Caro et al. 2006

Using a chalcone synthase gene to infer phylogenies in the genus Saintpaulia.

Bios 77: 72-76.

REFNO: 3498

KEYWORDS:

Molecular Systematics, Phylogeny, Saintpaulia

Research Article

Using a *chalcone synthase* gene to infer phylogenies in the genus *Saintpaulia*

Sarah E. Caro, Jessica M. Stampfle, Michael J. Greene, and Michael A. Kotarski

Niagara University, Niagara, New York 14109

Abstract. The plant genus *Saintpaulia* (African violet) is endemic to a small area in northeastern Tanzania and has undergone a recent radiation producing species which are difficult to distinguish morphologically. Previous molecular work identified a group of eleven species of the ionantha complex that are indistinguishable using sequence analysis of the internal transcribed spacer (ITS) of the ribosomal DNA gene. Exon 2 of the chalcone synthase (*CHS*) gene was PCR amplified producing 21 sequences from *Sa. ionantha* and 11 sequences from four other *Saintpaulia* species. The genus *Saintpaulia* contains two *CHS* genes (*SaCHSA* and *SaCHSD*). *SaCHSD* was used to construct a gene tree of six species. The sequence analysis clearly distinguishes four of the five closely related species of the ionantha complex.

Introduction

The chalcone synthase (*CHS*) gene is a small gene family that encodes the first enzyme in the production of flavonoids in green plants. The number of copies of the gene ranges from two in *Zea mays* (Franken et al., 1991) to eighteen in *Dendranthema* (Yang et al., 2002). There are multiple copies in most species except *Arabadopsis* and *Antirrhinum* (Forkman, 1993). There are two exons in the *CHS* gene. Exon 1 has a variable length (111–192 bp) depending upon the species and exon 2 (~1,200 bp) is highly conserved among many different plant genera (Wang et al., 2000) making it useful for evolutionary studies.

Correspondence to: Michael A. Kotarski, Department of Biology, PO Box 2032, Niagara University, Niagara, New York 14109; phone (716) 286–8250; fax: (716) 286–8254; e-mail: mmk@niagara.edu.

The plant genus *Saintpaulia* of the family Gesneriaceae is endemic to northeast Tanzania and coastal Kenya. The genus was first described as a group of nineteen species (Burtt et al., 1958) and later revised to twenty species (Burtt et al., 1964). Little morphological data have been added to this initial description. In 1997, DNA sequences of the internal transcribed spacer (ITS) region of the ribosomal RNA gene were used to produce a molecular phylogeny of seventeen species (Moller and Cronk, 1997). Eleven of these species could not be completely resolved by sequence analysis and are called the ionantha complex.

This study serves to take an alternative approach and examine the DNA sequence of the *CHS* gene in several species of *Saintpaulia*; *S. grandifolia, S. grotei, S. intermedia, S. ionantha,* and *S. orbicularis*. It is our goal to determine the number of *CHS* genes in the genus and use the sequences to construct a relatedness tree of *Saintpaulia*.

Bios 77(3) 72–76, 2006

Materials and Methods

Live plants were obtained from the American Gloxinia and Gesneriad Society members. Two grams of leaf tissue were ground in liquid nitrogen and solubilized in 3.43 M guanidinium thiocyanate, 0.04% polyvinylpyrrolidone, 0.12M beta mercaptoethanol, 1% Sarkosyl and extracted once with chloroform. Nucleic acids were ethanol precipitated, collected by centrifugation, and resuspended in 5 mM EDTA (7.8). RNA was removed by resuspension in guanidinium thiocyanate: 95% ethanol (1:0.75), and DNA was collected by centrifugation and resuspended in Tris-EDTA buffer.

PCR was performed using *Taq* polymerase (Promega Corp., Madison, WI) or iProof (Bio-Rad Laboratories, Inc., Hercules, CA) and two sets of primers. Redundant primers were prepared as outlined in Wang et al. (2000); CCK TCH YTG GAY GCN MGR CAR GAC and GG BCC RAA NCC RAA NAR MAC ACC. These primers correspond to the *CHS* conserved amino acid sequences of PSLDARQD and RVLFG-FGP. Primers specific for *S. ionantha CHS* exon 2 were also prepared; GCC CCT YGC CGG TCG AGC TC and GTG GTA GAG GTY CCC CGG CTC G. DNA was amplified using 34 cycles of 94° C (40 sec), 55° C (60 sec), and 72° C (60 sec) with a 72° C 10 minute final incubation. The amplified fragments are 861 base pairs long and were cloned into the pGEM-T Easy vector (Promega Corp., Madison, WI) or the *Eco* RV digested pBlueScript (Stratagene, Inc., La Jolla, CA) using modified versions of the manufacturers instructions. Both strands of individual clones were sequenced (Northwoods DNA, Inc., Soloway, MN) and aligned using the BCM Search Utilities software. Inconsistencies were corrected by examining sequence chromatograms using Finch TV software (Geospiza Inc., Seattle, WA). Amino acid sequences were produced and aligned using Clustal X software (Thompson et al., 1994). Clustal W and PAUP 4b10 (Swofford, 2003) were used for construction of cladograms. A BLAST search was conducted from the NCBI website to compare *Saintpaulia CHS* sequences among genera.

Results

Sequence analysis

Twenty-one double stranded sequences were obtained for *S. ionantha*. Each sequence is 757 base pairs in length and can be translated into 251 amino acids with *CHS* protein motifs. For the four other *Saintpaulia* species, a total of 11 *CHS* sequences were obtained. Conceptual translation of these sequences results in two distinct protein groups. Within each group the nucleotide sequences are translated into identical amino acid sequences. The protein sequences were aligned using Clustal X 1.83 (Figure 1). SaCHSA has five amino acids different from SaCHSD all near the carboxyl end of the protein. An analysis of the *CHS* sequences from five other *Saintpaulia* species fits this two-gene interpretation.

Eight *SaCHSD* nucleotide sequences representing five *Saintpaulia* species were used to construct a relatedness tree using PAUP 4.0 software. The cladogram shown in Figure 2 groups the species into clades using *Gloxinia lindeniana* and *Antirrhinum majus* as outgroups. The tree was made with 237 steps and bootstrap values

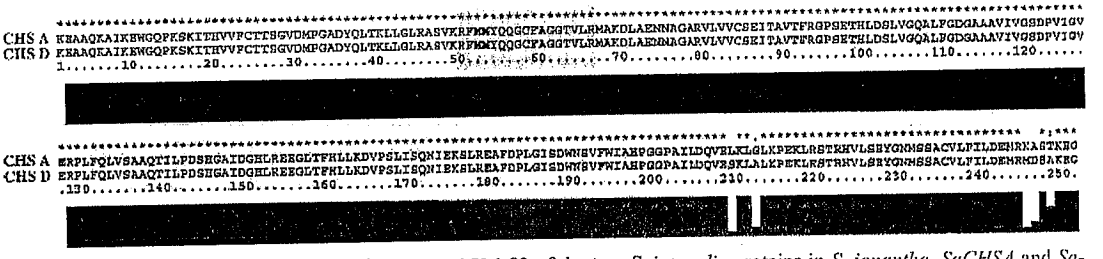

Figure 1. Multiple sequence alignment using Clustal X 1.83 of the two *Saintpaulia* proteins in *S. ionantha, SaCHSA* and *SaCHSD*. The boxed sequence shows the position of the synthase active site which is conserved between the CHS proteins. The shaded bar below the sequence represents the homology of the two proteins. The unshaded vertical bars highlight the amino acid substitutions between the sequences.

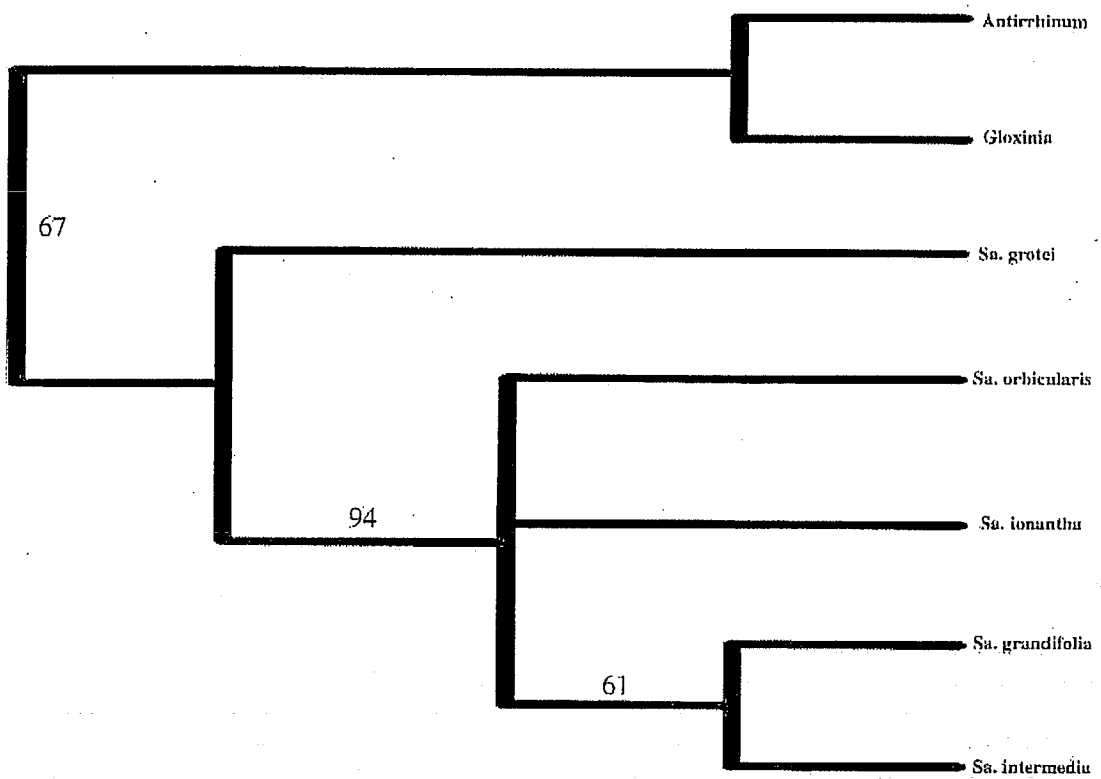

Figure 2. Cladogram derived from the DNA sequence comparison for CHSD from five representative *Saintpaulia* species and two outgroup species, *Gloxinia lindeniana* and *Antirrhinum majus*. The numbers on the cladogram are the bootstrap values which were produced using one thousand replicates.

were calculated using one thousand replicates. Only bootstrap values greater than 50 are used. While four of the five *Saintpaulia* species are resolved an additional node separating *S. ionantha* from *S. orbicularis* collapsed due to a low bootstrap value. The tree is supported by morphological data. *Gloxina* is a rhizadomis new world genus with an inferior ovary in the subfamily Gesneriodiae. The *Saintpaulia* are a fiborous old world genus with a superior ovary in the subfamily Cyntandroidaeae (Moore, 1957). Growth habit is a key characteristic dividing the genus. *Saintpaulia grotei* is well separated from the other *Saintpaulia* species and is the only species with a trailing habit.

Discussion

From the sequence analysis, two distinct *CHS* genes have been identified in the genus *Saintpaulia*. This suggests that an original *CHS* gene duplicated at least once and the gene copies have diverged by accumulating mutations. *CHS* gene duplication and divergence to produce a small gene family has been used to explain the multiple copies in *Dendranthema* (Yang et al., 2002), *Gerbera* (Helriutta et al., 1996), *Ipomea* (Durbin et al., 1995), *Petunia* (Koes et al., 1989), and leguminous plants (Ryder et al., 1987; Wingerder et al., 1989; Ito et al., 1997). The *Saintpaulia* SaCHSA protein is most similar to *Ipomea purpurea CHSD* (88% identity) and *Gerbera hybrida CHS1* (87% identity). SaCHSD is most similar to *Ipomea purpurea CHSD* (88% identity) and *Gerbera hybrida CHS1* (74% identity).

The amino acid substitutions in the amplified portion of the exon are all between amino acids 210–250, the carboxy terminal one-third of the protein. The consensus sequence of the CHS active site is R_MMY_QGCFAGG_VLR (Contessotto et al., 2001), which is conserved in the *Saintpaulia* CHS proteins and includes amino

acids 50–66 (Figure 1). This observation suggests that the *Saintpaulia* CHS proteins have not diverged to the point of acquiring new enzymatic functions as suggested in *Gerbera CHS2* (Helariutta et al., 1996) and *Saccharum CHS2* (Contessotto et al., 2001).

Arabidopsis thaliana and *Antirrhinum majus,* a close relative of *Saintpaulia*, diverged 120 million years ago. An average nucleotide substitution rate of 2.26×10^{-9} substitutions/site/year (r) can be calculated for *CHS* in these species. The *Saintpaulia* nucleotide sequence data can be used to estimate the time of divergence of *SaCHSA* and *SaCHSD*. The number of substitutions per site between the two genes (K) is 0.085 and the estimated divergence time (K/2r) is 18.8 million years. This divergence is similar to the *Ipomoea purpurea CHSA* and *CHSB* genes that have an estimated divergence of 21 million years (Durbin et al., 1995). In a later report, Durbin estimates that the divergence time between two less related *CHS* subgroups of *Ipomoea* (*CHSA-B* to *CHSD-E*) to be 146 million years (Durbin et al., 2000). In *Saintpaulia,* there is no compelling evidence to suggest more than two genes and there are no data yet to support the existence of more divergent subgroups.

The nucleotide sequences of *SaCHSD* used to construct the relatedness tree in Figure 2 can also be used to calculate the time of divergence of the five *Saintpaulia* species using the average substitution rate and time of divergence of *Arabidopsis* and *Antirrhinum*. The two most disparate *Saintpaulia* species (*S. grotei* and *S. grandifolia*) have been separated for an estimated 12.1 million years and the most closely related species (*S. ionantha* and *S. orbicularis*) have been separated by approximately 4.7 million years. A previous report using ITS sequence data suggests that *S. intermedia* was the most widely separated species from *S. ionantha* (Moller and Cronk, 1997). This observation is not confirmed here. The ITS data fail to resolve species within the ionantha complex, a group of 11 species including *S. grandifolia, S. grotei, S. ionantha,* and *S. orbicularis*. Presumably, this is due to the high copy number of ITS sequences and the slow spread of copies with nucleotide substitutions within the genome. By using the single copy nuclear gene, *SaCHSD*, the species *S. grotei, S. intermedia,* and *S. grandfolia,* are separated from the ionantha complex and a more accurate description of the *Saintpaulia* genus can be made.

Literature Cited

Burtt, B.L. 1958. Studies in the Gesneriaceae of the Old World XV: The genus Saintpaulia.—Notes of the Royal Botanic Garden Edinburgh **22**(6):547–568.

Burtt, B.L. 1964. Studies in the Gesneriaceae of the Old World XXV: Additional notes on Saintpaulia.—Notes of the Royal Botanic Garden Edinburgh **25**(3):191–195.

Contessotto, M.G.G., C.B. Monteiro-Vitorello, P.D.S.C. Mariani, L.L. Coutinho. 2001. A new member of the chalcone synthase (CHS) family in sugarcane. *Genetics and Molecular Biology.* **24**:257–261.

Durbin, M.L., B. McCaig, M.T. Clegg. 2000. Molecular evolution of the chalcone synthase multigene family in the morning glory genome. *Plant Molecular Biology.* **42**:79–92.

Durbin, M.L., G.H. Learn JR., G.A. Huttley, M.T. Clegg. 1995. Evolution of the chalcone synthase gene family in the genus *Ipomoea*. *Proc. Natl. Acad. Sci.* USA **92**:3338–3342.

Forkman, G. 1993. in The Flavonoids: Advances in Research Since 1986. pp. 537–564, Harborne, J.B. (ed.), Chapman & Hall, London.

Franken, P., U. Niesbach-Klosgen, U. Weydemann, L. Marechal-Drouard, H. Saedler, U. Wienand. 1991. The duplicated chalcone synthase genes C2 and Whp (white pollen) of *Zea mays* are independently regulated; evidence for translational control of Whp_expression by the anthocyanin intensifying gene. *EMBO J.* **10**:2605–2612.

Helariutta, Y., M. Kotilainen, P. Elomaa, N. Kalkkinen, K. Bremer, T.H. Teeri, V.A. Albert. 1996. Duplication and functional divergence in the chalcone synthase gene famile of Asteraceae: Evolution with substrate change and catalytic simplification. *Proc. Natl. Acad. Sci.* **93**: 9033–9038.

Ito, M., Y. Ichinose, H. Kato, T. Shiraishi, T. Yamada. 1997. Molecular evolution and functional revelance of the chalcone synthase genes of pea. *Mol. Gen. Genet.* **255**:28–37.

Koes, R.E., C.E. Spelt, P.J. van den Elzen, J.N.M. Mol. 1989. Cloning and molecular characterization of the chalcone synthase multigene family of *Petunia hybrida*. *Gene* **81**: 245–257.

Moller, M., W.C.B. Cronk. 1997. Origin and relationships of *Saintpaulia* (Gesneriaceae) based on ribosomal DNA internal transcribed spacer (ITS) sequences. *American Journal of Botany* **84**(7):956–965.

Moller, M., W.C.B. Cronk. 1997. Phylogeny and disjunct distribution: evolution of *Saintpaulia* (Gesneriaceae). *Proc. R. Soc. Lond.* **B 264**:1827–1836.

Moore, H.E., Jr., 1957. African Violets, Gloxinias, and their relatives. A Guide to the Cultivated Gesneriads. MacMillian Co., New York.

Ryder, T.B., S.A. Hedruck, J.N. Bell, X. Liang, S.D. Clouse, and C.J. Lamb. 1987. Organization and differential activation of a gene family encoding the plant defense en-

zyme chalcone synthase in *Phaseolus vulgaris*. *Mol. Gen. Genet.* **210**:219–233.

Swofford, D.L. 2003. PAUP. Phylogenetic Analysis Using Parsimony (and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**:4673–4680

Wang, J.L., L. Qu, J. Chen, H. Gu, Z. Chen. 2000. Molecular evolution of the exon 2 of CHS gene and the possibility of its application to plant phylogentic analysis. *Chinese Sci. Bull.* **45**:1735–1742.

Wingender, R., H. Rohrig, C. Horicke, D. Wing, J. Schell. 1989. Differential regulation of soybean chalcone synthase genes in plant defense, symbiosis and upon environmental stimuli. *Mol. Gen. Genet.* **218**:315–322.

Yang, J., J. Huang, H. Gu, Y. Zhong and Z. Yang. 2002. Duplication and adaptive evolution of the Chalcone synthase genes of *Dendranthema (Asteraceae)*. *Mol. Biol. Evol.* **19**(10):1752–1759.

Received 17 October 2005; accepted 5 July 2006.