> Harms	E Martinez 4051 (ASU)	Y	Y
<i>Pickeringia montana</i> Nuttall	Wojciechowski 883 (ASU)	Y†	Y
<i>Sophora stenophylla</i> A. Gray	RK Gierisch 4997 (ASU)	Y	Y
<i>Sophora nuttalliana</i> B.L. Turner	MA Baker 11465 (ASU)	Y†	Y
<i>Styphnolobium japonicum</i> (L.) Schott	RR Halse 4523 (ASU)	Y†	Y
<i>Styphnolobium conzattii</i> (Standl.) M. Sousa & Rudd	R Torres 5231 (MEXU)	Y†	Y
<i>Styphnolobium montevidis</i> M. Sousa & Rudd (2 access.)	E Bello 3 (MO); L Landrum 10506 (ASU)	Y, Y†	Y, Y
<i>Swartzia jojori</i> Harms	RT Pennington 938 (E)	Y	Y
<i>Swartzia flaemingii</i> Ducke	Ratter 7433 (K)	Y	Y
<i>Sweetia fruticosa</i> Spreng.	RT Pennington 897 (E)	Y	Y
Dalbergioid clade			
<i>Adesmia parvifolia</i> Phil.	Lavin 8269 (MONT)	Y	Y
<i>Adesmia volckmannii</i> Phil.	Lavin 8281 (MONT)	Y	Y

TAXON	Collection/accession	<i>clpP</i> intron presence	<i>rps12</i> intron presence
<i>Aeschynomene pfundii</i> Taub.	Lavin s.n. (MONT)	Y† pm?	Y
<i>Aeschynomene rudis</i> Benth.	DE Fairbrothers et al. 82 (LSU)	Y	Y
<i>Amicia glandulosa</i> H.B. & K.	RT Pennington 654 (E)	Y	Y
<i>Amorpha fruticosa</i> L.	Wojciechowski 1378 (ASU)	Y†	Y
<i>Arachis magna</i> Krapov., W.C. Greg. & C.E. Simpson	CIAT 22248 (MONT)	Y	Y
<i>Brya hirsuta</i> Borhidi	Lavin 7110 (MONT)	Y	Y
<i>Chaetocalyx nigricans</i> Burkart	Vanni 2955 (F)	Y	Y
<i>Dalea pulchra</i> Gentry	Wojciechowski 175 (ARIZ)	Y	Y
<i>Diphysa humilis</i> Oerst. ex Benth. & Oerst.	Haber 1322 (MO)	Y	Y
<i>Diphysa ormocarpoides</i> (Rudd) M. Sousa & R. Antonio	R Torres-C. 977 (MEXU)	Y	Y
<i>Eysenhardtia polystachya</i> Sarg.	ASU 218775	Y	Y
<i>Macherium</i> sp.	CE Hughes 4/89 (FHO)	Y	Y
<i>Marina parryi</i> (Torr. & A. Gray) Barneby	Wojciechowski 176 (ARIZ)	Y	Y
<i>Ormocarpopsis itremoensis</i> DuPuy & Labat	D DuPuy 2363 (K)	Y	Y
<i>Ormocarpopsis parvifolia</i> Dumaz-le-Grand	P Phillipson 3508 (K)	Y	Y
<i>Pictetia sulcata</i> (P. Beauv.) Beyra & Lavin	A Gentry 51037 (MO)	Y pm?	no rxn
<i>Platymiscium</i> sp.	RT Pennington 692 (E)	N pm?	Y
<i>Platypodium elegans</i> Vog. (2 access.)	RT Pennington 488, 688 (E)	Y, Y	Y, Y
<i>Pterocarpus indicus</i> Willd.	RT Pennington 718 (E)	Y	Y
<i>Stylosanthes hamata</i> Taub.	Beyra-M. s.n. (MONT)	Y pm?	Y
<i>Tipuana tipu</i> (Benth.) Kuntze	Lavin 6184 (MONT)	Y	Y
<i>Vatairea macrocarpa</i> Ducke	RT Pennington s.n. (E)	Y	Y
<i>Vatairea</i> sp.	RT Pennington 475 (E)	Y	Y
Genistoid clade			
<i>Anarthrophyllum desideratum</i> Reiche var. <i>mustersi</i> Speg.	RH Fortunato 4971 (ARIZ)	Y	Y
<i>Bowdichia virgilioides</i> Kunth (2 access.)	AM de Carualho 3981 (US); BA Perelra 1110 (US)	Y, Y	Y, Y
<i>Brongniartia peninsularis</i> Rose	J Rebman 4214 (ASU)	Y	Y
<i>Brongniartia ulbrichiana</i> Harms	CE Hughes 2459 (K)	Y	Y
<i>Crotalaria pumila</i> Ortega	Wojciechowski 930 (ASU)	Y	Y
<i>Crotalaria incana</i> L.	N Harriman 16341 (ASU)	Y	Y
<i>Crotalaria pumila</i> Ortega	Wojciechowski 930 (ASU)	Y	Y
<i>Crotalaria saltiana</i> Andrews	CC Chaung 4723 (ASU)	Y	Y
<i>Cyclolobium brasiliense</i> Benth. (2 access.)	Klittgaard & de Lima 86 (K); JRI Wood 17607 (K)	Y, Y	Y, Y
<i>Cytisus scoparius</i> Link	Wojciechowski 1000 (ASU)	Y	Y
<i>Dichilus lebeckioides</i> DC.	A Balsinhas 3464 (K)	Y	Y
<i>Dicraeopetalum mahafaliense</i> (M. Pelt.) Yakovlev	Capuron 28529-SF (TEF)	Y	Y
<i>Diptotropis martiusii</i> Benth.	Beck et al. 166 (US)	Y	Y
<i>Genista monspessulana</i> (L.) L.A.S. Johnson	Wojciechowski 897 (ASU)	Y	Y

TAXON	Collection/accession	<i>clpP</i> intron presence	<i>rps12</i> intron presence
<i>Harpalyce brasiliana</i> Benth.	Ganer 2674 (K)	Y	Y
<i>Lamprolobium fruticosum</i> Benth.	Clarkson & Nelder 8827 (K)	Y	Y
<i>Lebeckia sericea</i> Thunb	JJM van der Meruve 215 (K)	Y	Y
<i>Lupinus consentini</i> Walp.	CE Hughes 1974 (FHO)	Y	Y
<i>Lupinus sericeus</i> Pursh	Wojciechowski 1460 (ASU)	Y	Y
<i>Lupinus tegeticularis</i> var. <i>duranii</i> (Eastw.) Barneby	Wojciechowski 729 (DAV)	Y	Y
<i>Maackia amurensis</i> Rupr.	P Herendeen 1-V-2003-5 (US)	Y	Y
<i>Ormosia formosana</i> Kaneh.	JM Hu 1095 (DAV)	Y	Y
<i>Piptanthus nepalensis</i> Sweet	W Hodgson 10787 (ASU)	Y	Y
<i>Plagiocarpus axillaris</i> Benth.	Barrit 1406 (K)	Y	Y
<i>Poecilanthe falcata</i> (Vell.) Heringer	Klittgaard & de Lima 4 (K)	Y	Y
<i>Poecilanthe subcordata</i> Benth.	Harley 21205 (K)	Y	Y
<i>Spartium junceum</i> L.	Wojciechowski 999 (ASU)	Y	Y
<i>Templetonia retusa</i> (Vent.) R. Br.	Lewis et al. 3833 (K)	Y	Y
<i>Thermopsis rhombifolia</i> Nutt. ex Pursh	Wojciechowski 807 (MONT)	Y	Y
<i>Ulex europaeus</i> L.	D Damrel 2304 (ASU)	Y	Y
Mirbelioid & Indigoferoid clades			
<i>Aotus ericoides</i> G. Don	Wojciechowski 866 (ASU) (UCSCA 81.114)	N?	Y
<i>Bossiaea cordigera</i> Benth. ex Hook. (f.)	Wojciechowski 860 (ASU) (UCSCA 81.47)	Y	Y
<i>Daviesia latifolia</i> R. Br.	Wojciechowski 863 (ASU) (UCSCA 94.379)	Y†	Y
<i>Gastrolobium</i> sp.	MD Crisp 9294 (CANB)	Y	Y
<i>Gompholobium minus</i> Sm.	MD Crisp 9153 (CANB)	Y	Y
<i>Hovea purpurea</i> Sweet	Wojciechowski 869 (ASU)	Y	Y
<i>Hypocalyptus coluteoides</i> (Lam.) R. Dahlgren	C Burman 1236 (CANB)	N?	Y
<i>Indigofera sphaerocarpa</i> A. Gray	Van Devender 94–445 (ASU)	Y	Y
<i>Indigofera</i> sp.	Schrire 2556 (K)	Y?	no rxn
<i>Isotropis foliosa</i> Crisp	MD Crisp 9121 (CANB)	Y	Y
<i>Jacksonia</i> sp.	MD Crisp 9114 (CANB)	Y	Y
<i>Microcharis sessilis</i> (Thulin) Schrire	Thulin & Warfa 6263 (UPS)	Y	Y
<i>Phylloxylon spinosa</i> Du Puy, Labat & Schrire	Schrire 2531 (K)	Y	Y
Millettioids-Phaseoloids clade			
<i>Alysicarpus vaginalis</i> DC.	JB Nelson 702 (ASU)	Y	N
<i>Apios Americana</i> Medik.	LC Higgins 12117 (ASU)	Y	Y
<i>Austrosteensia blackii</i> (F. Muell.) R. Geesink	Pedley 5005 (K)	Y	Y
<i>Bituminaria bituminosa</i> (L.) C. H. Stirton	Soza & Gross 1140 (RSA)	Y	Y
<i>Campylotropis macrocarpa</i> (Bunge) Rehder	Wojciechowski 864 (ASU) (UCBBG 95.0459)	Y	N
<i>Clitoria mariana</i> L.	Wojciechowski & Lavin 1198 (ASU)	Y	Y

TAXON	Collection/accession	<i>clpP</i> intron presence	<i>rps12</i> intron presence
<i>Cologania angustifolia</i> Kunth. (2 access.)	W Hodgson 5573 (ASU); Wojciechowski 1412 (ASU)	no rxn, Y	Y, Y
<i>Cologania longifolia</i> A. Gray	Wojciechowski s.n.	Y	Y
<i>Dalbergiella nyasae</i> Baker f.	Muller 2686 (K)	Y	Y
<i>Derris laxiflora</i> Benth.	JM Hu 1081 (DAV)	Y	Y
<i>Desmodium angustifolium</i> DC.	AL Reina et al. s.n. (ASU)	Y	N
<i>Desmodium batocaulon</i> A. Gray	Wojciechowski 1474 (ASU)	Y	N
<i>Desmodium incanum</i> DC.	RD Worthington 32961 (ASU)	Y	N
<i>Desmodium psilocarpum</i> A. Gray	Wojciechowski 929 (ASU)	Y	N
<i>Desmodium rosei</i> B. G. Schub.	Wojciechowski s.n.	Y	N
<i>Desmodium tortuosum</i> (Sw.) DC.	WM Bush 368 (ASU)	Y	N
<i>Erythrina cristi-galli</i> L.	Wojciechowski 894 (ASU)	Y	Y
<i>Glycine max</i> Merr.	Univ. California Davis (cult.)	Y	Y
<i>Hoita macrostachya</i> (DC.) Rydb.	S Boyd 11472 (ASU)	Y	Y
<i>Kummerowia stipulacea</i> Makino	AR Diamond 15666 (ASU)	Y	N
<i>Kunstleria ridleyi</i> Prain	NN Kat 193 (L) (JM Hu 1230)	Y	Y
<i>Lablab purpureus</i> Sweet	Wojciechowski 1514 (ASU) (USDA 639019)	Y	Y
<i>Lespedeza cuneata</i> G. Don	Wojciechowski & Lavin 1182 (ASU)	Y	N
<i>Lonchocarpus eriocarinalis</i> Micheli	Lavin s.n. (MONT)	Y	Y
<i>Lonchocarpus phaseolifolius</i> Benth.	Hughes 7/89 (FHO)	Y	Y
<i>Macroptilium gibbosifolium</i> (Ortega) A. Delgado	ASU 242502	Y	Y
<i>Millettia leptobotrys</i> Dunn	JM Hu 1164 (DAV)	Y	Y
<i>Millettia thonningii</i> Baker	Lavin s.n. (MONT)	Y	Y
<i>Ostryocarpus stuhlmannii</i> (Taub.) R. Geesink	Corby 2162 (K)	Y	Y
<i>Otholobium glandulosum</i> (L.) J.W. Grimes	D Kelch 028 (UC)	Y	Y
<i>Pedimelum mephiticum</i> (S. Wats.) Rydb.	A Salywon & Wojciechowski 1058 (ASU)	Y	Y
<i>Phaseolus filiformis</i> Benth. (2 access.)	Wojciechowski 1513 (ASU) (DLEG 890043); D Damrel V55 (ASU)	Y, Y	Y, Y
<i>Phaseolus vulgaris</i> L.	Univ. California Davis (cult.)	Y	Y
<i>Philenoptera ericalyx</i> (Harms) Schrire (var. wankiensis)	JM Hu 1090 (DAV)	Y	Y
<i>Piscidia piscipula</i> Sarg.	Lavin & Luckow 5793a (TEX)	Y	Y
<i>Platycyamus regnellii</i> Benth.	HC de Lima 5 (RJ)	Y	Y
<i>Pongamiopsis amygdalina</i> (Baill.) R. Vig.	DuPuy 560 (K)	Y	Y
<i>Psophocarpus tetragonolobus</i> DC.	Wojciechowski 1512 (ASU) (DLEG 910202)	Y	Y
<i>Psoralea argophylla</i> Pursh	Wojciechowski 1436 (ASU)	Y	Y
<i>Pueraria montana</i> (Lour.) Merr.	Wojciechowski & Lavin 1189 (ASU)	Y	Y
<i>Rhynchosia senna</i> Gillies ex Hook. & Arn.	ASU 235516	Y	Y

TAXON	Collection/accession	<i>clpP</i> intron presence	<i>rps12</i> intron presence
<i>Rupertia physodes</i> (Douglas ex Hook.) J.W. Grimes	Wojciechowski & Steele 889 (ASU)	Y	Y
<i>Strophostyles helvola</i> (L.) Elliott	D Cothran 7 (MONT)	Y	Y
<i>Tephrosia leiocarpa</i> A. Gray	DLEG 880028	Y pm?	Y
<i>Tephrosia tenella</i> A. Gray	PD Jenkins s.n. (ARIZ)	Y	Y
<i>Tephrosia virginiana</i> (L.) Pers.	G. Neesom s.n.	Y pm?	Y
<i>Vigna subterranea</i> (L.) Verdc.	Lavin s.n. (MONT) (USDA 241993)	Y	Y
<i>Vigna unguiculata</i> (L.) Walp.	JC Baudet 114 (MONT)	Y	Y
<i>Wajira grahamiana</i> (Wight & Arn.) Thulin & Lavin	Lavin 1623 (MONT)	Y	Y
Loteae, Robinieae, and Sesbania clades			
<i>Anthyllis vulneraria</i> L.	G Allan 45 (RSA)	Y	Y
<i>Coronilla coronata</i> L.	A Mayer 39 (M/MSB)	Y	Y
<i>Coronilla minima</i> L.	D Podlech 54633 (M/MSB)	Y	Y
<i>Coronilla varia</i> L.	SP McLaughlin & JE Bowers 6823 (ARIZ)	Y	Y
<i>Dorycnium pentaphyllum</i> Scop.	F Schuhwerk 92/111 (M/MSB)	Y	Y
<i>Hammatolobium kremerianum</i> C. Muell.	G Allan 43 (RSA)	Y	Y
<i>Hippocrepis glauca</i> Ten.	Hepper 9239 (K)	Y	Y
<i>Hippocrepis unisiliquosa</i> L.	G Allan 55 (RSA)	Y	Y
<i>Lotus corniculatus</i> L.	Wojciechowski 1545 (ASU)	Y	Y
<i>Lotus creticus</i> L.	G Allan 5 (RSA)	Y†	Y
<i>Lotus grandiflorus</i> Greene	Wojciechowski 885 (ASU)	Y	Y
<i>Lotus purshianus</i> Clem. & E.G. Clem.	Wojciechowski 707 (DAV)	Y	Y
<i>Lotus rigidus</i> Greene	Wojciechowski & Sanderson 156 (ARIZ)	Y†	Y
<i>Coursetia glandulosa</i> A. Gray	Wojciechowski 1200 (ASU)	Y	Y
<i>Genistidium dumosum</i> I.M. Johnston	Lavin 210890 (MONT)	Y	Y
<i>Gliricidia maculata</i> (H.B. & K.) Steud.	CE Hughes 675 (FHO)	Y	Y
<i>Hebestigma cubense</i> Urb.	Lavin 7144a (MONT)	Y	Y
<i>Hybosema robustum</i> M. Sousa & Lavin	CE Hughes 92/92 (FHO)	Y	Y
<i>Lennea modesta</i> Standl. & Steyerf.	Lavin & Delgado 8210a (MEXU)	Y†	Y
<i>Olneya tesota</i> A. Gray	Wojciechowski 877 (ASU)	Y?	Y
<i>Peteria thompsonae</i> S. Watson	Wojciechowski, Steele, & Morse 1531 (ASU)	Y	Y
<i>Poissonia heterantha</i> (Griseb.) Lavin	Lavin 5856 (TEX)	Y	Y
<i>Poissonia hypoleuca</i> (Speg.) Lillo	Lavin 5787 (TEX)	Y	Y
<i>Poissonia weberbaueri</i> (Harms) Lavin	Hutchinson 7259 (F)	Y	Y
<i>Poitea glycyphylla</i> (Poiret) Lavin	Lavin 8030-4 (MONT)	Y	Y
<i>Poitea immarginata</i> (C. Wright) Lavin	Lavin 7105 (MONT)	Y	Y
<i>Robinia neomexicana</i> A. Gray	Wojciechowski 717 (ARIZ)	Y†	Y
<i>Robinia pseudoacacia</i> L.	Univ. California Davis, cult.	Y†	Y

TAXON	Collection/accession	<i>clpP</i> intron presence	<i>rps12</i> intron presence
<i>Sphinctospermum constrictum</i> (S. Watson) Rose	Lavin 5120 (MONT)	Y	Y
<i>Sesbania drummondii</i> (Rydb.) Cory	Lavin s.n. (TEX)	Y	Y
<i>Sesbania emerus</i> (Aubl.) Urb.	CE Hughes 80/87 (FHO)	Y	Y
<i>Sesbania punicea</i> Benth.	Wojciechowski & Lavin 1185 (ASU)	Y†	Y
IR-Lacking clade (IRLC)			
<i>Afgekia filipes</i> (Dunn) R. Geesink	JF Maxwell 90–246 (L); (Hu 1231)	N	Y
<i>Afgekia sericea</i> Craib	C Chermisirivanthana 996 (E)	N	Y
<i>Alhagi maurorum</i> Medik.	USDA 502281	N	N
<i>Astragalus americanus</i> (Hook.) M.E. Jones	Wojciechowski 851 (DAV)	N	N
<i>Astragalus canadensis</i> L. var. <i>brevidens</i> (Gand.) Barneby	Wojciechowski & Sanderson 302 (ARIZ)	N	N
<i>Astragalus garbancillo</i> Cav.	CE Hughes 2041 (FHO)	N	N
<i>Astragalus lonchocarpus</i> Torr.	Wojciechowski & Sanderson 143 (ARIZ)	N	N
<i>Astragalus nothoxys</i> A. Gray	Wojciechowski & Sanderson 177 (ARIZ)	N	N
<i>Astragalus pelecinus</i> (L.) Barneby	Wojciechowski & Sanderson 294 (USDA 186284)	N	N
<i>Callerya atropurpurea</i> (Wall.) Schot (2 access.)	OSC 322, OSC 323; Liston	N, N	Y, Y
<i>Callerya australis</i> (Endl.) Schot (2 access.)	OSC 326, OSC 327; Liston	N, N	Y, Y
<i>Callerya megasperma</i> (F. Muell.) Schot (2 access.)	OSC 325, Liston; Wilson 7646 (CANB); J Trusty	N, N	Y, Y
<i>Callerya pilipes</i> (F. M. Bailey) Schot	Gray 08360 (CANB); J Trusty	N	Y
<i>Callerya reticulata</i> (Benth.) Schot (4 access.) (=Milletia reticulata Benth.)	OSC 324, Liston; “HD”, and “OTG25”, J Trusty, cult.; Wojciechowski 1278 (ASU), cult.	N, N, N, N	Y, Y, Y, Y
<i>Calophaca hovenii</i> Schrenk	ID Baitulin et al. s.n. (K) DNA Bank 22172	N	N
<i>Caragana arborescens</i> Lam.	Wojciechowski & Sanderson 413 (ARIZ); USDA 310390	N	N
<i>Caragana pygmaea</i> DC.	Wojciechowski & Lavin 1134 (MONT)	N	N
<i>Chesneya elegans</i> Fomin	M Nydegger 43494 (M/MSB)	N	N
<i>Chesneya parviflora</i> Jaub et. Spach	J Leonard 5840 (K)	N	N
<i>Cicer arietinum</i> L.	Wojciechowski & Sanderson 189 (ARIZ; cult.)	N	N
<i>Cicer canariense</i> A. Santos Guerra & G. P. Lewis	KP Steele 38 (USDA 557453)	N	N
<i>Cicer macracanthum</i> Popow	KP Steele 75 (USDA 599080)	N	N
<i>Cicer pinnatifidum</i> Jaub. & Spach	Wojciechowski 409 (DAV) (USDA 458555)	N	N
<i>Cicer yamashitae</i> Kitam.	KP Steele 71 (USDA 504550)	N	N
<i>Clianthus puniceus</i> Lindl.	OSC 7140; Liston	N	N
<i>Colutea arborescens</i> L.	Wojciechowski & Sanderson 406 (ARIZ)	N	N

TAXON	Collection/accession	<i>clpP</i> intron presence	<i>rps12</i> intron presence
<i>Ebenus cretica</i> L. (2 access.)	R Gadringer et al KR36-1; N Bohling & T Raus 7262 (M/MSB)	N, N	N, N
<i>Ebenus longipes</i> Boiss. & Balansa	M Nydegger 45688 (M/MSB)	N	N
<i>Endosamara racemosa</i> (Roxb.) R. Geesink	JF Maxwell 90–202 (L)	N	Y
<i>Erophaca baetica</i> Boiss. subsp. <i>orientalis</i> (Chater & Meikle) Podlech	JR Edmonson & MAS McClintock 2803 (K), DNA Data Bank 22170	N	N
<i>Galega officinalis</i> L.	USDA 325341	N	N
<i>Glycyrrhiza acanthocarpa</i> J.M. Black	LAS Johnson & Constable 47187 (UC)	N	N
<i>Glycyrrhiza aspera</i> Pall.	D Podlech 31059 (M/MSB)	N	N
<i>Glycyrrhiza astragalina</i> Gill. (2 access.)	G Seijo 1511 (K); Weigand et al. 6894 (M/MSB)	N, N	N, N
<i>Glycyrrhiza glabra</i> L. (2 access.)	A Al-Harasi s.n. (K); J. Lamond 3228 (E)	N, N	N, N
<i>Glycyrrhiza lepidota</i> Pursh	Wojciechowski 714 (DAV)	N	N
<i>Glycyrrhiza pallidiflora</i> Maxim.	(K) DNA Data Bank 22180	N	N
<i>Glycyrrhiza triphylla</i> Fisch. & C.A. Mey.	TF Hewer H.4000 (E)	N	N
<i>Gueldenstaedtia himalaica</i> Baker	BN Starling et al. 241 (K) DNA Bank 22182	N	N
<i>Gueldenstaedtia verna</i> (Georgi) Boriss.	Ulanova & Bassargin s.n. (K) DNA Bank 22183	N	N
<i>Halimodendron halodendron</i> (L.) Voss	Gillis & Good (ASU)	N	N
<i>Hedysarum alpinum</i> L.	Riewe & Marsh 290 (ASU)	N	N
<i>Hedysarum boreale</i> Nutt.	Lavin s.n. (MONT)	N	N
<i>Hedysarum sulphurescens</i> Rydb.	Lavin s.n. (MONT)	N	N
<i>Lathyrus aphaca</i> L.	KP Steele 34 (USDA 286527)	N	N
<i>Lathyrus jepsonii</i> Greene	Pinkava 11905 (ASU)	N	N
<i>Lathyrus latifolius</i> L.	Wojciechowski 543 (DAV)	N	N
<i>Lathyrus odoratus</i> L.	KP Steele 46 (USDA 420254)	N	N
<i>Lathyrus sativus</i> L.	KP Steele 29 (USDA 283562)	N	N
<i>Lens ervoides</i> Grande	KP Steele 67 (USDA 572330)	N	N
<i>Lessertia annularis</i> Burch	H Merxmuller & W Giess (M/MSB)	N	N
<i>Lessertia benguellensis</i> Baker	AE vanWyk 8758 (M/MSB)	N	N
<i>Lessertia herbacea</i> Druce	Wojciechowski & Sanderson 299 (ARIZ); cult.	N	N
<i>Medicago monantha</i> Trautv.	KP Steele 177 (USDA)	N	N
<i>Medicago sativa</i> L.	Wojciechowski 561 (ARIZ)	N	N
<i>Medicago truncatula</i> Gaertn.	Wojciechowski 1014 (ASU); (cult "A17"; D Cook)	N	N
<i>Melilotus indica</i> All.	Wojciechowski 540 (DAV)	N	N
<i>Melilotus officinalis</i> Lam.	Wojciechowski 308 (DAV)	N	N
<i>Millettia japonica</i> A. Gray	Tsugaru 768 (KYO); J Trusty	N	Y
<i>Ononis arvensis</i> L.	KP Steele 32 (USDA 440578)	N	N
<i>Ononis biflora</i> Desf.	KP Steele 41 (USDA 244319)	N	N

TAXON	Collection/accession	<i>clpP</i> intron presence	<i>rps12</i> intron presence
<i>Onobrychis montana</i> DC.	CT Mason & PB Mason 3773 (ARIZ)	N	N
<i>Oxytropis deflexa</i> var. <i>sericea</i> Torr. & A. Gray	Wojciechowski & Sanderson 132 (ARIZ)	N	N
<i>Oxytropis lambertii</i> Pursh	Wojciechowski & Sanderson 155 (ARIZ)	N	N
<i>Parochetus communis</i> Buch.-Ham. ex D. Don	Wojciechowski 901 (ARIZ); cultivated (Liston)	N	N
<i>Pisum sativum</i> L.	Wojciechowski 1015 (ASU); cultivar "Sparkle" (N Weeden)	N	N
<i>Smirnowia turkestanica</i> Bunge	KH Rechinger 51917 (K)	N	N
<i>Sphaerophysa salsula</i> DC.	RR Halse 5170 (ASU)	N	N
<i>Spongiocarpella purpurea</i> (P.C. Li) Yakovlev	BN Starling et al. 146 (K)	N	N
<i>Sulla coronaria</i> (L.) Medik.	USDA 459103 (UC Riverside Ag Garden)	N	N
<i>Sutherlandia frutescens</i> R. Br.	Wojciechowski & Sanderson 266 (ARIZ)	N	N
<i>Swainsona campylantha</i> F. Muell.	JZ Weber 776 (M/MSB)	N	N
<i>Swainsona pterostylis</i> (DC.) Bakh. f.	Wojciechowski & Sanderson 296 (ARIZ)	N	N
<i>Taverniera cuneifolia</i> Arn.	M Thulin 10821 (K) DNA Bank 22569	N	N
<i>Taverniera lappacea</i> DC.	A Radcliffe-Smith 5474 (K) DNA Bank 22568	N	N
<i>Tibetia himalaica</i> (Baker) H.P. Tsui	TN Ho et al. 1867 (E)	N	N
<i>Trifolium dubium</i> Sibth.	KP Steele s.n.	N	N
<i>Trifolium nanum</i> Torr.	CU Boulder course # 4520 collection (ARIZ)	N	N
<i>Trifolium pratense</i> L.	KP Steele 129 (USDA 237713)	N	N
<i>Trifolium repens</i> L.	Wojciechowski 730 (DAV); cult.	N	N
<i>Trigonella foenum-graecum</i> L.	KP Steele 16 (USDA 567879)	N	N
<i>Trigonella kotschyi</i> Fenzl ex Boiss.	KP Steele 64 (USDA 206775)	N	N
<i>Vavilovia formosa</i> (Steven) Federov	Axbegob, Uupzoeba, & Panebapsh s. n. (K)	N	N
<i>Vicia faba</i> L.	Wojciechowski 998 (ASU) (USDA 469175)	N	N
<i>Vicia ludoviciana</i> Nutt. ex Torr. & A. Gray	S McLaughlin & JE Bowers 3185 (ARIZ)	N	N
<i>Vicia narbonensis</i> L.	KP Steele 122 (USDA 294300)	N	N
<i>Wisteria brachybotrys</i> Siebold & Zucc.	K DNA Bank 22664; J Trusty	N	Y
<i>Wisteria floribunda</i> (Willd.) DC.	P Herendeen 1-V-2003-8 (US)	N	Y
<i>Wisteria frutescens</i> (L.) Poir. (2 access.)	AL & HN Moldenke 29243 (ARIZ); USDA 2774	N, N	Y, Y
<i>Wisteria macrostachya</i> Nutt. ex Torr. & A. Gray	W Lathrup 081, cult.; J Trusty	N	Y

TAXON	Collection/accession	<i>clpP</i> intron presence	<i>rps12</i> intron presence
<i>Wisteria sinensis</i> Sweet (2 access.)	K DNA Bank 22082; "WT 20", J Trusty	N, N?	Y, ?

Intron present, "Y"; intron not present (i.e., loss), "N"; †, extra minor bands present; ?, reaction weak or inconclusive; "pm?", polymorphic for intron presence and loss. Clade descriptions follow Wojciechowski et al. (2004).

Table 2

Plastid genomic rearrangements among legumes

Rearrangement	Phylogenetic distribution	Reference
IR-loss	IRLC clade	Lavin et al. (1990), Liston (1995), and Wojciechowski et al. (2000)
50 kb inversion	All papilionoid tribes except Swartzieae, Sophoreae, and Dipterygeae	Doyle et al. (1996) and Pennington et al. (2001)
78 kb inversion	Subtribe Phaseolinae of the papilionoid tribe Phaseoleae	Bruneau et al. (1990)
<i>rpl22</i> gene loss	All legumes	Doyle et al. (1995)
<i>rps16</i> gene loss	15 of 28 papilionoid tribes	Doyle et al. (1995)
<i>ycf4</i> gene loss	15 of 28 papilionoid tribes	Doyle et al. (1995)
<i>clpP</i> intron loss	IRLC papilionoids	Current study
<i>srp12</i> intron loss	<i>Bauhinia</i> (tribe Cercideae), <i>Soemmeringia</i> (tribe Dalbergieae), <i>Mucuna</i> (tribe Phaseoleae), tribe Desmodieae	Doyle et al., (1995), Bailey et al. (1997) and Lai et al. (1997)
<i>rps12</i> intron loss	IRLC papilionoids, tribe Desmodieae	Current study