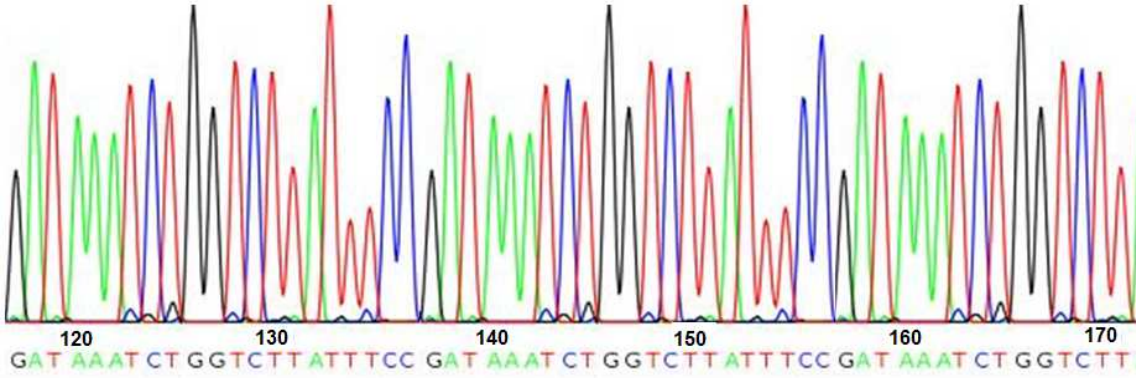
CHAPTER 4



Phylogenetic analysis of Trigonella

using nuclear ribosomal internal transcribed spacer and the plastid trnL-F sequences

4.1 INTRODUCTION

The genus *Trigonella* L. (Fabaceae) belongs to the subfamily Papilionoideae, tribe *Trifolieae*. Taxonomically, the tribe Trifolieae includes four genera, *Trifolium, Medicago, Melilotus and Trigonella.* Species of this group share morphological character states including leaves that are digitately trifoliate with stipules adnate to the stem, but not encircling it entirely. There are some species that appear to be intermediate between all the four closely allied genera (Heyn 1981). Trifolieae is a member of a large clade of legumes lacking one copy of the chloroplast inverted repeat, the IRLC (Lavin *et al.* 1990, Liston 1995). Molecular phylogenetic studies have identified a strongly supported "vicioid clade" within the IRLC comprising tribes Trifolieae and Fabeae (Liston and Wheeler 1994, Sanderson and Wojciechowski 1996, Wojciechowski *et al*. 2000, 2004). Within the "viciod clade", Fabeae and Trifolieae comprise a monophyletic group. Phylogenetic analysis within the "vicioid clade" conducted by Steele and Wojciechowski (2003) strongly supported the monophyly of *Trifolium* but this genus was resolved (with moderate bootstrap support) as a sister lineage to Fabeae, making Trifolieae paraphyletic. A more recent phylogenetic analysis within the "viciod clade" conducted by Ellison *et al* (2006) resolved *Trifolium* as a sister group to *Trigonella+Melilotus* clade but with a weak support. Thus, based on the present results a close relationship of *Trifolium* to other genera in Trifolieae is questioned indicating phylogenetic relationship of genera among Trifolieae still needs to be fully tested.

Analysis of both nuclear and chloroplast sequences from a variety of genes and genic regions strongly supported the monophyly of *Medicago* and *Trifolium* (Watson *et al.* 2000, Bena 2001, Steele and Wojciechowski 2003, Wojciechowski *et al.* 2004, Ellison *et al.* 2006 and Steele *et al.* 2010). "Medicagoid" species described earlier as transition between *Trigonella* and *Medicago* join with very good support to the *Medicago* clade rather than the *Trigonella/Melilotus* clade supporting the morphological based taxonomic transfer of the "medicagoids" *Trigonella* species to the genus *Medicago* (Bena 2001). In all these phylogenetic studies the monophyly of *Trigonella* as delimited by Small (1987-b) was in

question. *Trigonella* was always resolved paraphyletic with respect to *Melilotus* indicating the need for a critical evaluation of the monophyly of the genus.

According to Sinskaya (1961), Hutchinson (1964) and Tutin and Heywood (1964) the genus *Trigonella* contains mostly annual or perennial plants that are often scented. The exact number of species that comprise the genus *Trigonella* has been debated. Petropoulos (2002) indicated that earlier taxonomists like Linnaeus suggested existence of as many as 260 species of *Trigonella*. In contrast, about 128 species of *Trigonella* were reported by Vasil'chenko (1953), 97 by Fazli (1967) and 70 by Hector (1936), Rouk and Mangesha (1963) and Hutchinson (1964). According to species nomenclature in GRIN (GRIN taxonomy 2001) the genus *Trigonella* currently has 37 recognized species including the economically important species *T. foenum-graecum* (fenugreek).

The most elaborate taxonomic classification in *Trigonella* done to date is based on 54 morphological characters (Small 1987-b-b) which delimites the genus into 12 sections (table 3.1). Five of these sections are monotypic; three sections are made of 2 species each while the remaining species are distributed over the remaining 5 sections. Despite the increasing use of molecular markers in phylogenetic and systematic studies (Hillis 1995), the sectional delimitation in *Trigonella* is still based on morphology. The use of seed protein electrophoresis profiles in the taxonomy of *Trigonella* species has indicated that the nomenclature of *Trigonella* should be reassessed (Niknam 2004). A very recent classification based on seed characteristics did not support the sub generic classification proposed by Small (Ceter *et al.* 2012). Since no significant datasets with taxonomic relevance (other than morphological character) is available in *Trigonella*, one of the objectives was to use molecular sequence data to understand the relationships within and between the sections of the genus. Increased resolution of relationship within *Trigonella* will allow greater understanding of the evolution of morphological, biochemical and molecular characters in this genus. The objectives were therefore to use nucleotide sequence data

1. To examine the phylogenetic relationships in *Trigonella*

2. To assess its generic affinities in the tribe *Trifolieae*

For phylogenetic analysis regions from both nuclear and chloroplast genome were selected. Internal transcribed spacer (ITS) region of nuclear ribosomal DNA was chosen because it's utility for examining interspecific relationships in many plant families (reviewed in Baldwin *et al.* 1995) and specifically within Fabaceae (Wojciechowski 2003) has been well documented. The chloroplast *trn*L intron and *trn*L-F region was included because these are universally useful markers for the application in a broad spectrum of phylogenetic questions (Gielly and Taberlet 1994, Kores *et al*. 2001). Moreover, its uniparental, rather than biparental mode of inheritance can facilitate the detection of hybrids. Since combining data from multiple loci is an effective approach for avoiding problems associated with single locus estimates of phylogeny (Rokas *et al.* 2003, Gatesy and Baker 2005, Ellison *et al.* 2006), combined analysis of nrDNA and organelle sequences was carried out to reinforce phylogenetic signals.

4.2 MATERIALS AND METHODS

4.2.1 Plant material

Fifty six accessions representing 22 *Trigonella* species (table 3.1) and 5 *Medicago* species (medicagoids- *M. plicata, M. brachcarpa, M. pamphylica, M. lunata* and *M. rostrata*) were included in the present study. The *Trigonella* species chosen covered 11 of the 12 sections of the genus recognized (Sirjaev, 1929-1934, cited in Small 1987-b). Within *Trigonella*, 6 species (*T. anguina, T. calliceras, T. grandiflora, T. spinosa, T. stellata and T. strangulata*) were represented by a single accession while multiple accessions were present for the remaining 16 species. Species belonging to section *Ellipticae* could not be sampled. Prior to phylogenetic analysis, species identification was confirmed using the morphological characters used for species circumscription in *Trigonella* and generic separation in tribe *Trifolieae* (chapter 2).

4.2.2 DNA extraction and quantification

Total genomic DNA was isolated from leaf tissue as described in section 3.3.2 and quantified as described in section 3.3.3.

4.2.3 Amplification and sequencing

The entire nuclear ribosomal DNA, including both the spacers and the 5.8S cistron was amplified by PCR with primers ITS-F and ITS-R (table 4.1). These primers are distal to ITS4 and ITS5 of White *et al*. (1990, table) by a few base pairs. The amplifications were carried out in GeneAmp PCR system 9700 (Applied Biosystems, Foster city, CA, USA). The thermal cycling conditions were as follows:

Initial denaturation	: 4 min at 94° C
35 cycles of	: 30 s denaturation (94° C)
	: 30s annealing (50° C)
	: 1 min 30s elongation (72 °C)
Final extension	: 5 min at 72 °C

Table 4.1: Primer sequences

ITS		*trn*L-F	
ITS-F	5'CGTAACAAGGTTTCCGTAGGTGAACC3'	c	5' CGAAATCGGTAGACGCTACG3'
ITS-2	5'GCTGCGTTCTTCATCGATGC3'	d	5'GGGGATAGAGGGACTTGAAC3'
ITS-3	5'GCATCGATGAAGAACGCAGC3'	e	5' GGTTCAAGTCCCTCTATCCC3'
ITS-R	5'TTATTGATATGCTTAAACTCAGCGGG3'	f	5'ATTTGAACTGGTGACACGAG3'

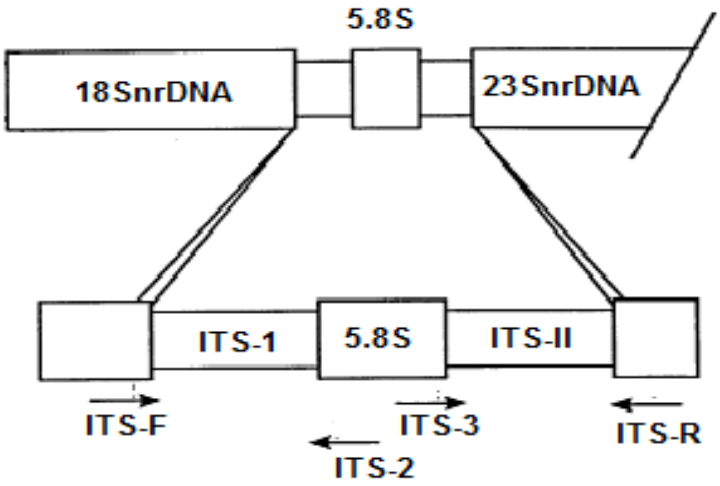


Fig. 4.1: Organization of the nr DNA region. Arrows indicate the approximate position of the primer site

The *trn*L-F region was amplified either as a single fragment (primers c and f) or in two shorter fragments using the primers c/d and e/f in the *trn*L 3' exon (Taberlet *et al*. 1991, table 4.1, Fig. 4.2).

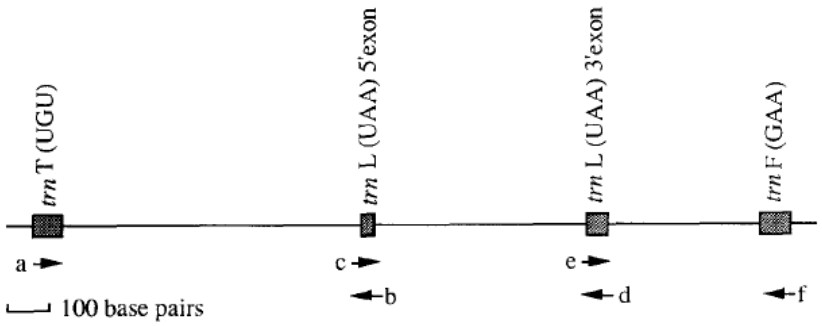


Fig. 4.2: Position and direction of the universal primers used to amplify non-coding regions of chloroplast DNA

The amplifications were carried out in GeneAmp PCR system 9700 (Applied Biosystems, Foster city, CA, USA). The thermal cycling conditions were as follows:

Initial denaturation	: 3 min at 94° C
35 cycles of	: 30 s denaturation (94° C)
	: 45s annealing (50° C)
	: 1 min 30s elongation (72 °C)
Final extension	: 5 min at 72 °C

Amplified products were purified by polyethylene glycol (PEG) precipitation (Sambrook and Russell 2001). Sequencing reactions were carried out directly on purified PCR products using the Big Dye®Terminator v3.1 Cycle sequencing Kit (Applied Biosystems Foster City, CA, USA), with 10-20 ng of template, 1 pmole of primer, 3.5 µl of 5X sequencing buffer, 1.0 µl of Ready Reaction mix in a 20 µl reaction volume.

Cycle sequencing conditions were as follows:

Initial denaturation	:96° C 1 min
25 cycles of	:96° C 10 sec. denaturation
	:50° C 5 sec. annealing
	:60° C 4 min elongation

Sequencing reactions were purified by ethanol/EDTA precipitation according to manufacturer's protocol. Both the regions were sequenced bi directionally using the same primer pair as for amplification. For ITS additional internal primers ITS-2 and ITS-3 were used in some cases (table 4.1). Dried pellets were suspended in 10 µl of Hi-Di Formamide and run on ABI 3100 *Avant* Genetic Analyzer as per the recommended protocol.

Sequences for ITS and *trn*L intron region were obtained from GenBank for species of *Melilotus alba* and *Melilotus officinalis* to be used as out group.

Table 4.2: Accessions, origin, Numbers of outgroup species

Outgroup	**Species**	**origin**	**SA**	**EC**
Medicagoids	*M. brachycarpa**	Turkey	12543	583556
	*M. pamphylica**	Turkey	34530	583562
	*M. rostrata**	Turkey	16856	583611
	*M. lunata**	Turkey	32224	583612
	*M. plicata**	Turkey	5056	583608
			GenBank Number	
		ITS	*trn*L-intron	*trn*L-F
Medicago	*M. sativa*	2995808		83700744
	M. lupulina	85724123		83700741
Melilotus	*M. alba*	85724127	83700746	
	M. officinalis	85724128	83700747	
Trifolium	*T. polyphyllum*	85724278		83700918
	T. lupinaster	85724226		83700860
	T. pseudostriatum	85724282		83700922
	T. acaule	85724132		83700751

*** Reclassified as outlined in chapter 2.**

Similarly, for *Medicago sativa, M. lupulina, Trifolium polyphyllum, T. lupinaster, T. pseudostriaitum and T. acaule* sequences for ITS and *trn*L-F regions were retrieved from GenBank as representatives of *Medicago* and *Trifolium,* respectively (table 4.2).

Two combined analysis were performed: one using the ITS and *trn*L intron data set (since only *trn*L intron region is reported for the two *Melilotus* species used as out group) and the other using the ITS and *trn*L-F data set (without the two species of *Melilotus*).

4.2.4 Phylogenetic analysis

Sequences were aligned using Clustal W (Thompson *et al*. 1994). Only Maximum parsimony analysis was performed for separate ITS-I and ITS-II using PAUP* beta version 4.0b8 (Swofford 2001). Maximum parsimony and maximum likelihood analysis were performed on ITS, *trn*L-F, ITS+*trn*L intron and ITS+*trn*L-F data set using PAUP* beta version 4.0b8. Gaps were treated as missing data. All parsimony analysis were simultaneous and unconstrained (Nixon and Carpenter 1996) with character state changes unordered and weighed equally. Analysis was conducted using an initial heuristic search comprising 1000 replicates of random stepwise addition using tree bisection and reconnection (TBR) branch swapping with MULTREES option on, but saving only 1 trees per replicate. Multiple most parsimonious trees resulting from this analysis were used to compute a strict consensus tree, which was then used as a constraint for another round of heuristic searches. The consistency index, CI (Kluge and Farris 1969) and retention index, RI (Farris 1989) were calculated. The robustness of the clade in the strict consensus tree was evaluated by non-parametric bootstrap analysis (Felsenstein 1985) and by computing decay values (Bremer 1994). The following general descriptions for categories of bootstrap support were used: poor <50%, weak, 50-74%, moderate, 75-84%, strong, 85-100% (Chase *et al*. 2000). Decay indices were obtained using the programs PAUP* and TreeRot.v2 (Sorenson 1999) and clades having DI greater than or equal to 4 were considered well supported (Marcilla *et al*. 2001). A suitable model for sequence

evolution was selected using jModel Test (version 0.11, Posada 2008) for ITS, *trn*L-F and combined data set. Once an explicit model of sequence evolution was selected, parameters were adjusted to those estimated during model testing. A final tree topology was then estimated in a new maximum-likelihood search implemented in PAUP* with TBR branch swapping. Due to the prohibitive time required for ML analysis only 100 replicates were performed.

The ITS region used in the present study is a part of a tandem repeat within the diploid nuclear region, while the *trn*L-F, is part of the haploid chloroplast genome. Because the modes of inheritance are different for these two genomes, biparental vs. uniparental, their evolutionary histories are not linked (Doyle 1992; Moore 1995). Thus, there is no reason to assume that the resulting gene tree will be identical. Possible conflicts between the two data sets were evaluated with an incongruence length difference test (ILD) (Farris *et al.* 1994, 1995) prior to combining the data. This test, implemented as the partition homogeneity test in PAUP* (Swofford 2001), determines whether the original data partitions differ significantly from randomly shuffled partitions of the combined data sets. One hundred replicates were performed on parsimony informative characters using TBR branch swapping, simple sequence addition, MULTREE on, Steepest Descent in effect, and MaxTrees set at 100. The trees obtained for each gene region were also examined for “hard” or “soft” incongruences (Seelanan *et al.* 1997) and the data was combined following the suggestion of Liu and Miyamoto (1999). This approach has also been used to increase the phylogenetic resolution in other plant groups (Cameron *et al*. 2002) including *Trifolium* (Ellison *et al.* 2006).

Bayesian analysis of the separate and the combined data sets were conducted with MrBayes 3.0b4 (Huelsenbeck *et al*. 2002; Ronquist and Huelsenbeck, 2003). The best fit model of sequence evolution was chosen using the hierarchical Likelihood Ratio test (hLRT) and Alkali Information Criterion (AIC), calculated with Mr. Model Test 2.3 (Pasoda and Crandall 1998). These models were applied to their respective partitions in the separate and combined analysis. In each analysis, a single run of 3,000,000 generations were conducted. In each

run trees were sampled every 100 generation and burn-in was determined by inspection of the log-likelihoods of the sample trees. Branch length information was recorded and averaged across all retained trees, and majority rule consensus tree were computed to obtain posterior probabilities (PP). Clades with > 0.90 posterior probabilities were considered strongly supported.

4.3 RESULTS

4.3.1 Amplification and sequencing

A total of 56 accessions were sequenced for both nrDNA and *trn*L-F regions (table 4.3). Clean ITS sequence could not be obtained for *T. spinosa* and hence this species was excluded from the ITS and the combined analysis. All sequences obtained were included in a series of parsimony analysis. In all species, multiple accessions of a given species clustered together, and a single accession was chosen for inclusion in the present study (fig. 4.3). The two accessions of *T. cylindracea*, although clustering in the same clade showed divergent ITS and *trn*L-F sequences. Since a single accession could not be determined as a "representative" of the species, both the accessions were included in the phylogenetic analysis.

4.3.2 Intraspecific sequence divergence

ITS sequences have been reported for *T. foenum-graecum T. caerulea, T. cretica, T. arabica, T. calliceras, T. stellata, T. spicata* and *T. kotschyi* (Bena 2001, Kakani *et al.* 2011). The ITS sequences obtained in the present study were compared for nucleotide divergence with those deposited in the GenBank. The nucleotide divergence between sequences of the same species was very low and ranged from 0 for *T. foenum-graecum* and *T. cretica* to 3 nucleotides for *T. arabica.* In the present analysis, *T. stellata* and *T. calliceras* were represented by a single accession. In these species also, the very low nucleotide divergence observed between the two studies supports the inclusion of a "representative" of these species. The intraspecific sequence divergence between accessions of species used in the present study was also very low (ranging from 0 for most

Table 4.3: Length of the sequenced regions

Species	EC No.	nrDNA	trnL (c/d)	trnL e/f)
T. anguina	583495	713	444	220
T. balansae	583507	714	444	220
T. balansae	583508	714	444	220
T. balansae	583509	714	444	220
T. balansae	546586	714	444	220
T. balansae	583543	714	444	220
T. maritima	583600	708	498	220
T. maritima	583601	708	498	220
T. stellata	583621	715	503	214
T. suavissima	583624	713	500	213
T. suavissima	583625	713	500	213
T. calliceras	583570	713	483	188
T. spicata	583614	715	514	215
T. spicata	583616	715	512	209
T. cylindracea	583578	714	407	220
T. cylindracea	583579	714	388	220
T. filipes	583580	714	407	220
T. filipes	583582	714	407	220
T. filipes	583584	714	405	220
T. kotschyi	583597	714	407	219
T. kotschyi	583598	713	407	288
T. kotschyi	583599	714	387	220
T. mesopotamica	583603	714	388	221
T. mesopotamica	583605	714	388	221
T. mesopotamica	583606	714	407	222
T. mesopotamica	583607	714	408	220
T. strangulata	583622	715	512	215
T. cretica	583575	711	509	219
T. cretica	583576	711	509	219
T. cretica	583577	711	509	219
T. caerulea	583567	730	685	208
T. caerulea	583568	725	695	214
T. caerulea	583569	730	690	210
T. arabica	583496	709	341	186
T. arabica	583498	715	341	186

Table 4.3 continued……

Species	EC number	nrDNA	trnL (c/d)	trnL (e/f)
T. schlumbergeri	583509	715	506	186
T. schlumbergeri	583610	715	506	186
T. coelesyriaca	583557	715	513	222
T. coelesyriaca	583562	715	513	222
T. coelesyriaca	583565	715	513	221
T. grandiflora	583595	714	513	221
T. spinosa	583620	-	551	186
T. coerulescens	583573	712	507	218
T. coerulescens	583574	712	507	218
T. foenum-graecum	583588	719	509	218
T. foenum-graecum	583589	719	509	218
T. foenum-graecum	583590	719	509	218
T. foenum-graecum	583591	719	509	218
T. foenum-graecum	583592	719	509	218
T. gladiata	583593	713	511	210
T. gladiata	583594	713	511	212
M. brachycarpa	583556	709	414	205
M. plicata	583608	709	424	163
M. pamphylica	583562	712	424	163
M. rostrata	583611	709	414	205
M. lunata	583612	709	424	150

species to 1.1% in *T. arabica*) with the exception of *T. cylindracea* which showed a sequence variation of 2.5% between the two accessions used (Table 4.4). Within ITS, the low levels of intra-specific difference observed was due to point mutations rather than length variation. Whenever present, intra-individual nucleotide polymorphism observed was associated only with ITS-II (Fig. 4.4)

For *trn*L-F sequences also very low intraspecific sequence divergence was observed ranging from 0% for most species to 1.1% in *T. cylindracea* and this variation was not enough to place the individuals of the same species in different clusters. Such low level of variation is not surprising since we do not expect a complete sequence identity among individuals of the same species. Whenever

present, intra-individual nucleotide polymorphism was associated with length variation resulting from insertions/deletions within the *trn*L-F region (Fig. 4.5)

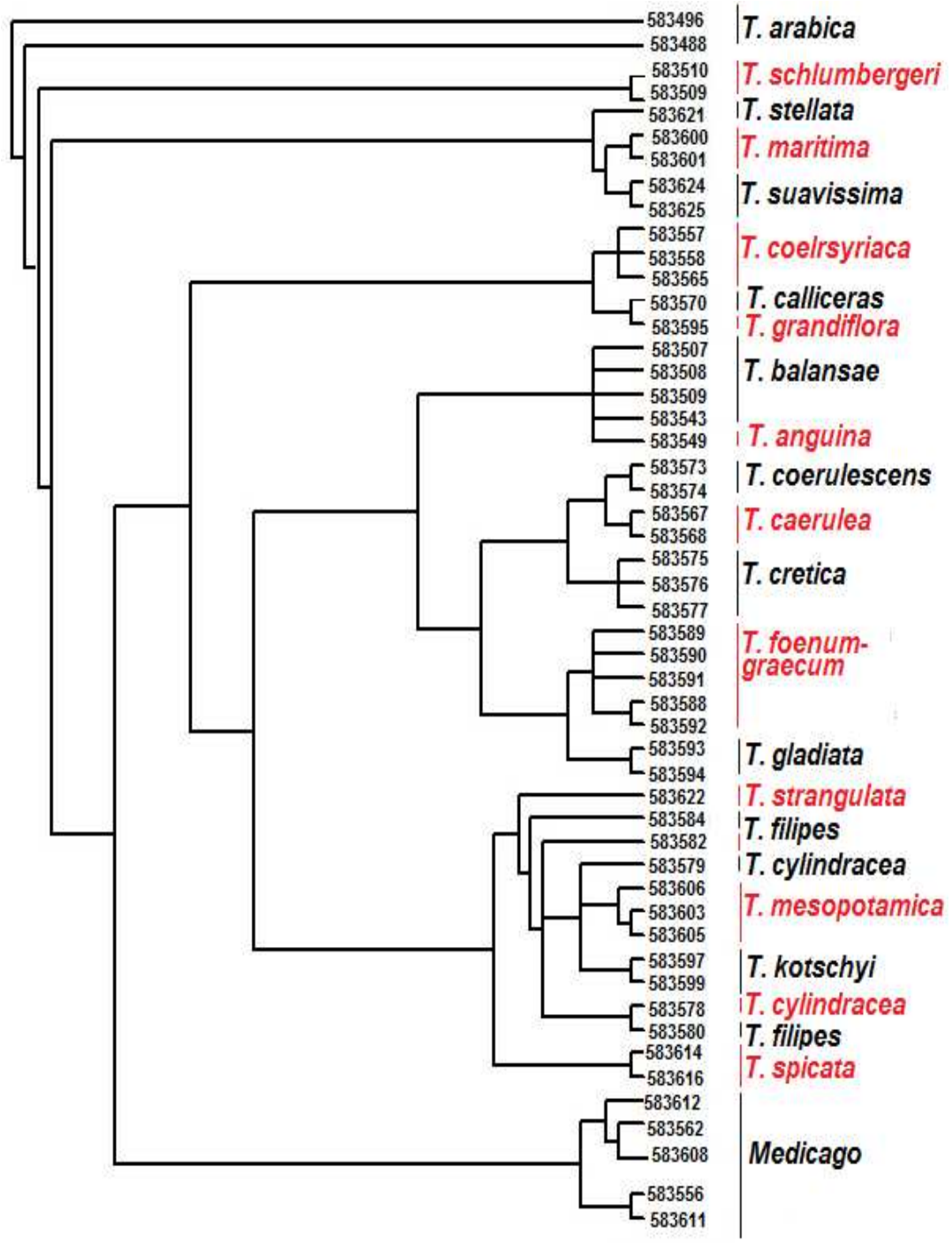
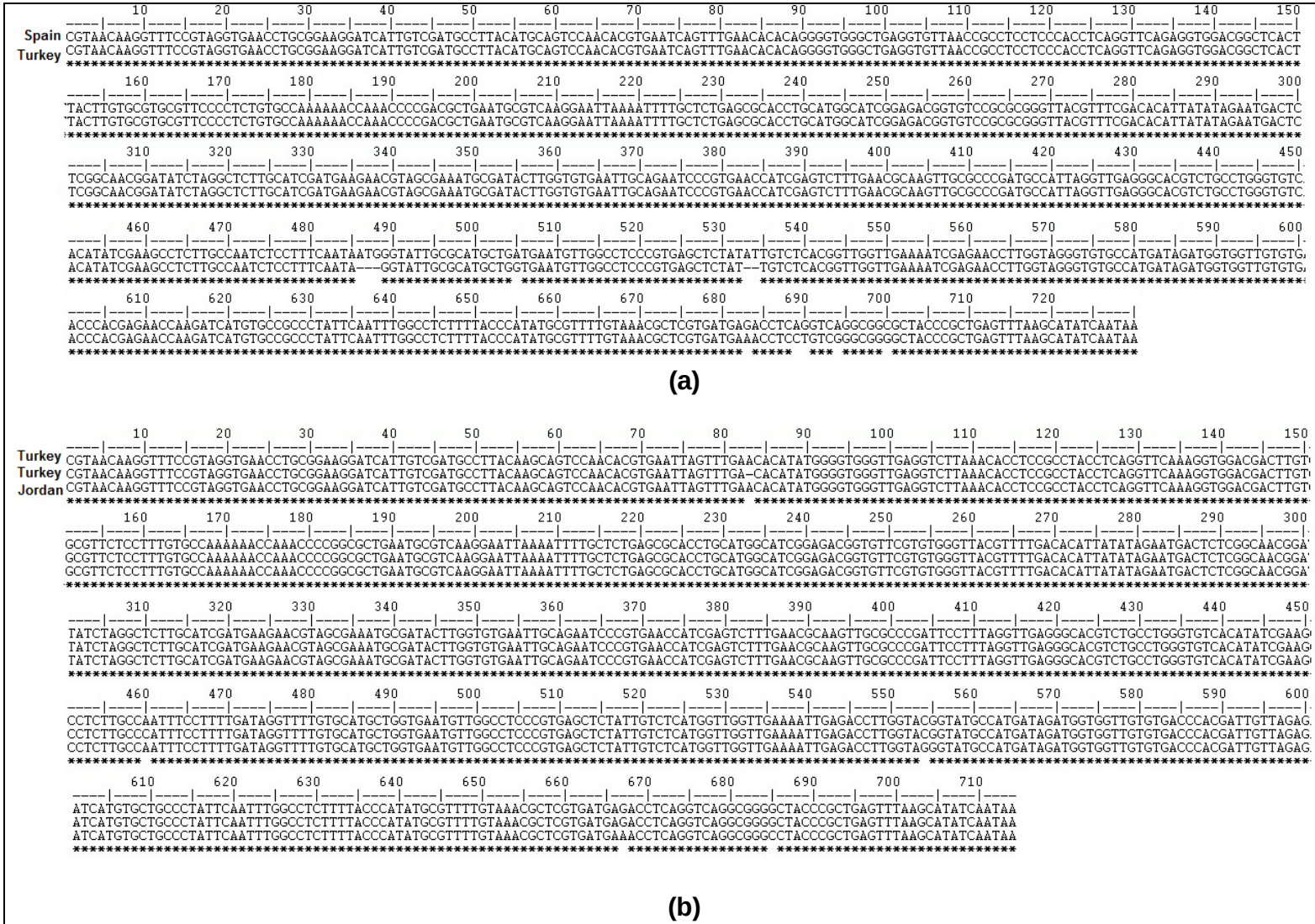


Fig. 4.3: Maximum parsimony analysis of combined ITS + *trn*L-F sequences. Multiple accessions of a given species clustered together.

Table 4.4: Intraspecific sequence divergence in ITS and *trn*L-F sequences

No.	*Trigonella species*	No. of accessions	Sequence divergence (%)	
			ITS	*trn*L-F
1	*T. arabica*	2	1.1	0
2	*T. balansae*	5	0-0.1	0-0.1
3	*T. coelesyriaca*	3	0	0
4	*T. caerulea*	3	0-0.9	0-1.5
5	*T. coerulescens*	2	0	0.4
6	*T. cretica*	3	0	0.1
7	*T. cylindracea*	2	2.5	1.15
8	*T. filipes*	3	0-0.2	0-0.8
9	*T. foenum-graecum*	5	0-0.1	0
10	*T. gladiata*	2	0.4	0.2
11	*T. kotschyi*	3	0-0.7	0.2-1.03
12	*T. maritima*	3	0.2	0.5
13	*T. mesopotamica*	4	0.2-1.4	0-0.4
14	*T. schlumbergeri*	2	0	0
15	*T. spicata*	2	0	0
16	*T. suavissima*	2	0	0
17	*T. calliceras*	1	0	-
18	*T. stellata*	1	0	-
19	*T. grandiflora*	1	-	-
20	*T. strangulata*	1	-	-
21	*T. anguina*	1	-	-
22	*T. spinosa*	1	-	-
	Medicago			
1	*M. brachycarpa*	1	-	-
2	*M. plicata*	1	-	-
3	*M. lunata*	1	-	-
4	*M. rostrata*	1	-	-
5	*M. pamphylica*	1	-	-

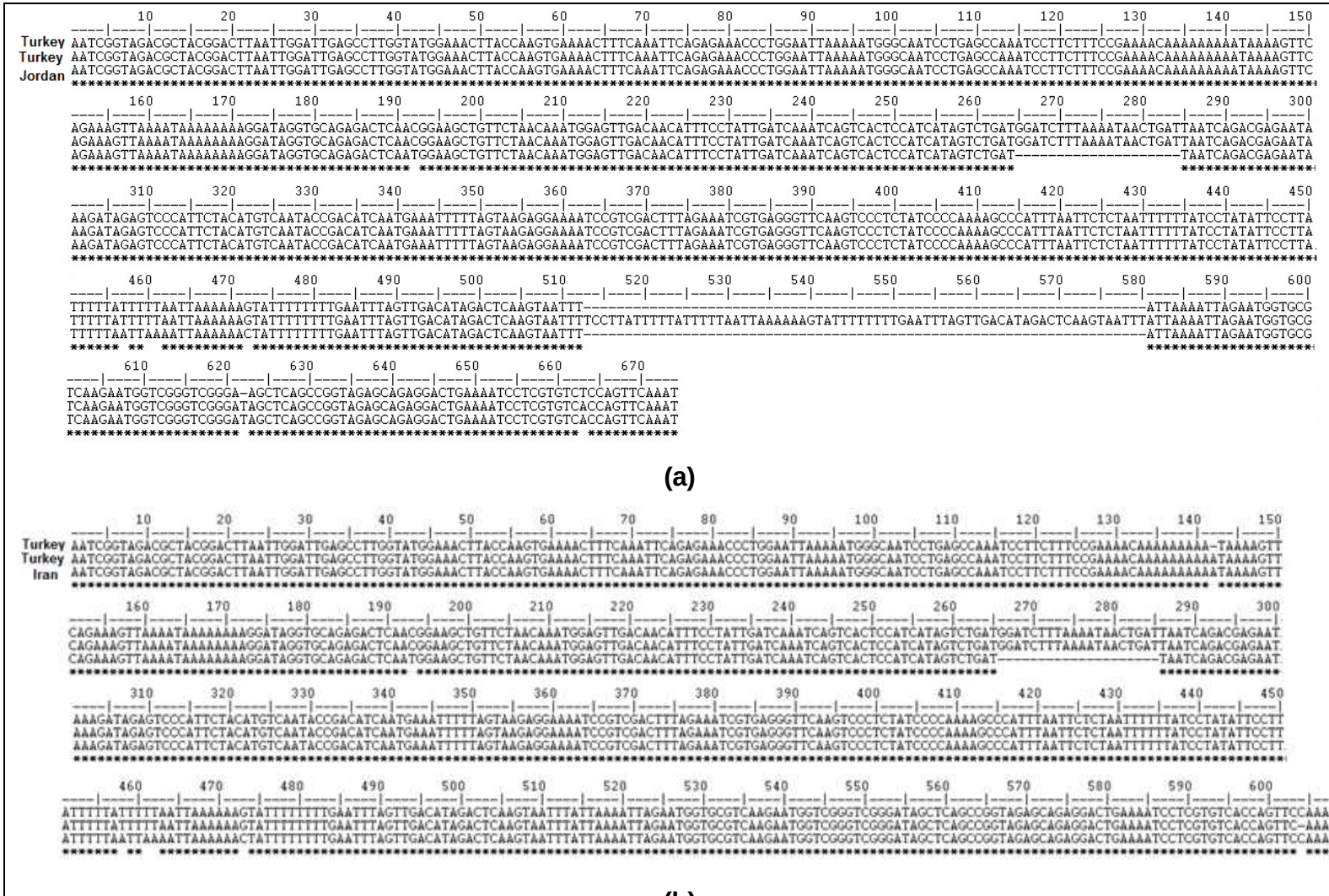


(a)

(b)

Fig. 4.4: Intraspecific nucleotide sequence divergence observed in ITS region.

(a) Accessions of *T. caerulea,* EC 583569 (Spain) and EC 583568 (Turkey) and (b) *T. kotschyi* EC 583597, EC 583598 (Turkey) and EC 583599 (Jordan)



(a)

(b)

Fig. 4.5: Intraspecific nucleotide sequence divergence observed in *trn*L-F region.
(a) Accessions of *T. kotschyi* EC 583597, 583598(Turkey) and EC 583599 (Jordan) and (b) *T. mesopotamica* EC 583606 and EC 583607 (Turkey) and EC 583605 (Iran)

4.3.3 Sequence characteristics

The length and the composition of each gene region sequenced, as well as the tree statistics from separate and combined analysis are summarized in Table 4.5.

Table 4.5: Sequence statistics for separate and combined internal transcribed spacer (ITS), and chloroplast *trn*L-F data sets used in the phylogenetic analysis

Description	ITS-I	ITS-II	ITS	*trn*L-F	ITS+*trn*L-F
Number of taxa included-ingroup	21	21	21	22	22
Number of taxa included-outgroup Included	34	34	34	35	33
Length range (bp)-ingroup	268-253	248-240	609-730	506-715	1307-1421
Aligned length (bp)-ingroup	245	237	698	466	1166
Aligned length(bp)-outgroup included	237	267	674	252	997
G+C content mean (%)- ingroup	67.5	68	48.6	34.8	40.0
G+C content mean (%)- outgroup included	68.1	69.6	48.7	34.6	42.1
Parsimony uninformative sites-ingroup	49	34	87	39	123
Parsimony uninformative sites-outgroup included	45	27	207	13	104
Potentially informative characters-ingroup	40	38	79	14	91
Potentially informative characters-outgroup included	64	60	127	10	220
CI of MPTs	0.71	0.68	0.69	0.92	0.77
RI of MPTs	0.85	0.82	0.83	0.89	0.83
Number of MPTs	20	1000	112	44	108
Length of MPTs	193	170	383	28	511

CI- Consistence Index; RI- Retention Index; MPTs- Maximum Parsimonious Tree

Nucleotide sequences are being deposited in GeneBank. Within *Trigonella,* (excluding the out group) the ITS-1 region varied from 248 for *T. caerulea* to 233 for *T. maritima*. Of the 237 aligned characters 45 were variable and parsimony uninformative and 64 were potentially informative for parsimony analysis. ITS-II varied from 226 for *T. caerulea* to 218 for *T. stellata, T. coelesyriaca, T. maritima* and *T. suavissima.* Of the 220 aligned characters 27 were variable and 60 were potentially informative for parsimony analysis. ITS-1 was longer as compared to ITS-II and within the limits reported by Baldwin (1992) for Fabaceae and Bena (2001) for *Trigonella*. Individual sequences across the ITS region varied from 609-730bp. For the final alignment, 94 bps corresponding to regions of ambiguous alignment were removed from the data set. This resulted in a final data set of 674 aligned characters of which 207 were variable but uninformative and 127 were potentially parsimony informative. The mean GC content of the ITS region was 48.7%. The diversity of the ITS sequences among species of *Trigonella, Trifolium, Melilotus and Medicago* was due more to nucleotide substitution variation than to insertions and deletions (indels).

The *trnL*-F region varied in length from 506 to 715 among the 35 taxa. The aligned sequences however included only 252 sites, of which 13 were variable but uninformative and 10 were potentially parsimony informative. The mean GC content of the *trn*L-F region was 35.4%.The *trn*L-F data set was characterized by numerous indels (gaps from one to 90 nucleotides). The combined ITS and *trn*L sequences included 862 aligned sites among the 35 taxa. Of these, 96 were variable but uninformative and 157 were potentially parsimony informative. The mean GC content was 42.1%. Similarly, for the combined ITS + *trn*L-F sequences of the 997 aligned sites among the 33 terminal taxa 104 were variable but uninformative and 220 were potentially parsimony informative. The mean GC content was 42.1%.

4.3.4 Phylogenetic Analysis

4.3.4.1 ITS analysis

The strict consensus of the trees obtained from parsimony analysis of ITS-I and ITS-II is presented in Fig. 4.6 and 4.7 with bootstrap values provided.

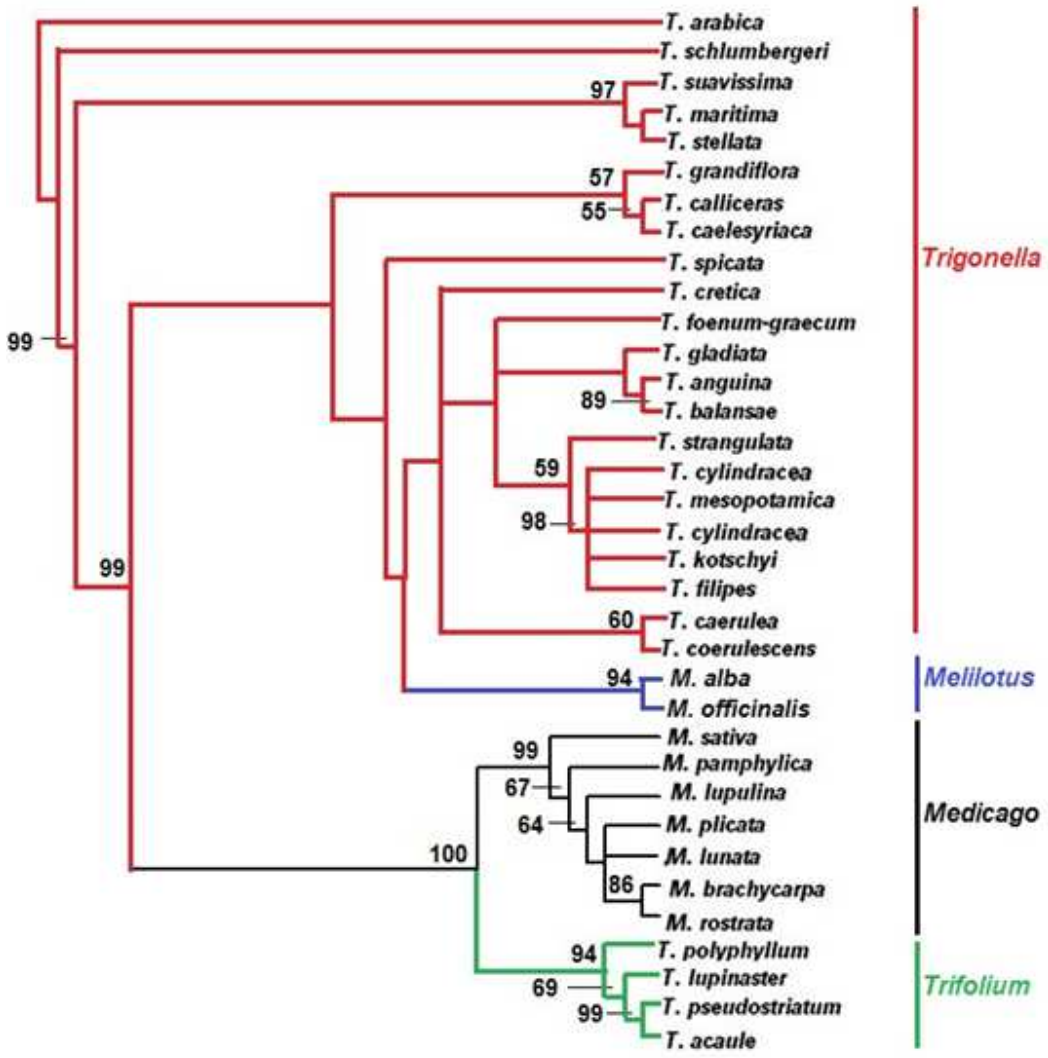


Fig. 4.6: Phylogenetic relationship in *Trigonella* based on the maximum parsimony analysis of ITS-I sequences. Tree shown is the strict consensus of 20 most parsimonious trees (CI=0.71, RI= 0.85, HI= 0.2850) from heuristic search analysis of 237 sequences. Numbers along the branch indicate bootstrap percentage above 50%.

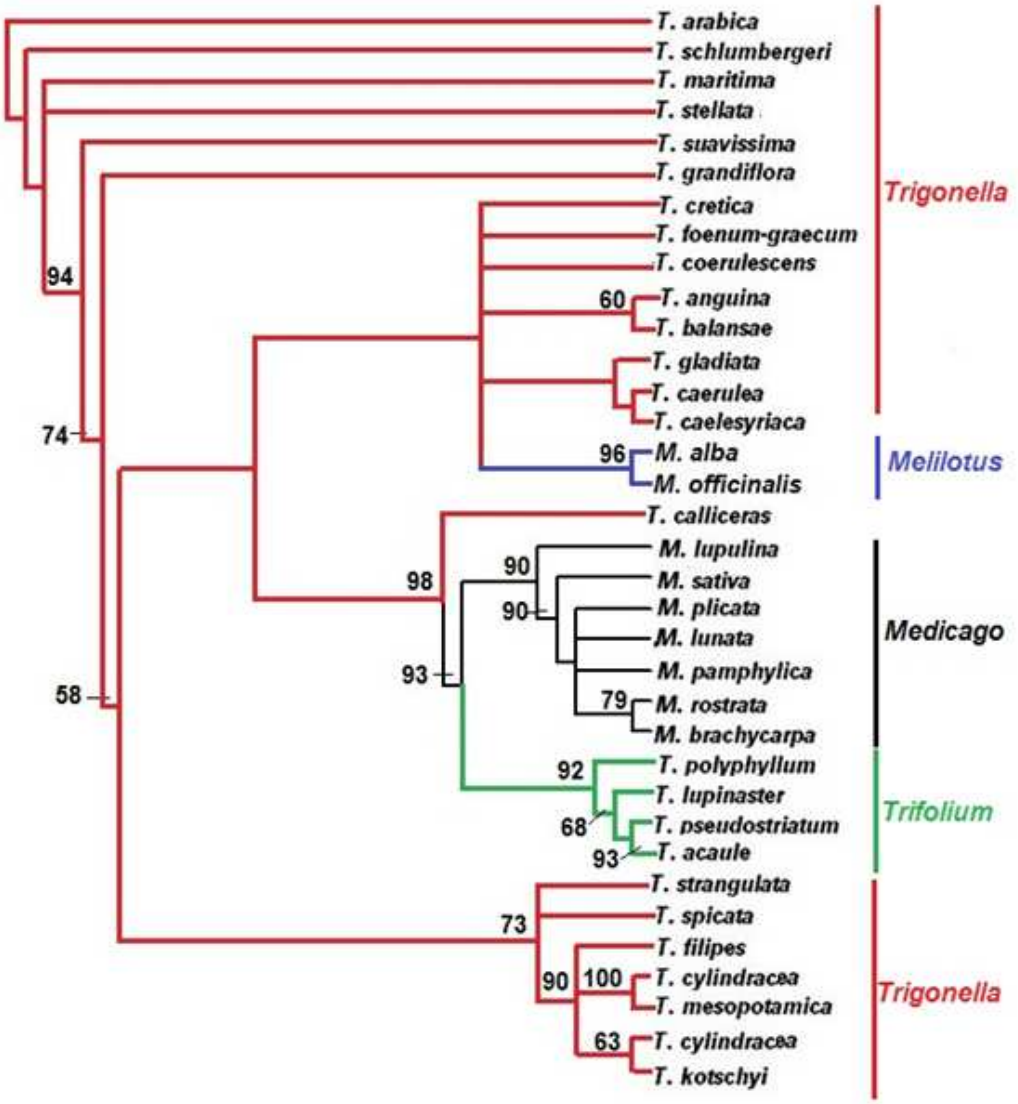


Fig. 4.7: Phylogenetic relationship in *Trigonella* based on the maximum parsimony analysis of ITS-II sequences. Tree shown is the strict consensus of 1000 most parsimonious trees (CI=0.680, RI= 0.820, HI= 0.3118) from heuristic search analysis of 267 sequences. Numbers along the branch indicate bootstrap percentage above 50%.

Parsimony analysis of the ITS data set resulted in 112 equally parsimonious trees of length 383 (CI=0.694, RI= 0.830, HI= 0.305). The strict consensus of these trees is presented in Fig. 4.9 with bootstrap (BP) and decay values (DI) provided. Using jmodel test the GTR+G model of sequence evolution with a discrete gamma rate distribution was selected for ITS data set. Maximum

likelihood (ML) analysis using these model parameters resulted in a single maximum likelihood tree of score –ln L 2133.26612.

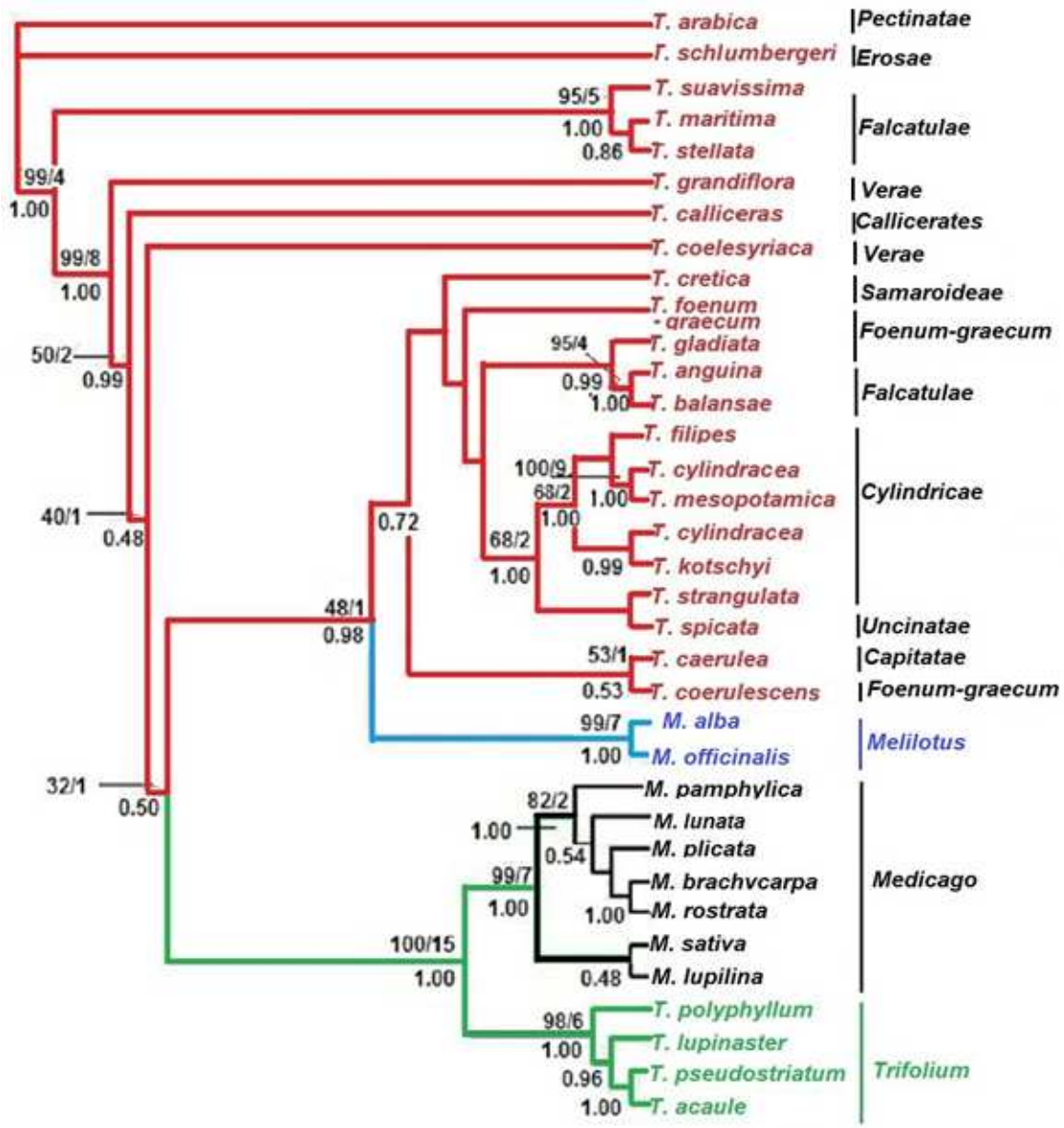


Fig. 4.8: Phylogenetic relationship in *Trigonella* based on the maximum parsimony analysis of ITS sequences. Tree shown is the strict consensus of 112 most parsimonious trees (CI=0.694, RI= 0.830, HI= 0.305) from heuristic search analysis of 674 sequences. Numbers along the branch indicate bootstrap percentage above 50% followed by the decay index. Numbers below the branch indicate Bayesian posterior probabilities. Section of *Trigonella* according to Small (1987-b) are indicated on the right.

The tree generated from the ML analysis of ITS data was highly congruent with the topology of the strict consensus from the parsimony analysis. The Bayesian

analysis of the ITS data set using an AIC selected SYM+G substitution model generated trees with a topology highly similar to that produced with MP and ML analysis (Fig. 4.8). The Bayesian posterior probability values (PP) and the ML, MP bootstrap values (BP) are well correlated with the PP values consistently higher. In the MP and ML analysis 12 clades have a BP≥ 90%, DI ≥4 while 3 clades have BP≥ 70%, DI ≥ 1. Analysis of the ITS sequence data confirms the monophyly of *Medicago* and *Trifolium* and the clustering of medicagoids with the monophyletic *Medicago* with a BP value of 100% and DI of 7 (Bena 2001). The 2 species of *Melilotus* form a part of a basal polytome within the clade of *Trigonella* species.

Within *Trigonella*, section *Falcatulae* is rendered paraphyletic by the position of the strongly supported *T. balansae* and *T. anguina* cluster (BP=95, DI=4) outside a strongly supported clade (BP=95%, DI=5) comprised of the remaining three representative of this section, *T. maritima, T, stellata and T. suavissima*. Section *Cylindricae* is rendered paraphyletic by the moderately supported sister group relationship of *T. strangulata* with a strongly supported clade consisting of *T. cylindracea,* T*. filipes,* T*. kotschyi* and *T. mesopotamica*, with the section *Uncinatae* represented by *T. spicata* as a sister group. There is a strong bootstrap support for *T. arabica* (section *Pectinatae*, Bp =94%, DI=4) as a sister group to *T. schlumbergeri* (section *Erosae*) and the strongly supported clade comprising of *T. maritima, T, stellata and T. suavissima.* Remaining species within *Trigonella* are largely unresolved with little BP support. In the tree depicted (Fig. 4.8) *T. foenum-graecum* is sister to clade comprising *T. gladiata* (section *Foenum-graecum*), with a weak support. *T. coerulescens*, the other representative from section *Foenum-graecum* is sister to *T. caerulea* (BP=53%). Despite the lack of resolution in this part of the tree, our result indicates that section *Foenum-graecum* is paraphyletic. The monophyly of section *Callicerates* (*T. calliceras*), *Uncinatae (T. spicata*), *Samaroideae* (*T. cretica*) and *Capitatae* (*T. caerulea*) is weakly supported. Section *Verae* is rendered paraphyletic by the separate position of *T. grandiflora* (BP=99%, DI=8) and *T. coelesyriaca* (BP=40%).

4.3.4.2 *trn*L-F analysis

Parsimony analysis of the *trn*L-F data set resulted in 44 equally parsimonious trees of length 28 (CI=0.92, RI= 0.89, HI= 0.07).

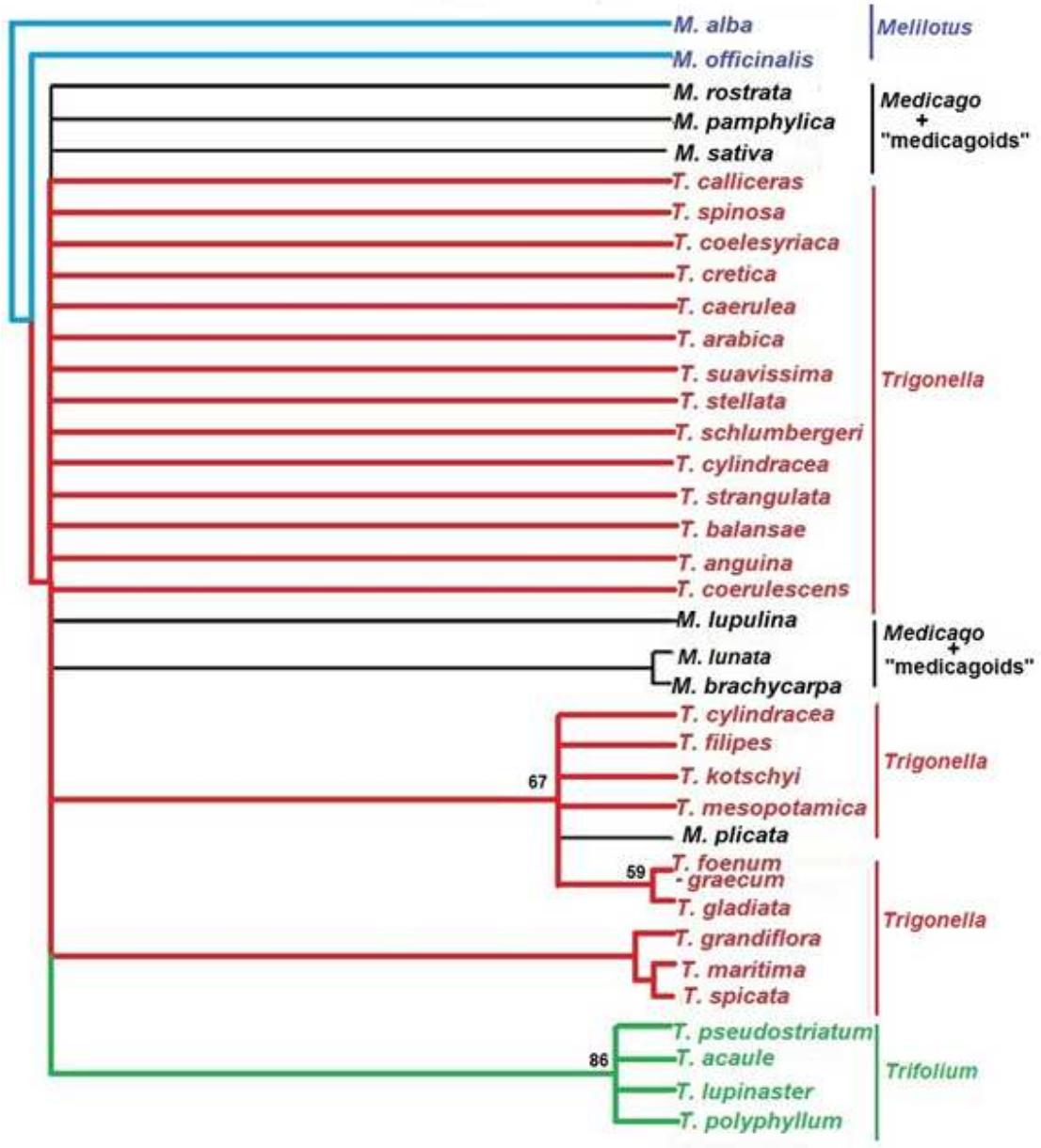


Fig.4.9: Phylogenetic relationship in *Trigonella* based on the maximum parsimony analysis of *trn*L-F sequences. Tree shown is the strict consensus of 44 most parsimonious trees (CI=0.920, RI= 0.890, HI= 0.305) from heuristic search analysis of 252 sequences. Numbers along the branch indicate bootstrap percentage above 50%.

Using jmodel test the GTR+G model of sequence evolution with a discrete gamma rate distribution was selected for *trn*L-F data set. Maximum likelihood (ML) analysis using these model parameters resulted in a single maximum likelihood tree of score –ln L 453.5424. Because of the low number of parsimony informative characters, the *trn*L-F data set resulted in trees that are poorly resolved and weakly supported (Fig. 4.9). The phylogenetic reconstruction confirms the monophyly of only *Trifolium.* However comparison of the poorly resolved *trn*L-F with ITS tree showed no strongly supported topological incongruence.

4.3.4.3 Combined analysis

Based on the ILD test, the two partitions are significantly different (p=0.01, table 4.2). Removal of non *Trigonella* sequences, where several topological discrepancies were observed, still resulted in significantly different partitions (p=0.01). Incongruences between ITS and *trn*L-F data have been reported previously in subtribe Trigonellinae (Ellison *et al*. 2006). Lack of resolution should not be interpreted to be lack of evidence for combining data (Cunningham 1997a and b); however it may simply be evidence of insufficient information and signals. The later may be the case with the *trn*L-F data set, in which there is an obvious deficit of discrete characters suitable for parsimony analysis. Because each data set showed no strongly supported conflicting groups, ITS and *trn*L -F data sets were analyzed simultaneously.

Parsimony analysis of the combined data set (ITS+ *trn*L) resulted in 328 equally parsimonious trees of length 448 (CI=0.70, RI= 0.820, HI= 0.381). The strict consensus of these trees is presented in Fig. 4.10 with bootstrap (BP) and decay values (DI) provided. Using jmodeltest the TrN+G+I model of sequence evolution with a discrete gamma rate distribution was selected for the combined ITS+ *trn*L data set. Maximum likelihood (ML) analysis using these model parameters resulted in a single maximum likelihood tree of score –ln L 2698.6895. The tree generated from the ML analysis of the combined data is highly congruent with the topology of the strict consensus from the parsimony analysis. In the MP and ML analysis 14 clades have a BP$\geq$ 90%, DI $\geq$4 while 3 clades have BP$\geq$ 70%, DI $\geq$ 1.

The topology of the Bayesian tree is in conflict with the strict consensus tree for MP analysis of the combined data with respect to the position of *T. cretica*. Apart from the position of this species in the parsimony strict consensus tree, which is poorly supported difference; the MP strict consensus tree and the Bayesian tree are consistent.

Table 4.6: Results of partition-homogeneity test

Sum of tree lengths	Number of replicates
5888*	1
6386	1
6396	1
6398	1
6400	2
6401	3
6402	3
6403	1
6404	3
6405	4
6406	5
6407	4
6408	4
6409	6
6410	6
6411	9
6412	7
6413	7
6414	8
6415	4
6416	9
6417	5
6418	3
6419	3

* = sum of lengths for original partition
P value = 1 - (99/100) = 0.010000

Partition-homogeneity test with heuristic search; Character partition = genes; Starting seed = 1; Number of replicates = 100; Optimality criterion = parsimony; Character-status summary: Of 1214 total characters- 182 characters are constant; 43 variable characters are parsimony-uninformative and 989 characters are parsimony-informative.

Combined analysis resolved *Trifolium* and *Trigonella* + *Melilotus* of Trifolieae as sister groups. The monophyly of *Trigonella* and *Melilotus* resolved in the combined analysis was not resolved by the separate ITS and *trn*L-F data sets.

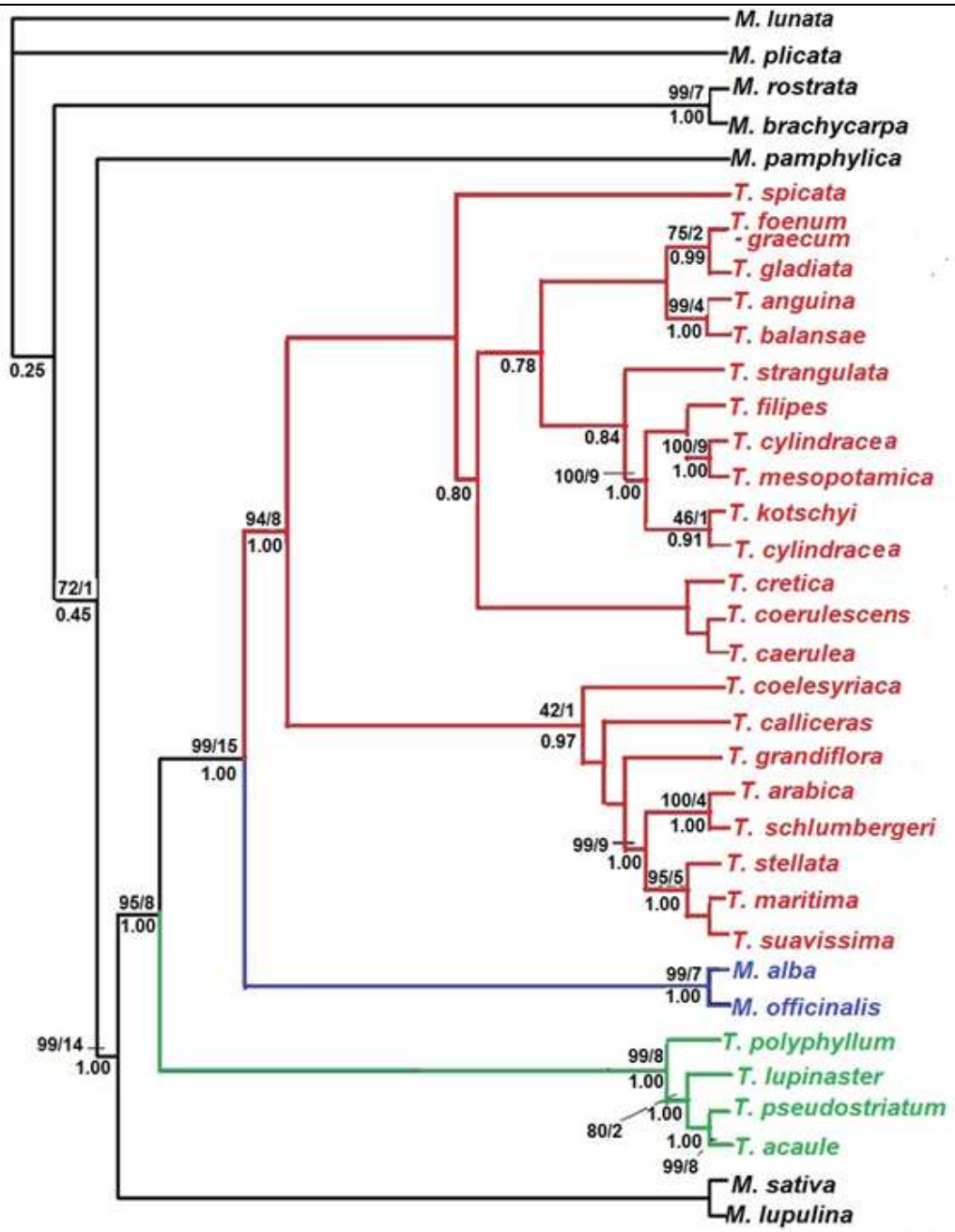


Fig. 4.10: Phylogenetic relationship in *Trigonella* based on the maximum parsimony analysis of ITS + *trn*L intron sequences. Tree shown is the strict consensus of 328 most parsimonious trees (CI=0.70, RI= 0.820, HI= 0.381) from heuristic search analysis of 862 sequences. Numbers along the branch indicate bootstrap percentage above 50% followed by the decay index. Numbers below the branch indicate Bayesian posterior probabilities.

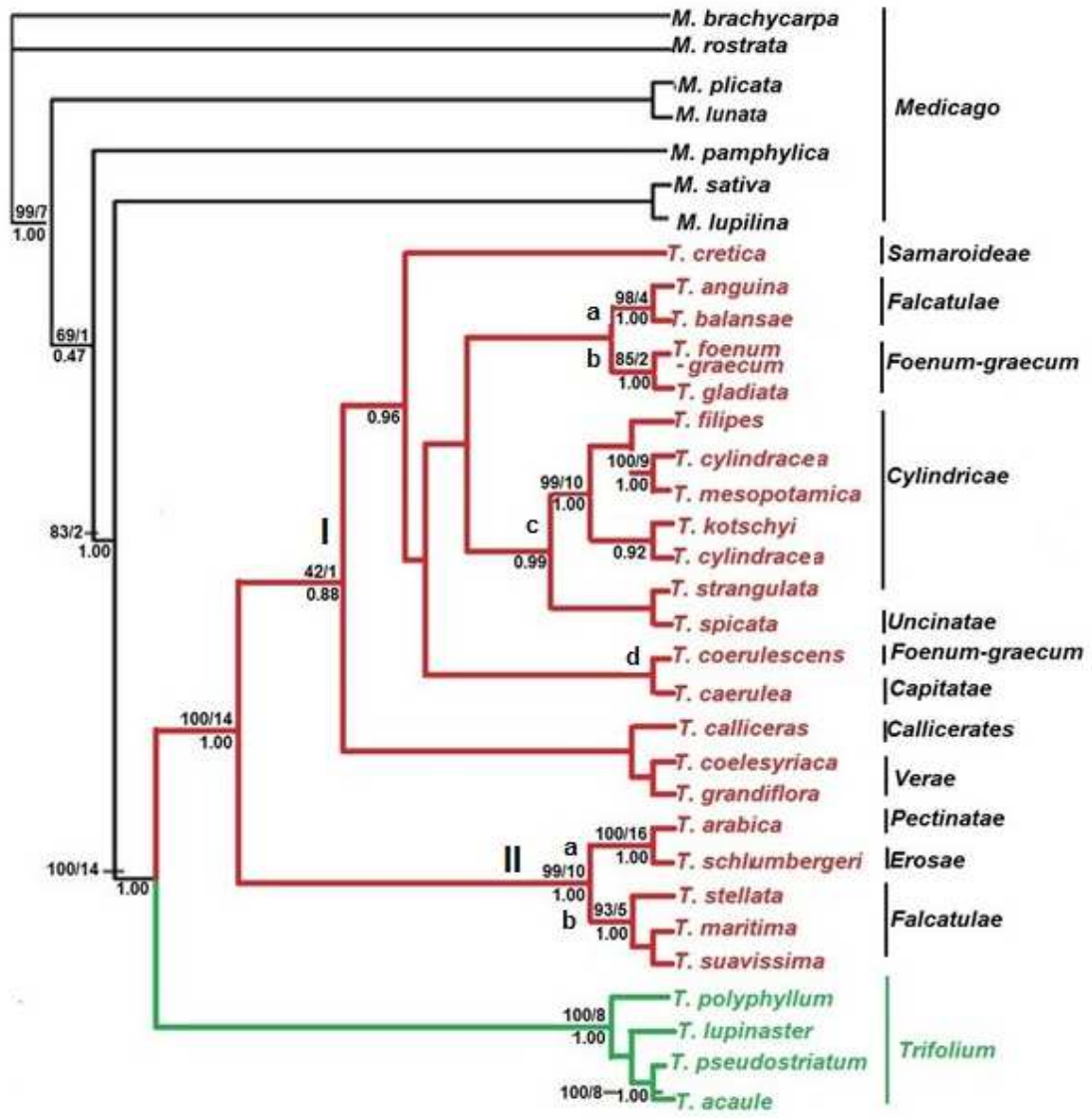


Fig.4.11: Phylogenetic relationship in *Trigonella* based on the maximum parsimony analysis of ITS+trn*L-F* sequences. Tree shown is the strict consensus of 108 most parsimonious trees (CI=0.77, RI= 0.873, HI= 0.291) from heuristic search analysis of 997 sequences. Numbers along the branch indicate bootstrap percentage above 50% followed by the decay index. Numbers below the branch indicate Bayesian posterior probabilities. Section of *Trigonella* according to Small (1987-b) are indicated on the right.

Parsimony analysis of the combined ITS+ *trn*L-F (without *Melilotus*) resulted in 108 equally parsimonious trees of length 511 (CI=0.77, RI= 0.873, HI= 0.291). The strict consensus of these trees is presented in Fig. 4.11 with bootstrap (BP) and decay values (DI) provided. Using jmodeltest the TrN+G model of sequence

evolution was selected for the combined ITS+ *trn*L-F data set. Maximum likelihood (ML) analysis using these model parameters resulted in a single maximum likelihood tree of score –ln L 2698.6895. The tree generated from the ML and MP analysis of the combined data is highly congruent with the topology of the strict consensus from the parsimony analysis. The Bayesian analysis of the combined data set using an AIC selected GTR+G+I substitution model for both the regions generated trees with a topology highly similar to that produced with MP. The topology of the Bayesian tree is in conflict with the strict consensus tree for MP analysis of the combined data with respect to the position of *T. spicata and T. strangulata*. Apart from the position of these species in the parsimony strict consensus tree, which is poorly supported difference; the MP strict consensus tree and the Bayesian tree are consistent.

Despite the conflicting signals in the 2 data sets, the combined analysis was better resolved and the BP support increased for some of the clades. The ITS sequences made a much greater contribution than the *trn*L-F sequences.

The topology of the combined analysis indicates that *Trigonella* is monophyletic and consists of two major lineages referred to as Clade I and II.

- Clade I: This clade is only weakly supported but contains well-supported subgroups referred to as subclade a-d. Section *Falcatulae* is rendered paraphyletic by the position of the strongly supported *T. balansae* and *T. anguina* cluster (subclade 1-a, BP=98, DI=4) outside a strongly supported subclade (BP=96%, DI=5) comprised of the remaining three representative of this section, *T. maritima, T, stellata and T. suavissima*. Unlike ITS, in the combined data set *T. foenum-graecum* forms a strongly supported clade (BP=86%, DI=2, subclade 1-b) with *T. gladiata* (section *Foenum-graecum*). However, section *Foenum-graecum* is again rendered paraphyletic by the position of *T. coerulescens* outside the clade comprising of the remaining two representative of this section. Like ITS, section *Cylindricae* is rendered paraphyletic with the clustering of *T. strangulata* with *T. spicata* outside the strongly supported clade made of the remaining representative of this section, that is, *T. cylindracea, T.*

filipes, T. kotschyi and *T. mesopotamica* (subclade 1-c). Unlike ITS, in the combined data set the monophyly of Section *Verae* is weakly supported by the clustering of *T. grandiflora* and *T. coelesyriaca* in the same clade (clade 1-d). The monophyly of section *Callicerates* (*T. calliceras*), *Samaroideae* (*T. cretica*) and *Capitatae* (*T. caerulea*) is weakly supported in the combined data set.

- Clade II: This well supported clade contains *T. arabica* (section *Pectinatae*) and *T. schlumbergeri* forming a strongly supported subclade (II-a, BP=99%, DI=4) with a sister group relationship with *T. maritima, T, stellata* and *T. suavissima* (II-b)

4.4 DISSCUSSION

4.4.1 Intraspecific sequence divergence

In the genus *Trigonella*, there is an opportunity to compare the ITS results from different studies, that is the present work and that by Bena (2001). It is quite reassuring that for all the seven species used by Bena, we observed a very low divergence among sequences obtained from the same species. The intraspecific variation observed also did not affect the overall phylogenetic position of species since the reconstruction always clustered the accessions of the same species in the same monophyletic group. In all the accessions divergence within the ITS sequence was observed in ITS-II. This suggests that in *Trigonella,* ITS-II changes faster as compared to ITS-I.

Various studies have documented that extensive sequence variation in ITS may arise from ancient duplication events and genomics harboring pseudo genes. Such variation may result in some unexpected placement of species, but this is not enough for their placement in clades of different sections. As seen in Fig 4.8, *T. cylindracea* accessions did not cluster together but were present in the same major clade which had a strong BP support. Thus the variation observed in the 2 accessions of *T. cylindracea* may have been caused by the ITS variation cited above. For all the species included in the current study, for the first time, the *trn*L-F region has been sequenced.

4.4.2 Monophyly of *Trigonella* and *Melilotus*

Various phylogenetic studies in tribe Trifolieae have reported that *Trigonella* is paraphyletic with regards to *Melilotus*. Phylogenetic analysis of tribe Trifolieae and Fabeae based on the sequence of *mat*K gene revealed that *Medicago* and *Trigonella* are sister taxa but *Melilotus* was nested with *Trigonella* (Steele *et al.* 2003). The ITS+ETS combined data of Bena (2001) positioned *Melilotus* as a sister group within the *Trigonella* clade. In the nuclear GA3ox1 sequence analysis *Melilotus* species formed a basal polytome within the clade of all *Trigonella* species (Steele *et al*. 2010). The *trn*K/*mat*K analysis also placed *Melilotus* species as a weakly supported group within *Trigonella* clade (Steele *et al*. 2010). Morphological tree also reflects that *Melilotus* is nested within *Trigonella* and that *Trifolium* is basal in this tree (Magda Gazara *et al.* 2001). The morphological characters examined by Small (1987-b) suggest that *Trigonella* and *Melilotus* are distinguishable on the basis of a combination of characters, but are not discontinuously separated. Independent analysis of ITS and *trn*L also showed that *Melilotus* is nested within the *Trigonella* clade. However, combined ITS and *trn*L data set resolved the monophyly of *Trigonella* and *Melilotus* with a high BP support. Generalized morphological distances of the genera of *Trifolieae* based on 54 morphological characters (Small 1987-b) showed that *Trigonella* is somewhat intermediate between *Medicago* and *Melilotus*. Although sampling within *Melilotus* is limited in the present study, results strongly indicate that the closest relative of *Trigonella* is *Melilotus*. This close relationship between the two genera is supported by a number of characters: stipules margin incised, stander apex notched, style longer than ovary and surface of seed coat smooth. Moreover, some species of *Trigonella* and nearly all species of *Melilotus* release coumarins upon maceration of leaf tissue while species of *Medicago* and *Trifolium* are coumarin negative (Ingham 1981). Similarity in pollen grain morphology of *Trigonella* and *Melilotus* further support their position together (Lashin 2006).

4.4.3 Tribe Trifolieae

Taxonomically, *Medicago* along with *Melilotus* (sweetclovers) and *Trigonella* were included in the tribe Trigonellinae, first recognized by Schultz (1901), but as circumscribed this tribe was not accepted by most taxonomists. Instead, most authors recognized the tribe Trifolieae, which included these three genera and *Trifolium* L. (Rechinger, 1984). The monophyly of *Trifolium* L. is strongly supported in the *mat*K analysis (Steele *et al*. 2010) and is apparent in the supertree which incorporates nrDNA ITS results from many Old World (Watson *et al*. 2000) and New World (Liston *et al*. 2001) species. Surprisingly the genus was resolved (with moderate bootstrap support) as a sister lineage to the Fabeae, making Trifolieae paraphyletic although the position was only weekly supported (Wojciechowski *et al*. 2000, 2004). Based on these results a close relationship of *Trifolium* to other genera in Trifolieae was questioned. A more recent phylogenetic analysis showing the position of *Trifolium* among the genera of the "vicioid clade" (Ellison *et al.* 2006) using combined nr DNA ITS and *trn*L resolved *Trifolium* and *Trigonella*+ *Melilotus* of Trifolieae as sister group. However this relationship was poorly supported. In the present study combined analysis of nr DNA ITS and *trn*L also resolved *Trifolium* and *Trigonella*+ *Melilotus* of Trifolieae as sister group with a high bootstrap support (BP 100, DI 4) in agreement with the traditional classification (Heyn 1981). Results of this analysis support the placement of *Trifolium* within Trifolieae as suggested by Ellison *et al.* (2006).

Medicago and *Trigonella,* as delimited by Small and Jomphe (1989 a and b) have always been strongly supported as sister genera based upon analyses of both the nr DNA ITS and the flanking external transcribed spacer region (nr DNA ETS) (Bena 2001), as well as the plastid-encoded *matK* gene (Steele and Wojciechowski 2003, Wojciechowski *et al*. 2004). However, in the present combined ITS+*trn*L-F analysis *Medicago* is resolved as a sister group to *Trifolium* + *Trigonella* clade. Present results are in accordance with study of Ellison *et al.* (2006). M*edicago* was resolved as a sister group to *Trifolium* and *Trigonella*+ *Melilotus.* Although sampling within *Medicago and Trifolium* is limited in the

present study, results indicate that the closest relative of *Trigonella* after *Melilotus* is *Trifolium*.

4.4.4 Classification

The phylogeny derived from the combined data sets provides strong support for the monophyly of the genus *Trigonella* as delimited by Small (1987). The *Trigonella* species sampled here represent 11 of the 12 sections recognized. Unfortunately, species belonging to section *Ellipticae* could not be included in the analysis. These perennial species are found in the mountains of Afghanistan and nearby areas; even herbarium specimens of these species are rare (Steele and Wojciechowski 2003). Phylogenetic analysis using ITS and *trn*L-F sequences has not been conducted previously in *Trigonella* and the relationship of several species require clarification. The circumscription of sections *Falcatulae*, *Cylindricae*, *Pectinatae, Erosae* and *Foenum-graecum* were not well defined and this suggests that these sections may not be natural groups.

In the separate ITS and combined analysis, section *Falcatulae* is paraphyletic. This second largest section in *Trigonella* forms two separate well supported clades. The first clade strongly clusters *T. balansae* and *T. anguina* (BP=99%, DI=4). Both these species are morphologically similar but readily distinguishable. *T. balansae* has the potential to complement the role of annual medics in alkaline soil farming system especially due to the more expensive seeds of annual medics (Howei *et al.* 2001). This cross pollinated species (Nair *et al.* 2004) is also compatible with *R. meliloti* associated with medic pastures. The present data indicated that the most closely related species to *T. balansae* is *T. anguina* indicating the use of the latter species in breeding programs in crosses with *T. balansae*. The second clade strongly clusters *T. maritima, T. stellata and T. suavissima* (BP=95%, DI=5). Strong incongruences of the phylogenetic analysis with the classification of the species belonging to sections *Falcatulae* suggest that changes in the current circumscription of section *Falcatulae* should be considered.

T. calliceras is the only species in section *Callicerates*. This species was not closely related to species of any other sections, in agreement with its placement

in a separate section*. T. calliceras* was a sister taxon to species of section *Verae.* Although weakly supported the data suggests that section *Callicerates* is more closely related to section *Verae* and possibly not to section *Uncinatae* (*T. spicata*) as proposed by Small.

T. spicata and *T. cephalotes* are two species of section *Uncinatae*. In agreement with morphology, phylogenetic results confirm the sister taxon relationship of *T. spicata* with species of section *Cylindricae*. Principal coordinate analysis by Small (1987-b) showed that *Trigonella* section *Uncinatae* is close to *Melilotus*, as postulated by Sirjaev (1935), a view not supported by the present phylogeny.

Five of the eleven species reported in section *Cylindricae* are sampled. This section is paraphyletic as *T. cylindracea, T. filipes, T. kotschyi* and *T. mesopotamica* form a strongly supported monophyletic clade that does not include *T. strangulata*. For section *Cylindricae* our result does not validate the sectional classification of Small.

T. cretica and *T. graeca* are two species of section *Samaroideae.* Lassen (in Greuter and Raus 1987) assigned the two species of *Trigonella* section *Samaroideae* to *Melilotus*. In cluster analysis by Small (1987-b), this section proved intermediate between the genera, while in the principal coordinate analysis it was closer to *Trigonella* than to *Melilotus*. In the maximum parsimony analysis of *trn*K/*mat*K sequence data *T. cretica* was sister to the clade comprising four *Melilotus* species, *M. alba, M. segetalis, M. sulcatus* and *M. indica*. Our phylogeny indicates that there is strong support for the inclusion of *T. cretica* within *Trigonella* (BP=100%, DI=14, PP=1.00). Although *T. cretica*, *T. balansae* and *T. anguina* showed similar chloroplast haplotypes, their sister taxon relationship is weakly supported within *Trigonella. T. cretica* did not form stable hybrids with *T. foenum-graecum* or *T. caerulea* (Singh 1973). The present phylogeny supports a basal position of *T. cretica* in the genus indicating that it is reproductively isolated and supports its placement in a separate section.

Section *Capitatae* includes two species namely, *T. capitata* and *T. caerulea*. A basal position and close affinity to *Melilotus*, once proposed for section *Capitatae* by Sirjaev, is not supported by the present phylogeny. *T. caerulea* is sister to *T.*

coerulescens (section *Foenum-graecum*) in ITS and the combined data. However, this relationship is only weekly supported by both ITS (BP=53%, DI=1, PP=0.53) and the combined analysis (BP=39%, DI=1, PP=0.52). According to Sirjaev (1935), the section *Foenum-graecum* of the genus *Trigonella* includes eight species which have been arranged in two subsections, *Biberstainianae* and *Gladiatae*. The first subsection includes a single species, *T. coerulescens*, which by its inflorescence and pod shape, seems to be related more to the section *Capitatae*, than to section *Foenum-graecum*. The phylogeny inferred from this study supports a close relationship between *T. caerulea* (from section *Capitatae*) and *T. coerulescens* as postulated by Sirjaev (1935). However, additional molecular data would be required to conclude if *T. coerulescens* should be placed in a new monotypic section within *Trigonella* or included in section *Capitatae*.

T. arabica, placed in monotypic section *Pectinatae*, clusters with a very strong support (BP=100%, DI= 6, PP1.00) with *T. schlumbergeri* (monotypic section *Erosae*) indicating closer relationship between the two species and thereby the sections. These species described earlier as transition between *Trigonella* and *Medicago*, have typical characters of *Trigonella*, such as simple corolla, simple androecium and cotyledonary thickening (Baum 1968). However, they possess neither a typical *Medicago* type pod nor a typical *Trigonella* type pod. In these sections the pods are strongly flattened. Based on morphology both *T. arabica* and *T. schlumbergeri* were placed in the same clade. However, they were placed in separate sections by Small (1987-b) due to difference in pod shape. The present clustering with a high BP support indicated the necessity for reconsideration of present taxonomic placement in different sections and the possible in one section.

Although weakly supported, the monophyly of section *Verae* is congruent with the sectional delimitation of Small (1987-b). Seven species in section *Foenum-graecum* subsection *Gladiatae* are characterized by relatively long and flat pods that contain several seeds. Of these *T. berythea*, considered much closer to cultivated *T. foenum-graecum* and *T. macrorryncha* are apparently endemic only

to South East Turkey (Huber-Morath 1970) while *T. cariensis, T. cassia* and *T. raphanina* are known only from herbarium material (Landizinsky 1979). *T. gladiata* has the widest distribution among the wild forms of section *Foenum-graecum* (Landizinsky 1979). In the separate ITS data, the sister taxa relationship between *T. gladiata* and *T. foenum-graecum* is weakly supported. In the separate *trn*L-F data set these two species clustered with a moderate support (BP=59%). In the combined ITS and *trn*L-F data also, these two species clustered with a high BP support (85%, PP=1.00). The phylogeny inferred from this study strongly supports the view of Sinskaya (1961) that T*. gladiata* and *T. foenum-graecum* share a common ancestral linkage. Only diploid counts, 2n=16 are known for *T. foenum-graecum* suggesting that it may have acquired the chloroplast of *T. gladiata* via cytoplasmic introgression, and not polyploidy speciation. *T. gladiata*, also called as sward fenugreek is a forage crop of great interest as suspected ancestor of cultivated fenugreek. This species is attractive for grazing but is characterized by low forage productivity. However, it possesses high drought resistance and grows on poor soil (Sinskaya 1961). The identification of closely related species for the widely cultivated *T. foenum-graecum* will ensure more efficient use of the wild species germplasm in the improvement of this crop. This is all the more important in the context of its susceptibility to various pests and disease, resulting in low yields (Acharya *et al.* 2010).

4.4.5 Pod character

In distinguishing *Trigonella* species legume characteristics are considered of great taxonomic value. These classification key fails in several instances to group phylogenetically related taxa. For example, recognition of sections *Uncinatae, Samaroideae* and *Capitatae* and their close proximity to *Melilotus* was in part attributed to the presence of one or two seeded indehiscent fruits, a character also involved in the delimitation of genus *Melilotus*. However, the species of section *Uncinatae*, *Samaroideae* and *Capitatae* were found clustering in different positions of the phylogram. Moreover, the close proximity to *Melilotus* reported for these species is also not supported by the present molecular data. Our

phylogeny suggests that legume characters like these are likely to be homoplastic in the genus having occurred independently in more than one lineage within *Trigonella* and *Melilotus*. *T. maritima*, *T. stellata* and *T. suavissima* possess a typical *Trigonella* pod (linear or oblong and curved, Small 1987-b) which in part led to their placement in section *Falcatulae* along with *T. balansae* and *T. anguina.* The separate clustering of *T. maritima*, *T. stellata* and *T. suavissima* and *T. balansae* and *T. anguina* indicate that the similarity in the shape of the pods in these species is due to parallel evolution. Although morphologically similar both *T. arabica* and *T. schlumbergeri* were placed in separate sections on the basis of the different morphological features of pods (Small 1987-b). Although legume characters were drastically different, *T. arabica*, *T. schlumbergeri, T. maritima, T. stellata and T. suavissima* form a strongly supported clade indicating that they share a common ancestral linkage. Our data suggests that although pod characters have been useful to assess infrageneric relationships in *Trigonella* these characters alone are not sufficient to define the sectional circumscription in *Trigonella*.

4.4.6 Comparison of Maximum parsimony vs. Bayesian analysis

Parsimony bootstrap and Bayesian posterior probabilities cannot be considered as equivalent. In agreement with previous reports, the posterior probabilities (PP) were higher than bootstrap support (BP). In all the analysis, clade support was evaluated by considering the bootstrap values because these values exhibit a more even distribution of support value across the clades. However, in all the analysis there were several clades which had little BP support but posterior probability value of >_0.90. Prominent examples in the combined analysis include clade showing sister taxa relationship of *T. strangulata* and *T. spicata* to section *Cylindricae* and the clade showing the position of *T. cretica* in the phylogram. The only example where the BP values were higher than PP values in the combined analysis (ITS+*trn*L and ITS+*trn*L-F) is the clade showing the position of *M. pamphylica*.