Sequence Analysis of the Chalcone Synthase Gene Intron in Four Petunia Jussieu Taxa

R.J. Griesbach and R.M. Beck
Floral and Nursery Plants Research, U.S. National Arboretum, U.S. Department of Agriculture, Agricultural Research Service, Beltsville, MD 20705-2350

ABSTRACT. The sequence of the intron within the chalcone synthase A gene (ChsA) was used to characterize Petunia integrifolia subsp. depauperata (Fries) Smith et Downs, P. littoralis Ando et Hashimoto, P. littoralis Smith et Downs, and an unknown taxon from the town of Torres in Brazil. Based upon the intron, the Torres taxon most closely resembled P. integrifolia. The unrooted phylogenetic tree suggested that P. integrifolia was more closely related to P. littoralis than P. altiplana.


H.J.W. Wijsman (personal communication) reported that the traits used to distinguish many of the P. integrifolia taxa were correlated with geographical distribution, concluding that various taxa were subspecies of a single broadly defined species. He recognized three different subspecies [P. integrifolia subsp. integrifolia (Fries) Wijsman, P. integrifolia subsp. inflata (Fries) Wijsman, and P. integrifolia subsp. occidentalis (Fries) Wijsman] and a single variety [P. integrifolia subsp. integrifolia var. depauperata (Fries) Wijsman]. A similar study was conducted by Ando et al. (1995) and concluded that all the species, except P. littoralis, were clearly distinct. Petunia littoralis could not be distinguished from P. integrifolia subsp. integrifolia var. depauperata. In addition, all of the P. integrifolia subspecies in their native habitats were separately distributed and readily distinguished from one another.

Additional taxonomic confusion resulted from the spontaneous appearance in botanical gardens of a purple-flowered intraspecific hybrid between P. axillaris and P. integrifolia (Loudon, 1840). All of these garden hybrids were called P. violacea. Vilmorin (1863) tried unsuccessfully to correct the problem in nomenclature by naming the intraspecific hybrid P. axillaris × P. integrifolia. By 1900, the true species had disappeared from cultivation and even one of the herbarium specimens of P. integrifolia in the Lindley Herbarium at the Royal Botanic Garden at Kew was reported to be a P. axillaris × P. integrifolia hybrid (Ferguson and Ottley, 1932).

In an effort to help solve the species identification problem, several molecular marker studies were conducted using species-specific repeated DNA (Shepherd et al., 1990), rDNA (Benabdelmouna and Abirached-Darmency, 1997; Kabbaj et al., 1995; Zeboudj et al., 1994), randomly amplified DNA (Benabdelmouna et al., 1999; Cerny et al., 1996), or the chalcone synthase intron (Griesbach et al., 2000). However, the taxa that were actually used in many of these studies are uncertain. For example, phylogenetic analysis using rDNA suggested that two taxa called “P. integrifolia” and “P. violacea” were distantly related (Kabbaj et al., 1995), but those names refer to the same taxon. Furthermore, the complex taxonomy of Petunia has caused many cultivated species to be misidentified.

The coding sequence of the chalcone synthase gene (Chs) is highly conserved among plants and has been used to define taxonomic relationships (Niesbach-Klösgen et al., 1987; Rauscher et al., 1999). There is a >66% DNA sequence similarity between monocots and dicots. Chs was used to study the phylogenetic relationships between seven different ecotypes of Ipomoea purpurea (Durbin et al., 2001). Within the coding sequence of the ecotypes, there were only 23 nucleotide substitutions out of 1104 base pairs (bp); while within the intron sequence of the ecotypes, there was significant variation in the number and type of transposable element insertions. Other studies have also shown that transposable elements can be integral parts of introns (Casacuberta and Santiago, 2003; Feschotte et al., 2002).
PLANT MATERIAL.

Four *Petunia* taxa were collected from the wild [Fig. 1; *P. integrifolia* subsp. *integri folia* var. *depauperata* (*Pi*; lat. 29°15′51″S, long. 49°41′33″W), *P. altiplana* (*Pa*; lat. 29°24′02″S, long. 50°26′57″W), *P. littoralis* (*Pl*; lat. 27°36′11″S, long. 48°26′09″W) and *Petunia* sp. Torres (*Pt*; lat. 29°21′26″S, long. 49°41′33″W)]. These plants were identified based upon the taxonomic key developed by Ando and Hashimoto (Fig. 2). The collected plants contained seed capsules that were used to raise populations. The plants within each of the populations were analyzed. Five randomly selected seedlings were used to obtain the morphological data and three randomly selected seedlings were used to obtain the DNA sequence data.

Plants were grown in a mixture of 50% perlite and peat under standard greenhouse conditions and fertilized weekly with a solution of 20N–8.7P–16.6K fertilizer at 200 mg·L−1 nitrogen. Leaf measurements were made on mature, fully expanded leaves before flowering.

**Material and Methods**

**DNA isolation and amplification.** Genomic DNA was isolated from either leaves or flowers using DNeasy Plant Mini Kit (Qiagen, Valencia, Calif.) and further purified with anion-exchange resin using the Genomic-tip Protocol of Qiagen’s Blood and Cell Culture DNA Kit. Purified DNA was precipitated with 50% isopropanol (v/v), washed with 70% ethanol, and resuspended in 30 to 50 µL water. We previously developed a polymerase chain reaction (PCR) system for amplifying the *ChsA* intron in several *Petunia* taxa (Griesbach et al., 2000). PCR primers for amplifying the chalcone synthase (*Chs*) intron were designed by hand from the reported sequences of eight (Niesbach-Klosgen et al., 1987). Primers were designed to include a highly conserved region of the coding sequence as an internal control for checking the accuracy of amplification. The two primers were initially designed to be complementary to 20 nucleotides before the intron (5′-GAGCCACACCTACAAAC-3′) and to 20 nucleotides after the intron (5′-AACCCCTGCTGTGAGCATC-3′).

Using these two primers it was possible to amplify the intron, but we could not completely sequence the product directly from the PCR product. Therefore, internal primers based upon the DNA sequence of *Pi* were designed for the entire length of the intron. The forward primers were 5′-CGATGGAGGAGGAAACACT-GACGTGAAGG-3′, 5′-GATGCTAGGAAGCATAGTG-3′, and 5′-GAAAATTCGTGTGTTTCTAC-3′. The reverse primers were 5′-CCACTATGTCCTTGGCAGACATC-3′, 5′-CTTTCAACGTCAAGTTTCCTTCTC-3′, 5′-CCTCTCTCCTC-TACCCTCCAT-3′, 5′-GCTATTGGAGACGTGAAAG-3′, and 5′-GCTAGAAGACGAGTCAACA-3′. These internal primers were used in various combinations to amplify partial sequences.

A Perkin-Elmer DNA Thermal Cycler (Perkin-Elmer, Norwalk, Conn.) was used to amplify the intron. The reaction mix contained 10 µL (1 mg·mL−1) genomic DNA, 1.5 µL AmpliTaq Gold Taq DNA polymerase (PE Applied Biosystems, San Jose, Calif.), 10 µL GeneAmp MgCl2-free 10X PCR Buffer (PE Applied Biosystems), 8 µL mixed dNTPs (10 mM), 15–20 µL MgCl2 (10 mM), 10 µL mixed primers (20 mM), and water to a total volume of 100 µL. An overlay of 20 µL mineral oil was used in each reaction tube.

The PCR cycle began with a Taq polymerase heat activation step of 95 °C for 12 min. This was followed by 92 °C for 2 min, 40 cycles of 92 °C for 30 s, 60 °C for 2 min, and 72 °C for 1 min, and 72 °C for 10 min.

1. Pedicel deflexed in fruiting stage
2. Stigma between anthers of longest stamen and shortest stamen
3. Main stem prostrate
   4. Leaves prostrate, broad spatulate, subglabrous; stem hirsute, branched at lower unflowered internodes  ———— *P. altiplana*
   5. Leaves erect, linear to oblong; stem unbranched at lower unflowered internodes
   6. Stem and leaves glabrous ———— *P. littoralis*
   7. Stem and leaves hirsute ———— *P. integrifolia*

Fig. 1. Location of *Petunia integrifolia* ssp. *depauperata*, *P. littoralis*, *P. altiplana*, and the taxon from Torres within Brazil.

Fig. 2. Taxonomic key for the identification of *Petunia altiplana*, *P. littoralis*, and *P. integrifolia* ssp. *integri folia* var. *depauperata* (modified from Ando and Hashimoto, 1993, 1994, 1995, and 1996; Smith and Downs, 1966).
Cloning and Sequencing. PCR products were cloned into plasmid DNA using the pGem-T Easy Vector System II (Promega, Madison, Wis.) according to the manufacturer’s protocol (Promega, 2004). Individual white-colored (clones carrying inserts) bacterial colonies were harvested and transferred into 15-mL Falcon snap-top tubes containing 5 mL LB medium supplemented with ampicillin (final concentration of 100 µg·mL⁻¹). The cultures were shaken (225 rpm) at 37 °C for 16 to 18 h. Plasmid DNA was isolated from the bacteria using the RMP Kit (Bio 101, Carlsbad, Calif.) according to the manufacturer’s protocol (Qbiogene, 2004).

Purified plasmid DNA containing cloned PCR products were cyclic sequenced on an ABI Prism 310 Genetic Analyzer using the ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems) in a GeneAmp 2400 PCR System (Perkin-Elmer) using the manufacturer’s protocol (Applied Biosystems, 2004).

DNA Analysis. The intron sequences were aligned by the Clustal method (Higgins and Sharp, 1988) as implemented in the DAMBE software (Xia and Xie, 2001). Repeated sequences were identified using REPutter software (Kurtz and Schleiermacher, 1999). The motif alignment and search tool (MAST) was used to identify motifs within the intron (Bailey and Gribskov, 1998). The E-value of a sequence is the expected number of sequences in a random database of the same size that would match the motif as well as the test sequence does.

Distance analysis was carried out using the PHYLIP software (Felsenstein, 2004). The Kimura two-parameter method (DNADIST program in PHYLIP) was used to calculate pair wise nucleotide divergence values (Kimura, 1980). For each data set, an unrooted neighbor-joining dendrogram was constructed from the distance matrix using the NEIGHBOR program in PHYLIP. Bootstrap values were calculated for 1000 replicates.

Results and Discussion

The most recent taxonomic key (Ando et al., 1995) places P. integrifolia subsp. integrifolia var. depauperata (Pi), P. altiplana (Pa), and P. littoralis (Pl) in the same grouping based upon floral structure. They are distinguished from each other by leaf morphology (Fig. 2). Based upon leaf shape, the taxon from Torres most closely resembles Pl with glabrous and succulent leaves (convergent evolution). Leaf morphology may not accurately represent the true taxonomic relationship of three species. DNA markers may be a more accurate means of defining the taxonomic relationships between these species.

In Petunia x hybrida ‘V30’, there are eight complete copies of the chalcone synthase gene; however, only two (ChsA and ChsJ) are expressed during normal plant development (Koes et al., 1989a). ChsA and ChsJ contain introns that differ in size and are highly divergent in sequence. About 90% of the total Chs mRNA pool is transcribed from ChsA (Koes et al., 1989b). The primers used in this study amplify a single PCR product that was specific expressed within floral tissue (data not shown). Therefore, the PCR product corresponds to ChsA.

The PCR products from Pi, Pa, Pl, and PT were cloned and sequenced. At least three different plants within the individual populations of the species were analyzed. In each instance, there was no difference in the DNA sequence of the ChsA intron between plants within a population.

The intron in Pi was 1275 bp with two different repeated segments of 29 and 38 bp (Figs. 3 and 4). The intron in PT was identical to that in Pi. In contrast, the Pl intron differed from that of Pi with six nucleotide substitutions, two insertions of 1 and 2 bp each, and 1 deletion of 1 bp. Five of the substitutions were transversions and one was a transition.

The Pa intron was very different from Pi with 21 base substitutions, 11 of which were transversions. Besides the substitutions, Pa had three insertions of 1, 2, and 349 bp each and three deletions of 1, 1, and 133 bp each. The 349 bp insertion was within the 133 bp deletion. It is likely that a single event led to deletion/insertion in Pa. A DNA substitution of unequal length would result in a deletion/insertion event. Most of the 349 bp insertion was also present in P. x hybrida (Ph; Fig. 5). The first 33 bp did not match very closely with only 30% homology; however, the remainder of the sequence was highly homologous. The insertion in Ph differed from the Pa insertion by having four substitutions and three insertions of 1, 2, and 3 bp relative to Pa (Fig. 5).

The 29 and 38 bp repeat sequences found in Pa, Pi, and Pl are also in Ph, but not repeated (Fig. 4). Within the 29 bp sequence, Ph differed from Pi in one substitution and one 1 bp insertion. There were four substitutions and one 2 bp insertion within the 38 bp sequence from Ph.

A single motif (TATCACAATTGACTCTTCATT) with an E = 7.4e⁻¹⁰ was identified within both repeated sequences starting at positions 40, 175, 210, and 240. This suggests that these two different repeated sequences may have a common origin. This motif did not correspond to any known transposable element or repeated sequence.

Neither repeated sequences were flanked by an inverted repeat, preventing them from being classified as a miniature inverted repeat (MITE; Feschotte at al., 2002). In addition, they were too short to be considered a short interspersed element (SINE; Feschotte at al., 2002). A potential MITE was found starting at position 44, where an inverted repeat (TTAATTGACTCTTT) was found flanking a 356 bp sequence. Although the flanking invert repeat and the 356 bp sequence did not correspond to any known MITE family, it is still possible that this sequence is a MITE. MITEs have only recently been discovered and new families are still being identified.

In all three species, the intron contained one open reading frame with a TATA box (position 359) and transcription start signal (position 371) as described by Joshi (1987). However, the translation product did not correspond to any known protein.

Based upon the ChsA intron, the Torres taxon most closely resembled Pi. The intron sequences of the two taxa were identical. The unrooted gene tree suggests that Pi is more closely related to Pl than Pa (Fig. 6). The gene trees based upon parsimony and maximum likelihood methods had the same topology (data not presented). Bootstrap values were calculated for 1000 replicates.

Table 1. Leaf morphology of Petunia integrifolia ssp. depauperata, P. littoralis, P. altiplana, and the taxon from Torres. Differences between the means within each column that are followed by the same letter are nonsignificant (Tukey’s HSD test).

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Width/length (cm)</th>
<th>Length (cm)</th>
<th>Trichomes (no./cm²)</th>
<th>Thickness (µm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P. integrifolia</td>
<td>2.3 a</td>
<td>3.5 a</td>
<td>21.0 a</td>
<td>459 a</td>
</tr>
<tr>
<td>P. littoralis</td>
<td>3.2 b</td>
<td>4.1 b</td>
<td>3.0 b</td>
<td>728 b</td>
</tr>
<tr>
<td>P. altiplana</td>
<td>1.3 c</td>
<td>7.1 a</td>
<td>6.0 a</td>
<td>478 ba</td>
</tr>
<tr>
<td>Torres</td>
<td>1.5 d</td>
<td>2.6 b</td>
<td>0 c</td>
<td>879 c</td>
</tr>
</tbody>
</table>
shown). The genetic distance between \( Pl \) and \( Pi \) was 0.0116, as compared with 0.0275 for \( Pi \) and \( Pa \).

The \( Chs \) phylogeny supports the conventional classification of these three species, in that \( Pl \) and \( Pi \) are more closely related than \( Pl \) and \( Pa \) (Fig. 2). \( Petunia altiplana \) was separated from \( Pi \) and \( Pl \) based upon its distinct growth habit and flower structure (Ando and Hashimoto, 1993). Unlike \( Pi \) and \( Pl \), \( Pa \) forms stems with short internodes, lateral branches, and adventitious roots. \( Pa \) produces flowers with a lighter corolla tube than limb, whereas \( Pi \) and \( Pl \) produce flowers with a darker tube than limb. \( Petunia littoralis \) was separated from \( Pi \) based upon its glabrous stems and lanceolate succulent leaves (Smith and Downs, 1966). A principal component analysis of 33 different morphological traits showed that \( Pl \) and \( Pi \) formed a distinct cluster separated from \( P. integrifolia \) subsp. \( integrifolia \) var. \( integrifolia \) (Ando et al., 1995b). However, neither leaf shape nor stem trichome density were included in the traits studied.

Our study suggests that the chalcone synthase intron would be a good candidate for determining the taxonomic relationship of closely related taxa. There are many types of gene mutations (single base pair insertions, deletions, and substitutions; and multiple base pair insertions, deletions, and rearrangements) that occur naturally (Gustafsson and Ekberg, 1977). If a DNA coding sequence is used to study evolutionary relationships, then many of these types of mutations would not be screened because they lead to nonfunctional gene products and are potentially lethal. One of the most common mutations is ultraviolet light-induced deletions at TT sites (Drake, 1970).

In order to study the phylogenetic relationships of closely related species, one should analyze non-coding sequences that have accumulated a wide array of mutations. The difficulty is in finding a highly variable homeologous sequence in multiple species. If this sequence is an intron in a highly conserved gene, then it can be easily found, PCR amplified, and sequenced. Besides chalcone synthase, histones, tubulin, and actins are conserved genes with introns (Doyle et al., 1996; Drouin and Dover, 1990; Wu et al., 2003).

When using conserved genes, one needs to make sure that a pseudogene is not being analyzed. Many conserved genes like \( Chs \) exist as a gene family with several members of the family being nonfunctional paralogous loci (i.e., pseudogenes). In a study of the internal transcribed spacer region (ITS) of \( Quercus \), several independent reports disagreed as to the phylogenetic trees generated from supposedly the same ITS sequence. The difference in the results was found

**Fig. 3 (left).** The DNA sequence of the intron from the chalcone synthase A gene (\( ChsA \)) in \( Petunia integrifolia \) subsp. \( integrifolia \) var. \( depauperata \) (\( Pl \)), \( Petunia \) sp. from Torres (\( PT \)), \( P. littoralis \) (\( Pl \)), and \( P. altiplana \) (\( Pa \)). Only differences from the \( Pi \) sequence are shown; “-” represents nucleotides not present in a particular species.
Fig. 6. An unrooted neighbor-joining dendrogram of sequence in the initial PCR amplification product.

This study demonstrates that sequence heterogeneity of introns may be useful in elucidating the evolutionary relationships between species. By comparing different gene trees, it should be possible to establish an accurate evolutionary tree of a genus. Further studies of additional introns with appropriate outgroups included could be used to develop an evolutionary tree of the *Petunia*.

**Fig. 5.** The sequence of the 177 bp insertion found in *Petunia integrifolia* (Pi) and *P. xyhbrida* 'V30' (Ph). The Ph sequence was reported in Koes et al. (1989). Only differences from the Pi sequence are shown; “-” represents nucleotides not present in a particular species.

**Fig. 4.** A comparison of the two different repeated sequences in *Petunia integrifolia* var. depauperata (P), *Petunia* sp. from Torres (PT), *P. littoralis* (Pl), *P. altiplana* (Pa), and *P. xyhbrida* 'V30' (Ph). The Ph sequence was reported in Koes et al. (1989). Only differences from the Pi sequence are shown; “-” represents nucleotides not present in a particular species.

**Literature Cited**


Benabdelmouma, A., D. Pelletier, C. Humbert, and M. Abirached-Darmency. 1999. Southern hybridization and fluorescent in situ hybridization detect...