Phylogenetic relationships in Solanaceae and related species based on cpDNA sequence from plastid trnE-trnT region

Danila Montewka Melotto-Passarin, Irving Joseph Berger, Keini Dressano, Valentina de Fátima De Martin, Giancarlo Conde Xavier Oliveira, Ralph Bock, and Helaine Carrer

Received 22 November 2007
Accepted 09 February 2008

ABSTRACT - Intergenic spacers of chloroplast DNA (cpDNA) are very useful in phylogenetic and population genetic studies of plant species, to study their potential integration in phylogenetic analysis. The non-coding trnE-trnT intergenic spacer of cpDNA was analyzed to assess the nucleotide sequence polymorphism of 16 Solanaceae species and to estimate its ability to contribute to the resolution of phylogenetic studies of this group. Multiple alignments of DNA sequences of trnE-trnT intergenic spacer made the identification of nucleotide variability in this region possible and the phylogeny was estimated by maximum parsimony and rooted with Convolvulaceae Ipomoea batatas, the most closely related family. Besides, this intergenic spacer was tested for the phylogenetic ability to differentiate taxonomic levels. For this purpose, species from four other families were analyzed and compared with Solanaceae species. Results confirmed polymorphism in the trnE-trnT region at different taxonomic levels.

Key words: intergenic spacer, multiple alignment of DNA, nucleotide polymorphism, taxonomic levels, maximum parsimony.

INTRODUCTION

Chloroplast genes have been extensively used to reconstruct the phylogeny of related species. In order to gain more information for phylogenetic reconstruction many chloroplast regions have been analyzed, including genes and intergenic spacers. Because of the low frequency of structural changes in the chloroplast DNA molecule (cpDNA) together with a low rate of sequence evolution, the plastome (plastid genome) is considered a useful tool for plant phylogenetic studies, especially above the species level (Chiang and Schaal 2000, Ingvarsson et al. 2003, Kelchner 2000, Neves et al. 2005, Odintsova and Yurina 2003). The sequences of the chloroplast rbcL gene (coding the large subunit of Rubisco) as well as a few other genes (i.e. matK, ndhF, psaB and trnL-trnF) have been widely used for inferring phylogeny in plants (Chiang and Schaal 2000, Oxelman et al. 1999, Soltis et al. 2000, Miz et al. 2008). Supposed to evolve more rapidly than coding regions, chloroplast non-coding
sequences such as the intergenic spacer between the trnL (UUA) 3’ exon and the trnF (GAA) gene (Neves et al. 2005, Miz et al. 2008), the atpB-rbcL spacer (Chiang and Schaal 2000, Soltis et al. 2000) and trnC-trnD spacer (Lee and Wen 2004) have been used to address questions concerning relationships among related species or also related genera. Clarkson et al. (2004) studied the phylogenetic relationships in Nicotiana (Solanaceae) inferred from the plastid DNA regions trnL intron, trnL-F spacer, trnS-G spacer and two genes, ndhF and matK.

In some cases the phylogenetic information contained in just one chloroplast genetic marker does not have enough resolving power to separate close taxa, especially at low taxonomic levels. The increase in the number of characters usually improves the resolution of phylogenetic analysis. Since the chloroplast genome does not usually undergo systematic recombination, all characters in different chloroplast markers can be pooled in a single analysis as belonging to the same haplotype.

Data from the study of cpDNA have contributed to resolve some long-standing problems in Solanaceae systematics. The Convolvulaceae are now established as a sister group of Solanaceae; together with Hydroloaeae, Montiniaceae and Sphenocleaceae, these families represent the order Solanales (APG 2003).

Garcia and Olmstead (2003) reconstructed phylogenies of tribe Anthocereideae (Solanaceae) based on sequence variation of two cpDNA markers ndhF and trnL-F. The gene ndhF encodes a subunit of the NADH dehydrogenase complex. Its average substitution rate is almost twice as high as that of rbcL (Soltis and Soltis 1998). Previously it was used to determine phylogenetic relationships in Solanaceae (Bohs and Olmstead 1997). The trnL-F intergenic spacer used as well in previous phylogenetic analyses of other genera in Solanaceae (e.g., Fukuda et al. 2001), includes an intron and spacer, flanking the 39 bp exon of the trnL gene. Because it is mostly non-coding, its substitution rate is higher than that of ndhF.

The Solanaceae family is extremely diverse and widespread containing around 85-90 genera and 1400-1700 species, making it one of the largest flowering plant families (Woodland 1997). Members of the Solanaceae are extremely diverse; in terms of habit, ranging from trees to small annual herbs; in habitat, from deserts to the wettest tropical rain forests; and in morphology, with astounding variation in many characters of both flowers and fruits (Knapp 2001, Knapp 2002). The family contains many common crop plants such as the tomato (Lycopersicon esculentum), aubergine (Solanum melongena, S. aethiopicum), chilli peppers (Capsicum spp.) and husk tomato (Physalis spp., about 100 species), the last perhaps better known as the ornamental garden plant “Chinese lanterns”. Other members such as deadly nightshade (Atropa belladonna), bittersweet (Solanum dulcamara), henbane (Hyoscyamus niger) and thornapple (Datura stramonium) are more renowned for their poisons, whilst some, Petunia spp., Brugmansia spp. (jungle trumpets) and Nicotiana spp. are grown as garden ornamentals. The largest and economically most important genus Solanum is estimated to contain over 1,000 species which, with the exception of the extreme northern and southern latitudes, extend into every continent and climatic region of the world (Hawkes et al. 1991, 1992). This hyper diversity in one genus is unusual in angiosperms, making Solanum interesting from an evolutionary standpoint as well as for its usefulness to humans.

A molecular phylogenetic framework and a provisional reclassification of Solanaceae are now available for the family (Olmstead et al. 1999). Molecular studies have also confirmed that Convolvulaceae represent the sister group of Solanaceae (Soltis et al. 2000). As noted above, tomato (formerly Lycopersicon) is clearly embedded within the large genus Solanum, which also includes potatoes. Thus, the linkage maps of potato and tomato are very similar (e.g. Doganlar et al. 2002) because they share a recent common ancestor.

Sequencing and mapping analysis of trnE-trnT intergenic spacer of tomato cpDNA revealed a 419 bp deletion between trnE (tRNA-Glu (UUC)) and trnT (tRNA-Thr (GGU)) which is not present in the tobacco plastome (GenBank/EMBL/DDBJ #Z00044). The tobacco fragment has 829 bp while the tomato has only 392 bp. Odintsova and Yurina (2003) studied the plastid genome structure of higher plants and affirmed that plastid identity (proportion of identical sites; percentage of nucleotide similarity) among related species is often very high. Based thereon, it was hypothesized that the tomato deletion could be a marker and not only an isolated occurrence during plastid evolution within the family, considering that both tomato and tobacco belong to the Solanaceae family.
Based on the utility of intergenic spacers of cpDNA in resolving phylogenetic and population genetic studies of plant species, it was decided to analyze the potential of the non-coding \textit{trnE-trnT} intergenic spacer to be integrated in phylogenetic analysis. Therefore this cpDNA region was analyzed with the following objectives: 1) to assess the variability of plastid \textit{trnE-trnT} intergenic spacer based on some Solanaceae species; 2) to estimate the phylogeny of this group by maximum parsimony using sweet potato (\textit{Ipomoea batatas}), a Convolvulaceae species, as an outgroup; 3) to test the phylogenetic utility of a relatively small segment of the chloroplast, the \textit{trnE-trnT} region, in resolving taxonomic levels. For this purpose species from other families were analyzed and compared with Solanaceae species.

**MATERIAL AND METHODS**

**Plant Material**

A total of 23 species from the Centro de Biotecnologia Agrícola (CEBTEC, Escola Superior de Agricultura “Luiz de Queiroz”/USP) were analyzed in this study. The \textit{trnE-trnT} intergenic spacer of chloroplast genome were isolated and sequenced from: (a) 16 species of Solanaceae: two varieties of cultivated tomato (\textit{Lycopersicon esculentum} cv. IAC-Santa Clara and cv. Moneymaker), peruvianum tomato (\textit{L. peruvianum}), potato (\textit{Solanum tuberosum}), aubergine (\textit{S. melongena}), “jílo” (\textit{S. gilo}), bittersweet nightshade (\textit{S. dulcamara}), glossy nightshade (\textit{S. americanum}), chinese lantern (\textit{Physalis alkekengi}), sweet pepper (\textit{Capsicum annuum}), pepper (\textit{C. frutescens}), belladonna (\textit{Atropa belladonna}), petunia (\textit{Petunia} sp. obtained at the local market), garden petunia (\textit{Petunia} sp. hybrid), scopolia (\textit{Scopolia carniolica}) and \textit{Physochlainaorientalis}; the tobacco (\textit{Nicotiana tabacum}) sequence was taken from the GenBank database (GenBank/EMBL/DDBJ#Z00044); (b) 1 species of Convolvulaceae: sweet potato (\textit{Ipomoea batatas}); (c) 1 species of Plantaginaceae: \textit{Plantago major}; (d) 1 species of Buddlejaceae: \textit{Buddleja davidii}; and (e) 3 species of Scrophulariaceae: \textit{Antirrhinum majus} (snapdragon), \textit{Lathraea squamaria} and \textit{Scrophularia nodosa}.

**DNA Isolation and PCR Amplification**

DNA preparations were carried out following the method described by Doyle and Doyle (1990). The \textit{trnE-trnT} region of all species was amplified by polymerase chain reaction (PCR) using two specific primers: \textit{trnE-5′} (forward): \textit{CCT TTC GTA GTA CCC TAC CCC 3′} and \textit{trnT-3′} (reverse): \textit{AGC CCC TTA TCG GAT TTG AAC 3′}.

The primers were designed on the 5′ region of \textit{trnE-5′}, which codes for \textit{tRNA-Glu} (UUC) and the 3′ region of \textit{trnT}, which codes for \textit{tRNA-Thr} (GGU) sequences. Amplifications were performed in 50 µL reactions containing 10mM Tris-HCl, pH 9.0, 0.1% Triton X-100, 1.5 mM MgCl₂, 50 mM KCl, 150 µM each dNTP, 10 pmoles of each primer, 50-100 ng of total DNA and 5U of \textit{Taq} polymerase (Gibco BRL). The PCR reaction was performed as specified by the manufacturer (Invitrogen) for \textit{Taq} polymerase reactions (35 PCR cycles; 96, 50 and 72 °C).

**DNA Sequencing**

The forward and reverse primers described in the above section (PCR amplification) were used for amplification and sequencing of the \textit{trnE-trnT} region. The PCR products were used as template for direct sequencing (BigDyeTM Terminator Cycle Sequencing \textit{v 3.0} kit, Applied Biosystems, according to the manufacturers’ instructions) on an ABI-3100 automated sequencer (Applied Biosystems).

**Phylogenetic reconstruction and sequence analysis**

Sequences were aligned using the multiple alignment algorithm CLUSTAL W (Thompson et al. 1994), with subsequent manual correction following the guidelines of Kelchner (2000). The aligned matrix was imported into PAUP*4.0b10 for phylogenetic analysis (Swofford 2001). Maximum parsimony trees were obtained from the resulting matrices using heuristic search options. Searches with 1,000 replicates of random addition sequence (saving no more than 30 trees per replicate to reduce time spent swapping large islands of trees) with the tree bisection reconnection (TBR) branchswapping algorithm and MulTrees on (keeping multiple equally most-parsimonious trees).

Internal support was assessed using 1,000 bootstrap replicates (Felsenstein 1985). Groups with bootstrap percentages of 90-100 were considered to be strongly supported, 80-89 moderately supported and 50-79 weakly supported. Only groups with bootstrap > 50 that are consistent with the strict consensus tree are shown.
DM Melotto-Passarin et al.

Two distinct phylogenetic analyses were estimated by maximum parsimony according to the following strategies: (a) The Solanaceae species, *Lycopersicon esculentum* cv. IAC-Santa Clara and cv. Moneymaker, *L. peruvianum*, *Solanum tuberosum*, *S. melongena*, *Solanum gilo*, *S. dulcamara*, *S. americanum*, *Physalis alkekengi*, *Capsicum annuum*, *C. frutescens*, *Atropa belladonna*, *Nicotiana tabacum*, *Petunia sp.*, *Petunia hybrida*, *Scopolia carniolica* and *Physocodium orientalis*, rooted with *Ipomoea batatas* which is a Convolvulaceae species; and (b) species from three other families, *Ipomea batatas* (Convolvulaceae), *Plantago major* (Plantaginaceae), *Buddleja davidii* (Buddlejaceae), *Antirrhinum majus*, *Lathraea squamaria* and *Sportunaria nodosa* (Scrophulariaceae) were analyzed and compared with Solanaceae species, based on the unrooted tree method.

These two phylogenetic analyses were performed to estimate the phylogeny of some Solanaceae species by maximum parsimony using *Ipomoea batatas* as an outgroup (analysis a) and to test the phylogenetic utility of the *trnE-trnT* chloroplast region in resolving taxonomic levels (analysis b). For this purpose, species from other families were analyzed and compared with Solanaceae species. The analysis of nucleotide sequence identity was based on alignment using Vector NTI Suite v.6.0 (Informax Inc., USA). The percentage of similarity is the proportion of identical sites (for macromolecule sequences).

**RESULTS AND DISCUSSION**

Characterization of *trnE-trnT* intergenic spacer of Solanaceae species

The length of the spacers ranged from 392 bp (*L. esculentum*) to 849 bp (*Scopolia carniolica*) with an average length of 653 bp (Table 1). Multialignment of sequences (not shown, supplementary data on-line) showed that at least part of the nucleotide deletion is present in all studied species together with a range of other small deletions, insertions, duplications and substitutions (both transitions and transversions), indicating that the *trnE-trnT* intergenic spacer could be included in an analysis with multiple chloroplast regions, thus improving the interspecific phylogenetic inferences. Moreover, nucleotide polymorphism was found between *L. esculentum* varieties (IAC-Santa Clara and cv. Moneymaker) indicating variation of the region at the intraspecific level. GC contents of the *trnE-trnT* region are between 26 and 31% in *Solanaceous* species (Table 1) which had been expected since it is an intergenic spacer region (AT rich).

In agreement with Ingvarsson et al. (2003), insertions, deletions and inversions at the chloroplast genome of higher plants were shown to be extremely useful to shed light on the phylogenetic relationships among closely related taxa. Introns and intergenic spacers of plastomes are very useful in studies on phylogenetics (Kelchner 2000) and population genetics (Ingvarsson et al. 2003) of species. It would be interesting to know whether the indels (insertions and deletions of nucleotides) could provide data to increase the robustness of inraspecific studies.

As commonly observed in cpDNA sequences (as an example, in *Oryza, Poacea*, Oliveira 2003), the indels play an important role in the differentiation between *Solanaceous* species (Table 1). The two tomato cultivars are the only species with the same length. *Lycopersicon peruvianum* is just one bp away. *Solanum tuberosum* has an intergenic spacer only slightly longer than *Lycopersicon peruvianum*, while *S. melongena* and *S. gilo* form a distinct group with 529 and 528 bp, respectively.

According to the spacer length the sequences could be divided in three groups based on the general pattern of deletions and similarities as shown in multialignment of sequences (not shown, supplementary data on-line). Interestingly, *S. tuberosum* (399 bp) was more similar to *Lycopersicon* species (392 bp), with 96.2% identity, and less to other *Solanum* species (71.7 and 72.2% identity with *S. melongena* and *S. gilo*, respectively) with slightly longer sequences. *L. esculentum* cultivars (with 99.7% identity) are differentiated by a single substitution (A/T transversion) as shown in sequence multialignment (not shown, supplementary data on-line) (position 349). At this position *L. peruvianum* agrees with cv. Moneymaker with a T instead of an adenine (A) in the cv. IAC-Santa Clara. Two other single points can distinguish *L. peruvianum* from *L. esculentum*: a T insertion/deletion and a G/T transversion (positions 166 and 232, respectively, supplementary data on-line). As expected, *C. annuum* and *C. frutescens* are highly similar (97.8% identity), composing one of the distinct groups of the multialignment. *Physalis alkekengi* is grouped
Table 1. Intergenic trnE-trnT sequence lengths and CG contents (%) in the studied species

<table>
<thead>
<tr>
<th>Species</th>
<th>trnE-trnT Length (bp)</th>
<th>GC content (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antirrhinum majus</td>
<td>536</td>
<td>31</td>
</tr>
<tr>
<td>Atropa belladonna</td>
<td>848</td>
<td>28</td>
</tr>
<tr>
<td>Buddleja davidii</td>
<td>514</td>
<td>31</td>
</tr>
<tr>
<td>Capsicum annuum</td>
<td>682</td>
<td>29</td>
</tr>
<tr>
<td>Capsicum frutescens</td>
<td>676</td>
<td>29</td>
</tr>
<tr>
<td>Ipomoea batatas</td>
<td>731</td>
<td>33</td>
</tr>
<tr>
<td>Lathraea squamaria</td>
<td>569</td>
<td>32</td>
</tr>
<tr>
<td>Lycopersicon esculentum cv IAC-Santa Clara</td>
<td>392</td>
<td>29</td>
</tr>
<tr>
<td>Lycopersicon esculentum cv Moneymaker</td>
<td>392</td>
<td>29</td>
</tr>
<tr>
<td>Lycopersicon peruvianum</td>
<td>393</td>
<td>29</td>
</tr>
<tr>
<td>Nicotiana tabacum</td>
<td>829</td>
<td>29</td>
</tr>
<tr>
<td>Petunia hybrida</td>
<td>848</td>
<td>28</td>
</tr>
<tr>
<td>Petunia sp.</td>
<td>847</td>
<td>28</td>
</tr>
<tr>
<td>Physalis alkekengi</td>
<td>847</td>
<td>29</td>
</tr>
<tr>
<td>Physochlaina orientalis</td>
<td>748</td>
<td>28</td>
</tr>
<tr>
<td>Plantago major</td>
<td>744</td>
<td>30</td>
</tr>
<tr>
<td>Scopolia carniolica</td>
<td>849</td>
<td>28</td>
</tr>
<tr>
<td>Scrophularia nodosa</td>
<td>505</td>
<td>29</td>
</tr>
<tr>
<td>Solanum americanum</td>
<td>835</td>
<td>29</td>
</tr>
<tr>
<td>Solanum dulcamara</td>
<td>470</td>
<td>26</td>
</tr>
<tr>
<td>Solanum gilo</td>
<td>528</td>
<td>31</td>
</tr>
<tr>
<td>Solanum melongena</td>
<td>529</td>
<td>31</td>
</tr>
<tr>
<td>Solanum tuberosum</td>
<td>399</td>
<td>28</td>
</tr>
</tbody>
</table>

with Capsicum species, due to the similarities in their sequences (75.6% identity) (sequence multialignment not shown, supplementary data on-line).

In the trnE-trnT intergenic spacer region, Nicotiana tabacum and Atropa belladonna were 94.8% identical. Compared with Petunia species, the identity of N. tabacum and A. belladonna in nucleotide sequence was 93.1 and 90.4%, respectively. Petunia sp. (obtained at the local market) and Petunia hybrida are very close to each other and have only one indel (position 171) in which Petunia sp. does not have a T. The trnE-trnT sequences of these two species and Scopolia carniolica are 99.8% identical (sequence multialignment not shown, supplementary data on-line).

Physochlaina orientalis has the most divergent sequence of Solanaceae species. Because of the accumulated deletions, this is the most divergent species of the Lycopersicin clade, with only 31% identity. Also, compared with Lycopersicin sequences, the sequence identity of Ipomoea batatas is 33.8% in the trnE-trnT spacer. P. orientalis is closer to I. batatas (51.8% identity in the nucleotide sequence) than to Solanaceae species.

Convolvulaceae has been the subject of only two family-wide molecular phylogenetic studies (Stefanovic et al. 2002, Stefanovic et al. 2003). The first study (Stefanovic et al. 2002) was based on sequences of four chloroplast loci: rbcL, atpB, psbE-J operon, and the trnL-F region. These sequences were obtained from a broad sample of taxa within the family, including 102 species of all nine traditionally recognized non-parasitic tribes (Austin 1998), seven Cuscuta species, as well as three outgroups. The cpDNA results confirmed that Convolvulaceae are sister to Solanaceae, with 100% bootstrap support for each family and the clade comprising both families.

The results presented here suggest that the trnE-trnT intergenic spacer provides nucleotide polymorphism in indels and mutations (transition and transversion) that are interesting to increase the robustness of phylogenetic studies of Solanaceae and related species.

Phylogenetic Analysis

The phylogeny of some Solanaceae species was estimated based on the multialignment of trnE-trnT
Three major clades were found in Solanaceae species: the Petunia clade with Petunia species and S. carniiola (100% bootstrap support), the Capsicum clade with C. frutescens and C. annuum (98%), and the Solanum clade that includes Solanum and Lycopersicon species (61%).

S. tuberosum was shown to be more related to Lycopersicon species than to other Solanum species, in agreement with Borisjuk et al. (1994) who pointed out that the New World Solanum species is more related to Lycopersicon than to other Solanum species. Nucleotide sequence analysis of a phylogenetically informative part of the 3’ end of 25S rDNA confirmed a closer relationship between tomato and potato than between tomato and tobacco, previously detected by Southern hybridization with an intergenic spacer element fragment of Solanum tuberosum as hybridization probe (Borisjuk et al. 1994).

Examples abound of the importance of a phylogenetic framework for diverse areas of plant research (for review, see Daly et al. 2001). One obvious example is the value of placing model organisms in the appropriate phylogenetic context to obtain a better understanding of both patterns and processes of evolution. Tomato (Lycopersicon esculentum) and other species of this small genus are actually embedded within the well-marked subclade Solanum (and, hence, are more appropriately referred to as a species of Solanum, so that tomato was renamed Solanum lycopersicon; e.g., Olmstead et al. 1999, Spooner et al. 1993). This is an important statement for geneticists, molecular

Figure 1. One of the 12 most-parsimonious trees resulted from the parsimony analysis of cpDNA trnE-trnT intergenic spacer sequences. Ipomoea batatas was used as outgroup. Numbers indicate bootstrap percentages > 50%
Phylogenetic relationships in Solanaceae and related species based on cpDNA sequence from plastid trnE-trnT region

biologists, and plant breeders, with a view to comparative genetic/genomic research and crop improvement.

These results are also in agreement with previous data obtained by calculation of nucleotide substitution rates in cpDNA (Kawagoe and Kikuta 1991), which demonstrated that evolutionary separation between Solanum and Lycopersicon occurred considerably later than between these two genera and Nicotiana (as discussed by Borisjuk et al. 1994).

The phylogenetic analysis based on the trnE-trnT spacer presented here (Figure 1) shows clearly that Solanum, as defined traditionally, is a paraphyletic taxon. Thus, the absorption of Lycopersicon by Solanum seems justifiable.

The sequences of Petunia hybrida, Petunia sp. and Scopolia carniolica are almost 100% similar, apart from only a few mutations. They were grouped in a trichotomy clade. The sequences of Nicotiana tabacum and Atropa belladonna were very similar.

Physochlaina orientalis is the most divergent species of the Solanaceae family based on the trnE-trnT region, which could be directly related with the geographical location of the genus Physochlaina (D’Arcy 1991).

The trnD-trnT (which contains trnE-trnT) non-coding chloroplast region was used for some successful phylogenetic studies but always in association with other regions, eg., the internal transcribed spacer region (ITS) of the nuclear ribosomal DNA (rDNA) or the trnT-trnF non-coding chloroplast region (Friesen et al. 2000, Yang et al. 2002).

The results presented here suggest that nucleotide sequence polymorphism of trnE-trnT should be included in analysis with multiple chloroplast regions and can contribute to the resolution of phylogenetic studies of Solanaceae and related species.

Comparative analysis of Solanaceae, Convolvulaceae, Plantaginaceae, Buddlejaceae, and Scrophulariaceae species using the trnE-trnT region

To reinforce the usage range of the trnE-trnT region at different taxonomic levels, some sequences of species of five different families were compared by sequence multialignment (not shown, see supplementary data on-line) and identified by phylogenetic reconstruction.

The second phylogenetic analysis included 17 species of Solanaceae and 1 species of Convolvulaceae (Ipomoea batatas), 1 species of Plantaginaceae (Plantago major), 1 species of Buddlejaceae (Buddleja davidii), and 3 species of Scrophulariaceae (Antirrhinum majus, Lathraea squamaria and Scrophularia nodosa). Multialignment of entire sequences (not shown, supplementary data on-line) showed divergent regions among different family species. Some deletions, insertions, duplications and substitutions (both transition and transversions) were detected. Physochlaina orientalis, which was the most divergent Solanaceae species, was grouped in the same clade as Plantago major (Plantaginaceae family), which may be an indicator, if confirmed, that Solanaceae is polyphyletic. Buddlejaceae and Scrophulariaceae have long been accepted as close families (Dahlgren 1983). When a restricted set of Buddlejaceae is considered (Buddleja, Emorya, Gomphostigma and Nicodemus) in molecular phylogenetic studies, both families fall in the same clade, with strong support (Bremer et al. 2001, Oxelman et al 1999). A third family, Myoporaceae, was added to the clade by a three-gene reconstruction (Olmstead et al. 2001). Buddleja davidii and Scrophularia nodosa were grouped in the same clade in the trnE-trnT IGS tree, corroborating the recent trend to cluster this group of families on a monophyletic basis. Lathraea squamaria is a parasite plant and did not group in a clade. Of the 1,425 characters from aligned trnE-trnT matrix, 1,111 are constant. Among the 314 variable characters, 138 are parsimony-uninformative and 176 are parsimony-informative. Figure 2 shows one of the 180 equally most parsimonious trees for trnE-trnT sequence data in which the topology of the tree under parsimony was essentially the same as the ones under maximum likelihood and neighbor-joining criteria (data not shown).

Snapdragon (Antirrhinum majus) was historically part of a broadly defined Scrophulariaceae, a family that is now known to be grossly polyphyletic (i.e., not a single clade). Phylogenetic studies indicate that Scrophulariaceae should be broken up into several families (Olmstead et al. 2001), and snapdragon and its closest relatives are part of a clade recognized as the family Plantaginaceae.

Although Antirrhinum spp. has long been placed in the family Scrophulariaceae, molecular phylogenetic studies indicate that the traditionally recognized
Scrophulariaceae are not a single clade but actually represent a number of distinct clades: Scrophulariaceae in the strict sense; Plantaginaceae, which includes *Antirrhinum, Plantago, and Veronica*; Orobanchaceae, which contains all of the parasitic taxa formerly placed in either Orobanchaceae or Scrophulariaceae; the new family Calceolariaceae; an expanded Stilbaceae; and an expanded Phyrmaceae (Olmstead et al. 2001). The *trnE-trnT* phylogeny corroborates the proposal for the disintegration of Scrophulariaceae by placing *Antirrhinum* and *Plantago* in the same clade (Plantaginaceae). Given the absence of other parasitic species in the data matrix, it is impossible to know whether *Lathraea* would be placed with Orobanchaceae or Scrophulariaceae (sensu Olmstead et al. 2001). Daniell et al. (2006) sequenced the cpDNA of *Solanum bulbocastanum, Solanum lycopersicum* for a comparative analysis with other Solanaceae genomes. They affirmed that only four spacer regions are fully conserved (100% sequence identity) among all genomes; deletions or insertions within some intergenic spacer regions result in less than 25% sequence identity, underscoring the importance of choosing appropriate intergenic spacers for plastid transformation and providing valuable new information for phylogenetic utility of the chloroplast intergenic spacer regions. These results show the importance of studying the potential of integration of an intergenic spacer of cpDNA in phylogenetic analysis.

Samson et al. (2007) compared the coffee chloroplast genome with sequenced genomes of the closely related family Solanaceae and identified large indels (> 500 bp) in several intergenic spacer regions and introns in the Solanaceae, including *trnE* (UUC)-*trnT* (GGU) spacer, *ycf4-cemA* spacer, *trnl* (GAU) intron and *rrn5-trnR* (ACG) spacer. Our results are in agreement with Samson et al. (2007) and we assume that the tomato deletion represents a marker and not only an isolated occurrence during plastid evolution within the family.

**CONCLUSION**

Sequencing of the Solanaceae species studied showed that chloroplast sequences are highly conservative between related species. However, a number of mutations such as indels and base substitutions were observed. This relative conservatism of the genome allows the investigation of the phylogenetic relationships among divergent plants. The non-coding *trnE-trnT*
Phylogenetic relationships in Solanaceae and related species based on cpDNA sequence from plastid \textit{trnE-trnT} region

intergenic spacer thus showed nucleotide sequence polymorphism and variation in sequence length between Solanaceae and related species. It could be included in analysis with multiple chloroplast regions to improve the resolution of phylogenetic studies of the species studied.

ACKNOWLEDGEMENTS

We thank the FAPESP (Fundação de Amparo a Pesquisa do Estado de São Paulo) for granting D.M.Melotto-Passarin a graduate scholarship (no. 02/04797-6).

Relações filogenéticas em Solanaceae e espécies relacionadas baseado em sequência de cpDNA da região plastidial \textit{trnE-trnT}

RESUMO - Espaçadores intergênicos do DNA cloroplastidial (cpDNA) são muito úteis em estudos filogenéticos e de genética de populações de espécies de plantas, sendo interessante estudar o seu potencial em análise filogenética. O espaçador intergênico não codificado \textit{trnE-trnT} do cpDNA foi analisado com o objetivo de avaliar o polimorfismo da sequência de nucleotídeos de dezenas de espécies de Solanaceae e estimar a sua capacidade em auxiliar na resolução dos estudos filogenéticos deste grupo. Alinhamentos múltiplos das sequências de DNA do espaçador intergênico \textit{trnE-trnT} possibilitou a identificação de variabilidade de nucleotídeo nesta região e a filogenia foi estimada pela máxima parcimônia e enraizada com Ipomoea batatas, uma Convolvulaceae, a família mais relacionada. Também foi testada a utilidade filogenética deste espaçador intergênico em resolver níveis taxonômicos. Para este propósito, espécies de quatro outras famílias foram analisadas e comparadas com as espécies Solanaceae. Este resultado de análise filogenética reforçou o polimorfismo da região \textit{trnE-trnT} em níveis taxonômicos diferentes.

Palavras-chave: espaçador intergênico, alinhamento múltiplo de DNA, polimorfismo de nucleotídeos, níveis taxonômicos, máxima parcimônia.

REFERENCES


DM Melotto-Passarin et al.


Crop Breeding and Applied Biotechnology 8: 85-95, 2008
Phylogenetic relationships in Solanaceae and related species based on cpDNA sequence from plastid $\text{trnE-trnT}$ region
