



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

Danila Montewka Melotto-Passarin¹, Irving Joseph Berger², Keini Dressano¹, Valentina de Fátima De Martin¹, Giancarlo Conde Xavier Oliveira³, Ralph Bock⁴, and Helaine Carrer¹

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

¹ Laboratório de Biotecnologia Agrícola, Departamento de Ciências Biológicas, Escola Superior de Agricultura "Luiz de Queiroz" (ESALQ), Universidade de São Paulo (USP), Avenida Pádua Dias, 11, 13.418-900, Piracicaba, SP, Brasil. *E-mail: hecarrer@esalq.usp.br

² Companhia de Cigarros Souza Cruz, Av. Gal Pinto Tourinho, 3200, 83.880-000, Bom Jesus, Rio Negro, PR, Brasil

³ Departamento de Genética, ESALQ, USP

⁴ Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, D-14476, Potsdam, Golm, Germany

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 *Hydroleaceae*, *Montiniaceae* and *Sphenocleaceae*, these families represent the order *Solanales* (APG 2003).

Garcia and Olmstead (2003) reconstructed phylogenies of tribe *Anthocercideae* (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 *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 *trnE-trnT* intergenic spacer based on some Solanaceae species; 2) to estimate the phylogeny of this group by maximum parsimony using sweet potato (*Ipomoea batatas*), a Convolvulaceae species, as an outgroup; 3) to test the phylogenetic utility of a relatively small segment of the chloroplast, the *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 *trnE-trnT* intergenic spacer of chloroplast genome were isolated and sequenced from: (a) 16 species of Solanaceae: two varieties of cultivated tomato (*Lycopersicon esculentum* cv. IAC-Santa Clara and cv. Moneymaker), peruvianum tomato (*L. peruvianum*), potato (*Solanum tuberosum*), aubergine (*S. melongena*), “jiló” (*S. gilo*), bittersweet nightshade (*S. dulcamara*), glossy nightshade (*S. americanum*), chinese lantern (*Physalis alkekengi*), sweet pepper (*Capsicum annuum*), pepper (*C. frutescens*), belladonna (*Atropa belladonna*), petunia (*Petunia* sp. obtained at the local market), garden petunia (*Petunia hybrida*), scopolia (*Scopolia carniolica*) and *Physochlaina orientalis*; the tobacco (*Nicotiana tabacum*) sequence was taken from the GenBank database (GenBank/EMBL/DBJ#Z00044); (b) 1 species of Convolvulaceae: sweet potato (*Ipomoea batatas*); (c) 1 species of Plantaginaceae: *Plantago major*; (d) 1 species of Buddlejaceae: *Buddleja davidii*; and (e) 3 species of Scrophulariaceae: *Antirrhinum majus* (snapdragon), *Lathraea squamaria* and *Scrophularia nodosa*.

DNA Isolation and PCR Amplification

DNA preparations were carried out following the method described by Doyle and Doyle (1990). The *trnE-*

trnT region of all species was amplified by polymerase chain reaction (PCR) using two specific primers: *trnE*-5' (forward): 5' CCT TTC GTA GTA CCC TAC CCC 3' and *trnT*-3' (reverse): 5' AGC CCC TTA TCG GAT TTG AAC 3'.

The primers were designed on the 5' region of *trnE*-5', which codes for tRNA-Glu (UUC) and the 3' region of *trnT*, which codes for 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 *Taq* polymerase (Gibco BRL). The PCR reaction was performed as specified by the manufacturer (Invitrogen) for *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 *trnE-trnT* region. The PCR products were used as template for direct sequencing (BigDye™ Terminator Cycle Sequencing 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.

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 *Physochlaina orientalis*, rooted with *Ipomoea batatas* which is a Convolvulaceae species; and (b) species from three other families, *Ipomoea batatas* (Convolvulaceae), *Plantago major* (Plantaginaceae), *Buddleja davidii* (Buddlejaceae), *Antirrhinum majus*, *Lathraea squamaria* and *Scrophularia 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 intraspecific studies.

As commonly observed in cpDNA sequences (as an example, in *Oryza*, Poaceae, 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 GC contents (%) in the studied species

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

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 *Lycopersicon* clade, with only 31% identity. Also, compared with *Lycopersicon* 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*

spacer sequences (not shown, supplementary data online) using sweet potato (*Ipomoea batatas*), a Convolvulaceae species, as an outgroup.

The first analysis includes 17 species of Solanaceae and the outgroup *Ipomoea batatas* (Convolvulaceae). Of the 1,352 characters from aligned *trnE-trnT* matrix, 1,142 are constant. Among the 210 variable characters, 156 are parsimony-uninformative and 54 are parsimony informative. The analysis of this data set identified 12 equally most-parsimonious trees. Figure 1 shows the parsimony strict consensus tree, which was essentially equal to the maximum likelihood and neighbor-joining trees (data not shown). The ingroup is divided into three main clades, one containing *Physalis*, *Capsicum frutescens* and *C. annuum*; another one with *Scopolia carniolica*, *Petunia hybrida* and *Petunia* sp.; the third clade is composed by *Solanum* and *Lycopersicon* species. The *Atropa belladonna* and *Nicotiana tabacum* clade was grouped with the *Scopolia carniolica*, *Petunia hybrida* and *Petunia* sp. Clade, with a rather weak bootstrap support (less than 50%). The two varieties of *L. esculentum* were grouped in one clade, as a sister group to *L. peruvianum*. Five clades had robust bootstrap support (95-100% bootstrap support; Figure 1). The support of the clades that grouped *C. frutescens*, *C. annuum* and *P. alkekengi* (53%), of the *Solanum* clade that includes *Solanum* and *Lycopersicon* species (61%) and of the clade that grouped the two *L. esculentum* cultivars (65%) was intermediate and five clades were poorly supported (Figure 1).

Three major clades were found in Solanaceae species: the *Petunia* clade with *Petunia* species and *S. carniolica* (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

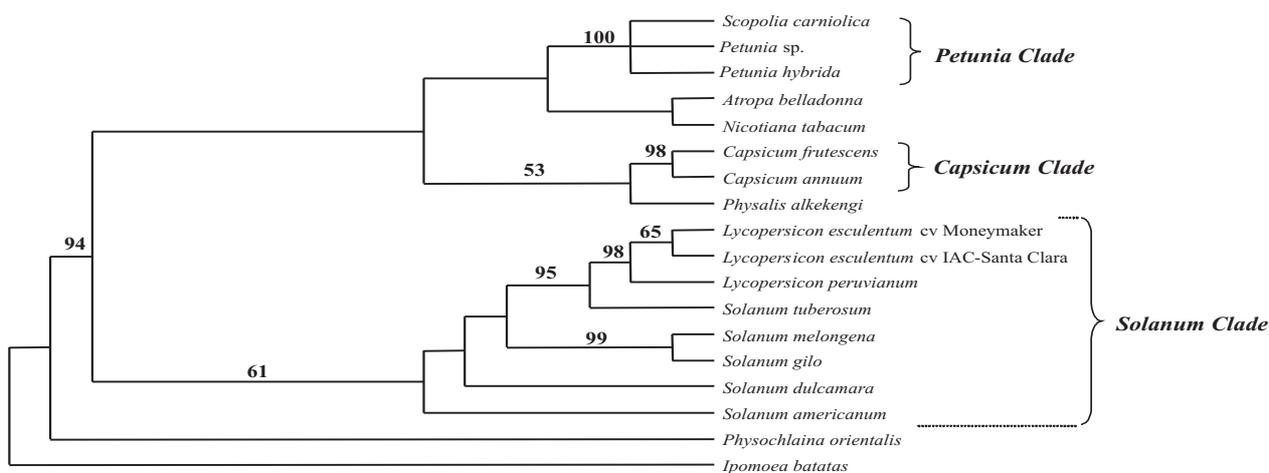


Figure 1. One of the 12 most-parsimonious trees resulted from the parsimony analysis of cpDNA *trnE-trnT* intergenic spacer sequences. *Ipomoea batatas* was used as outgroup. Numbers indicate bootstrap percentages > 50%

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 (nrDNA) 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 *Nicodemia*) 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

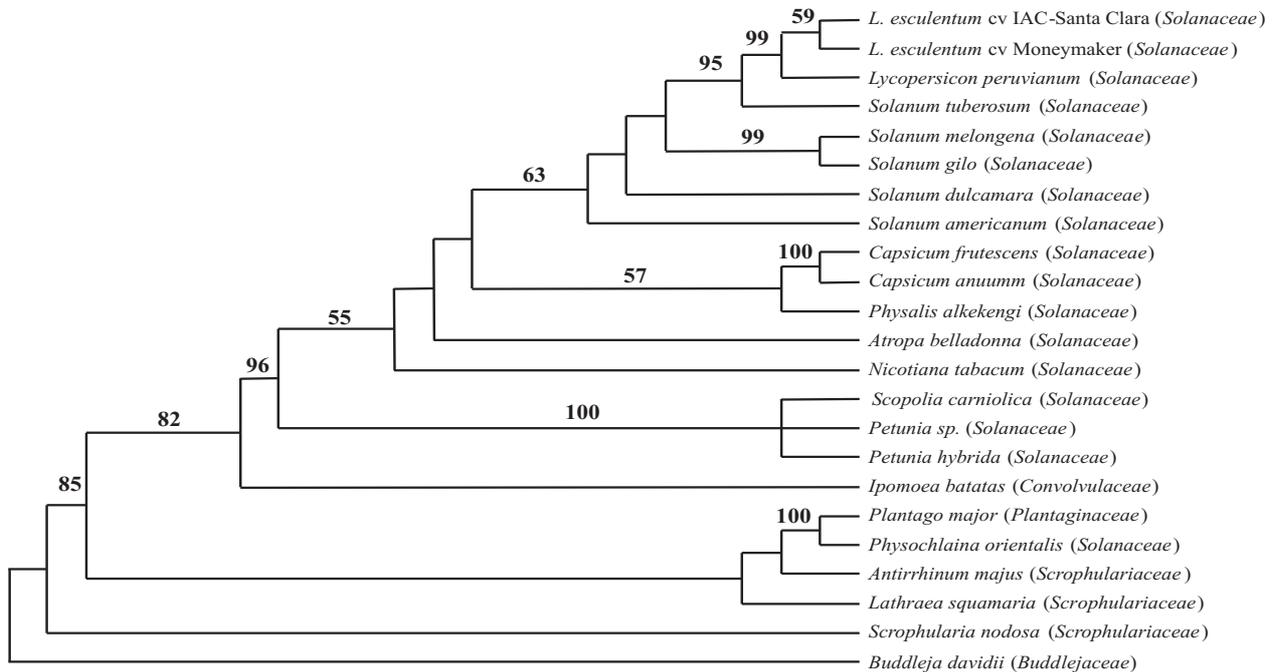


Figure 2. The strict consensus of the 180 equally maximum parsimonious unrooted trees obtained from *trnE-trnT* intergenic spacer sequences of 17 Solanaceae species, 1 Plantaginaceae, 1 Buddlejaceae and 3 Scrophulariaceae species. Numbers indicate bootstrap percentages > 50%

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 Phymaceae (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, *trnI* (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*

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 seqüência de cpDNA da região plastidial *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 *trnE-trnT* do cpDNA foi analisado com o objetivo de avaliar o polimorfismo da seqüência de nucleotídeos de dezesseis espécies Solanaceae e estimar a sua capacidade em auxiliar na resolução dos estudos filogenéticos deste grupo. Alinhamentos múltiplos das seqüências de DNA do espaçador intergênico *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 *trnE-trnT* em níveis taxonômicos diferentes.

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

REFERENCES

- APG (2003) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. **Botanical Journal of the Linnean Society** **141**: 399-436.
- Austin DF (1998) Parallel and convergent evolution in the Convolvulaceae. In: Mathews P and Sivadasan M (eds.) **Biodiversity and taxonomy of tropical flowering plants**. Mentor Books, Calicut, p. 201-234.
- Bohs L and Olmstead RG (1997) Phylogenetic relationships in *Solanum* (Solanaceae) based on *ndhF* sequences. **Systematic Botany** **22**: 5-18.
- Borisjuk N, Borisjuk L, Petjuch G and Hemleben V (1994) Comparison of nuclear ribosomal RNA genes among *Solanum* species and other Solanaceae. **Genome** **37**: 271-279.
- Bremer K, Backlund A, Sennblad B, Swenson U, Amdreassen K, Hjertson M, Lundberg J, Backlund M and Bremer B (2001) A phylogenetic analysis of 100 genera and 50 families of euasterids based on morphological and molecular data with notes on possible higher level morphological synapomorphies. **Plant Systematics and Evolution** **229**: 137-169.
- Chiang T-Y and Schaal BA (2000) Molecular evolution and phylogeny of the *atpB-rbcL* spacer of chloroplast DNA in the true mosses. **Genome** **43**: 417-426.
- Clarkson JJ, Knapp S, Garcia VF, Olmstead RG, Leitch AR and Chase MW (2004) Phylogenetic relationships in *Nicotiana* (Solanaceae) inferred from multiple plastid DNA regions. **Molecular Phylogenetics and Evolution** **33**: 75-90.
- Dahlgren R (1983) General aspects of angiosperm evolution and macrosystematics. **Nordic Journal of Botany** **3**: 119-149.
- Daly DC, Cameron KM and Stevenson DW (2001) Plant systematics in the age of genomics. **Plant Physiology** **127**: 1328-1333.
- Daniell H, Lee SB, Grevich J, Saski C, Quesada-Vargas T, Guda C, Tomkins J and Jansen RK (2006) Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analyses with other Solanaceae genomes. **Theoretical and Applied Genetics** **112**: 1503-1518.
- D'arcy WG (1991) The Solanaceae since 1976, with a review of its biogeography. In: Hawkes JG, Lester RN, Nee M and Estrada RN (eds.) **Solanaceae III: taxonomy, chemistry, evolution**. The Royal Botanical Gardens, Kew, Surrey, p.75-137.

- Doganlar S, Frary A, Daunay MC, Lester RN and Tanksley SD (2002) A comparative genetic linkage map of aubergine (*Solanum melongena*) and its implications for genome evolution in the Solanaceae. **Genetics** **161**: 1697-1711.
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. **Evolution** **39**: 783-791.
- Fukuda T, Yokoyama J and Ohashi H (2001) Phylogeny and biogeography of the genus *Lycium* (Solanaceae): inferences from chloroplast DNA sequences. **Molecular Phylogenetics and Evolution** **19**: 246-58.
- Garcia VF and Olmstead RG (2003) Phylogenetics of Tribe *Anthocercideae* (Solanaceae) based on *ndhF* and *trnL/F* sequence data. **Systematic Botany** **28**: 609-615.
- Hawkes JG, Lester RN, Nee M and Estrada-RN (1991) **Solanaceae III, Taxonomy, Chemistry, Evolution**. The Royal Botanical Gardens, Kew, Surrey. 492p.
- Hawkes JG (1992) Biosystematics of the potato. In: Harris PM (ed.) **The Potato crop**. Chapman and Hall, London, p.909.
- Ingvarsson PK, Ribstein S and Taylor DR (2003) Molecular evolution of insertions and deletions in the chloroplast genome of *Silene*. **Molecular Biology and Evolution** **20**: 1737-1740.
- Kawagoe Y and Kikuta Y (1991) Chloroplast DNA evolution in potato (*Solanum tuberosum* L.). **Theoretical Applied Genetics** **81**: 13-20.
- Kelchner SA (2000) The evolution of noncoding chloroplast DNA and its application in plant systematics. **Annual Missouri Botanical Garden** **87**: 482-498.
- Knapp S (2001) Is morphology dead in *Solanum* taxonomy? In: van der Werden G, Barendse G and van den Berg R (eds.) **Solanaceae V**. University of Nijmegen; Nijmegen, The Netherlands, p.23-38.
- Knapp S (2002) Floral diversity and evolution in the Solanaceae. In: Cronk QCB, Bateman RM and Hawkins JA (eds.) **Developmental genetics and plant evolution**. Taylor and Francis, London, p.267-297.
- Lee C and Wen J (2004) Phylogeny of *Panax* using chloroplast *trnC-trnD* intergenic region and the utility of *trnC-trnD* in interspecific studies of plants. **Molecular Phylogenetics and Evolution** **31**: 1-10.
- Miz RB, Mentz LA and Souza-Chies TT (2008) Overview of the phylogenetic relationships of some southern Brazilian species from section *Torva* and related sections of "spiny *Solanum*" (*Solanum* subgenus *Leptostemonum*, Solanaceae). **Genetica** **132**: 143-158.
- Neves SS, Swire-Clark G, Hilu KW and Baird WV (2005) Phylogeny of Eleusine (Poaceae: Chloridoideae) based on nuclear ITS and plastid *trnT-trnF* sequences. **Molecular Phylogenetics and Evolution** **35**: 395-419.
- Odintsova MS and Yurina NP (2003) Plastid genomes of higher plants and algae: structure and functions. **Molecular Biology** **37**: 649-662.
- Oliveira GCX (2003) **A molecular phylogenetic analysis of *Oryza L.* based on chloroplast DNA sequences**. PhD Thesis, Washington University, Saint Louis, USA.
- Olmstead RG, Sweere JA, Spangles RE, Bohs L and Palmer JD (1999) Phylogeny and Provisional Classification of the Solanaceae Based on Chloroplast DNA. In: Nee M, Symon DE, Jessup JP, Hawkes JG (eds.) **Solanaceae IV, Advances in Biology and Utilization** The Royal Botanical Gardens, Kew, Surrey., p. 111-137.
- Olmstead RG, Depamphilis CW, Wolfe AD, Young ND, Elisons WJ and Reeves A (2001) Disintegration of the Scrophulariaceae. **American Journal of Botany** **88**: 348-361.
- Oxelmann B, Backlund M and Bremer B (1999) Relationships of Buddjaceae s.l. investigated using parsimony jackknife and branch support analysis of chloroplast *ndhF* and *rbcL* sequence data. **Systematic Botany** **24**: 164-182.
- Samson N, Bausher MG, Lee SB, Jansen RK and Daniell H (2007) The complete nucleotide sequence of the coffee (*Coffea arabica* L.) chloroplast genome: organization and implications for biotechnology and phylogenetic relationships amongst angiosperms. **Plant Biotechnology Journal** **5**: 339-353.
- Soltis DE and Soltis PS (1998) Choosing an approach and an appropriate gene for phylogenetic analysis. In: Soltis DE, Soltis PS and Doyle JJ (eds.) **Molecular systematics of plants II: DNA sequencing**. Kluwer Academic Publishers, Boston, p. 1-42.
- Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoot SB, Fay MF, Astell M, Swensen SM, Prince LM, Kress WJ, Nixon KC and Farris JS (2000) Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. **Botanical Journal of the Linnean Society** **133**: 38-461.
- Spooner DS, Anderson GJ and Jansen RK (1993) Chloroplast DNA evidences for the interrelationships of tomatoes, potatoes and pepinos (Solanaceae). **American Journal of Botany** **80**: 676-688.
- Stefanovic S, Krueger L and Olmstead RG (2002) Monophyly of the Convolvulaceae and circumscription of their major lineages based on DNA sequences of multiple chloroplast loci. **American Journal of Botany** **89**: 1510-1522.
- Stefanovic S, Austin DF and Olmstead RG (2003) Classification of Convolvulaceae: A Phylogenetic Approach. **Systematic Botany** **28**: 791-806
- Swofford DL (2001) **PAUP*: Phylogenetic Analysis Using Parsimony**, version 4. Sinauer, Sunderland, MA.

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

Thompson JD, Higgins DG and Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. **Nucleic Acids Research** **22**: 4673-4680.

Woodland DW (1997) **Contemporary plant systematics**, 2nd ed., Andrews University Press, Barron Springs, Michigan. 619p.