WWU Graduate School Collection   WWU Graduate and Undergraduate Scholarship

2012

TTS protein orthologs as interspecific reproductive barriers in the solanaceae

Tara D. Callaway
Western Washington University

Follow this and additional works at: https://cedar.wwu.edu/wwuet

Part of the Biology Commons

**TTS PROTEIN ORTHOLOGS AS INTERSPECIFIC REPRODUCTIVE BARRIERS
IN THE SOLANACEAE**

By

Tara D. Callaway

Accepted in Partial Completion
Of the Requirements for the Degree
Master of Science

Kathleen L. Kitto, Dean of the Graduate School

ADVISORY COMMITTEE

Chair, Dr. Anu Singh-Cundy

Dr. Lynn Pillitteri

Dr. Dietmar Schwarz

MASTER'S THESIS

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Western Washington University, I grant to Western Washington University the non-exclusive royalty-free right to archive, reproduce, distribute, and display the thesis in any and all forms, including electronic format, via any digital library mechanisms maintained by WWU.

I represent and warrant this is my original work, and does not infringe or violate any rights of others. I warrant that I have obtained written permissions from the owner of any third party copyrighted material included in these files.

I acknowledge that I retain ownership rights to the copyright of this work, including but not limited to the right to use all or part of this work in future works, such as articles or books.

Library users are granted permission for individual, research and non-commercial reproduction of this work for educational purposes only. Any further digital posting of this document requires specific permission from the author.

Any copying or publication of this thesis for commercial purposes, or for financial gain, is not allowed without my written permission.

Tara Callaway
May 3, 2012

**TTS PROTEIN ORTHOLOGS AS INTERSPECIFIC REPRODUCTIVE BARRIERS
IN THE SOLANACEAE**

A Thesis
Presented to
The Faculty of
Western Washington University

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science

By

Tara D. Callaway
May 3, 2012

ABSTRACT

TTS (transmitting tissue-specific) proteins are abundant in the extracellular matrix (ECM) of the transmitting tissue, which forms the pollen tube pathway in *Nicotiana* pistils. These arabinogalactan proteins stimulate pollen tube growth and are vital for optimal seed set. I have cloned and sequenced two putative orthologs, *Pi*PRP1 and *Pa*PRP1, which are expressed in the pistils of *Petunia integrifolia* and *Petunia axillaris*, respectively. Comparison of the domain architecture and cross-reactivity with anti-TTS protein antibodies confirm that the proteins encoded by these *Petunia* cDNA clones are orthologs of TTS proteins (TTSPs) from *Nicotiana* species. Using immunological detection methods, I have shown that TTSP orthologs are present in the pistils of three subfamilies within the Solanaceae: Nicotianoideae, Petunioideae, and Solanoideae. Surprisingly, the proteins were also detected in leaves and roots of *P. integrifolia* seedlings. I cloned the TTSP ortholog expressed in seedling leaves (*Pi*PRP2) and found it to be nearly indistinguishable from *Pi*PRP1, encoded by the pistil cDNA. Like the TTSPs from *Nicotiana*, *Pa*PRP1, *Pi*PRP1, and *Pi*PRP2 are histidine-domain arabinogalactan proteins with a highly variable proline-rich domain containing KPP repeats that vary in number and location among solanaceous taxa.

Multiple alignments were used to deduce the effect of natural selection on the conserved and hypervariable domains of this multidomain subfamily of arabinogalactan proteins. For each pairwise comparison, I deduced the Ka/Ks ratio, which expresses the nucleotide substitutions per synonymous site (Ks) and non-synonymous site (Ka) in the two sequences. My analysis indicates that the two hypervariable domains of these proteins have

iv

undergone positive selection (Ka/Ks>1), whereas the conserved domains are under purifying selection (Ka/Ks<1). The differential selective pressure on the protein domains suggests that the hypervariable domains are involved in species-specific interactions with an unidentified pollen tube partner, and the conserved domains have general functions that are invariant. I propose that sequence divergence in the hypervariable domain reinforces speciation by generating a post-pollination prezygotic breeding barrier between incipient species.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Angiosperms have evolved a variety of reproductive isolating mechanisms that maintain species boundaries by blocking hybridization between sympatric sister species. Interspecific hybrids are predicted to display intermediate traits that have lower fitness if the two parental species are highly adapted to different ecological niches (Rieseberg and Willis 2007). To prevent the loss of reproductive potential to maladaptive hybrids, natural selection favors reduced gene flow between recently diverged or emerging species through reinforcement (Hopkins and Rausher 2012). Reinforcement, natural selection that strengthens genetic isolation by generating mating barriers, reduces gene flow between naturally occurring sympatric species (Dobzhansky 1940, Grant 1965, Rieseberg and Willis 2007, Yost and Kay 2009). However, the molecular mechanisms underlying such reinforcement of speciation are poorly understood, especially in plants.

My thesis research focuses on a reproductive gene, *TTSP* (transmitting tissue-specific protein), which may account for the gametic barriers known to exist between closely related species of *Petunia* and may be implicated in species barriers between other solanaceous taxa as well. To investigate the role of this class of reproductive protein in the Solanaceae, I used immunological techniques to look for the protein in the female reproductive tissues (pistils) of representative taxa. I found putative TTSP orthologs in three of the five subfamilies I surveyed. In the course of this analysis, I made the remarkable discovery that these proteins are not confined to the transmitting tissue of the pistil as has been reported for the *Nicotiana* TTSPs (Cheung *et al.* 1993), but that the *Petunia* orthologs are also expressed in roots and leaves of seedlings. The same gene, as opposed to a different paralog, appears to be expressed in the pistils and the vegetative tissue of seedlings. I inferred this from the cDNAs

I cloned from vegetative tissue of *P. integrifolia* seedlings: the pistil and vegetative tissue cDNA are identical except for a single synonymous substitution that most likely represents population-level allelism.

To understand evolutionary divergence among solanaceous TTSP orthologs, I cloned the cDNA encoding the TTSP ortholog, *Pa*PRP1, in *P. axillaris* pistils and the cDNA encoding the orthologous protein, *Pi*PRP1, in *P. integrifolia* pistils. Sequence comparisons between *Pa*PRP1 and *Pi*PRP1, and five other TTSP orthologs, revealed that the proteins evolve rapidly. Comparison of synonymous and non-synonymous substitutions in pairwise comparisons of eight solanaceous TTSP orthologs revealed that the two hypervariable domains (HVI and HVII) are under positive selection, while the signal and conserved domains (CVI and CVII) of the protein show strong evidence of being subjected to purifying selection. In contrast to the sequence variability seen in HVI and HVII among TTSP orthologs from different species, I found very little population-level polymorphism along the length of the entire protein among the different accessions of *Petunia* available to me (Figure 1-2).

I propose that the hypervariable domains of TTSP orthologs are critical for species-specific interaction with a pollen ligand, and that any genetic lesion that disrupts this interaction leads to reduced fertility or complete sterility. The functional constraints operate over much of the gene sequence to keep within-population polymorphisms to a minimum. However, after an initial speciation event, there is strong selection for sequence divergence in HVI or HVII domains of the proteins expressed in the pistils of the diverging incipient species. Sequence divergence in these domains would disrupt mate recognition between the

2

two populations, generating a gametic barrier that would reinforce isolation and strengthen speciation. This working model implies that evolutionary change in solanaceous TTSP orthologs occurs in bursts, followed by periods of evolutionary stasis that lasts as long as the species does. The footprint of speciation events in the Solanaceae should therefore be recorded in the phylogeny of TTSP orthologs.

Part I. Reproductive Barriers in the Solanaceae

Researchers have been studying pollen-pistil interactions in the Solanaceae for many decades with a view to deciphering the limitations that prevent closely related species from hybridizing, with the focus on several tobacco, petunia, tomato, and potato species (Cheung *et al.* 1995, Hiscock and Allen 2008, Lowry *et al.* 2008). The Solanaceae family consists of over 3,000 species in 99 recognized genera (Knapp 2010) and includes important crop plants (e.g. tomato, potato, eggplant, tobacco, and peppers), ornamentals (e.g. *Petunia*, *Nicotiana*, and *Calibrachoa*), and medicinal plants (such as *Datura* and *Capsicum*) (Hawkes 1999). I chose to focus on two of the 16 species out of the solanaceous genus *Petunia* (2n=14) that are known to occur sympatrically in South America yet rarely interbreed in their native habitats (Lorenz-Lemke *et al.* 2006). My research focuses on orthologs of *TTSP,* a gene that has been demonstrated to play a key role in facilitating pollen tube growth in tobacco (Cheung *et al.* 1993, Cheung *et al.* 1995, Wang *et al.* 1993, Wu *et al.* 1995). My working model is that sequence divergence in the *TTSP* orthologs can generate gametic barriers that reinforce speciation by preventing hybridization between emergent sister species.

3

Pollen-pistil interactions in the Solanaceae

In solanaceous species, pollen is received on a wet-type stigma that exudes copious amounts of a sticky fluid rich in carbohydrates and unevenly covered by a waxy layer (Heslop-Harrison 1981, Heslop-Harrison and Shivanna 1977). Pollen grains become hydrated as they absorb water, and the hydrated grain produces a cellular extension, the pollen tube, in a process known as pollen germination (Lord and Russell 2002, Swanson *et al.* 2004). Pollen tubes grow between the cells of the solid style through the extracellular matrix (ECM), which is rich in pectins, proteins, and nutrients such as sugars and amino acids (de Nettancourt 2001, Herrero and Dickinson 1979, 1981, Heslop-Harrison 1987, Lord and Sanders 1992, Sanders and Lord 1989). Optimal pollen tube growth is known to be critically dependent on the complex ECM environment as pollen tube growth *in vivo* is five-to-ten times greater than in the best *in vitro* media that have been developed (Lee *et al.* 2008)

Pollen tube growth in the pistil is biphasic: the tubes extend relatively slow at first, and then exhibit rapid growth as they enter the upper portion of the style, a region called the transition zone (Herrero and Dickinson 1980, Lubliner *et al.* 2003, Mulcahy and Mulcahy 1982, Singh *et al.* 1992). The first, slow phase of growth is largely autonomous and fueled mostly by the food storage reserves held within the pollen grain; as the pollen tubes enter the transition zone, the growth rate increases as much as two-to-five fold (Lubliner *et al.* 2003). This second, rapid phase of growth appears to have an obligate dependence on the pistil environment because pollen tubes grown *in vitro* do not exhibit this second phase of growth nor do they ever reach the length typical of those grown *in vivo* (Lee *et al.* 2008, Lubliner *et al.* 2003, Sanchez *et al.* 2004). An approximation of the pre-transition growth phase is seen

pollen cultured in an artificial medium with mineral nutrients, amino acids, and sucrose; however, simulating the post-transition rapid growth phase *in vitro* remains an elusive goal (Wu *et al.* 2000). The pollen tube elongates and travels the distance from the stigmatic surface to the ovary, enters the embryo sac in an unfertilized ovule (by-passing any previously fertilized ovules), and bursts to release the male gametes for double fertilization (Clark *et al*. 2006, Heslop-Harrison 1987, Hiscock and Allen 2008).

Reproductive isolating barriers in plants

Speciation, the process that generates sister species from an ancestral population, is spurred by the formation of barriers that prevent gene flow between two groups that could interbreed previously (Coyne and Orr 1998, Coyne and Orr 2004, Rieseberg and Willis 2007). When pollen from one species fails to pollinate and fertilize the pistil of a closely related species, the two species have a barrier to gene flow known as incongruity (Hogenboom 1975). Interspecific or intergeneric incongruity is a consequence of reproductive isolating barriers that are either prezygotic or postzygotic in their timing (Lowry *et al.* 2008). Postzygotic isolation acts after fertilization and encompasses isolating mechanisms such as hybrid inviability, hybrid sterility, reduced fitness in subsequent generations, and decreased hybrid vigor (Coyne and Orr 2004, Dobzhansky 1937, Mayr 1942).

There are prezygotic and postzygotic barriers that have the same individual strength, but collectively, prezygotic barriers were found to be twice as powerful as postzygotic barriers in angiosperms (Lowry *et al*. 2008). Early-acting barriers are more powerful in terms of total isolation because reproductive barriers act sequentially and cumulatively to prevent a

greater amount of gene flow than late-acting barriers (Rieseberg and Willis 2007). In *Petunia* specifically, prezygotic reproductive barriers have been shown to be a much more powerful force driving reproductive isolation than postzygotic reproductive barriers and ultimately make a larger contribution to the total reproductive isolation between taxa (Dell'Olivo *et al.* 2011, Lowry *et al.* 2008).

Prezygotic isolation barriers act prior to fertilization by preventing the delivery of male gametophytes to female gametophytes and can be further divided by defining whether the barrier is premating or postmating (Christie and Macnair 1987, Grant 1981, Mayr 1963). Premating barriers prevent pollen from being deposited on the stigma, and postmating barriers act after pollen has been deposited on the stigma but before an egg is fertilized. Premating barriers include isolating events that are spatial, temporal, and ethological, while major postmating barriers include pollen competition and gametic incongruity (Coyne and Orr 1998, Dell'Olivo *et al.* 2011, Rieseberg and Willis 2007, Yost and Kay 2009).

Temporal isolation acts as a premating barrier in plants when a mismatch between the flowering schedules of two populations of closely related species results in a failure to reproduce (Coyne and Orr 2004). Ethological isolation is another premating barrier that has been extensively studied in angiosperms and is predominantly pollinator-driven (Mather and Edwardes 1943). Co-evolution between pollinator and plant pollination syndromes can lead to highly specialized relationships that isolate plant populations that do not share pollinators. A specific suite of floral traits determines whether flowers are pollinated abiotically (wind or water) or biotically (animals); if pollination occurs biotically, floral traits can restrict pollen delivery to a particular animal pollinator (birds, bees, beetles, moths, or ants) (Hiscock and

6

Allen 2008). Much of the interspecific gene flow can be stopped if species advertise to different pollinators through the use of various corolla color and size, nectar source, and scent.

If the premating barriers fail to completely isolate species reproductively, postmating barriers can act by inhibiting the delivery of the male gametes to the ovary through such mechanisms as conspecific pollen precedence, pollen competition, pollen attrition, and/or gametic incongruity (Dell'Olivo *et al*. 2011, Rieseberg and Willis 2007, Yost and Kay 2009). Postmating barrier strength was shown to be higher between recently diverged species in close proximity than in taxa separated by greater distances. This is due to reinforcement in tension zones, or overlapping regions, where the recently diverged species co-occur (Butlin 1987, Kay and Schemske 2008). Reinforcement, a term coined by Alfred Russell Wallace, encompasses novel prezygotic barriers that prevent hybridizations between sympatric populations of newly formed species or diverging populations (1889). The Wallace effect, as reinforcement is now known, describes selection against low-fitness interspecific hybrids by reinforcing isolation after speciation has begun (Hopkins and Rausher 2012). Closely related species can remain as separate species in sympatry by forestalling gene flow and hybridization.

If premating reproductive barriers fail and pollen transfer between taxa occurs, postmating reproductive barriers will rely on the mate-recognition system involved in the pollen-pistil interaction. Such a system is able to differentiate between conspecific and heterospecific pollen and may involve one to several genes. Divergence in any one of these genes can create and/or reinforce speciation between populations (Yost and Kay 2009).

7

Identifying a single gene that functions in reinforcement of species isolation, especially one that is part of a mate-recognition system, has been a major challenge in plant evolutionary biology (Rieseberg and Blackman 2010, Rieseberg and Willis 2007, Yost and Kay 2009). In fact, only one postmating prezygotic gene, the S-RNase, has been identified and functionally verified as a gene that may contribute to reproductive isolation (Nasrallah 2002).

Natural selection on gene domains

Natural selection can favor variation or constancy of a gene, or domains within a gene, as a consequence of selective pressure. Gene structure, function, and nucleotide sequence tends to be conserved between extremely diverged taxa if change is deleterious and therefore selected against. Genes can also be under selective pressures to change, as is well-known for the major histocompatibility genes (MHC genes) in mammals (Yuhki and O'Brien 1990). The influence of natural selection on DNA sequences can be inferred from pairwise comparisons of the nucleotide sequences. The Ka/Ks ratio can be employed to determine the type of selection that is at work: whether the gene is undergoing positive or purifying selection or if it is under neutral evolution (Clark *et al*. 2006, Hurst 2002). Natural selection favoring nucleotide substitutions may produce synonymous (silent) substitutions or non-synonymous substitutions. Synonymous substitutions change a nucleotide that does not alter the amino acid sequence, while non-synonymous substitutions change a nucleotide or nucleotides that alter the amino acid encoded and so change the coding information (Higgs and Attwood 2005). If the Ka/Ks rate is greater than one, than positive selection is acting and that signifies that natural selection favors amino acid diversity at that nucleotide site, gene domain, or gene (Higgs and Attwood 2005, Hurst 2002). If the site is under purifying

selection, the Ka/Ks ratio is less than 1, signifying that changes have deleterious consequences and are selected against.

Positive selection in a gene region can indicate a functionally important gene domain whose rapid divergence over evolutionary time is favored because it generates new advantageous phenotypes, such as resistance to a new strain of pathogen (Swanson and Vacquier 2002). Selecting for change in the amino acid sequence of a reproductive protein, such as TTSP, may confer an advantage, such as species-specificity in mating interactions, especially as sister species diverge in sympatry (Swanson *et al.* 2004). Though it is often difficult to demonstrate, positive selection has been detected in plant genes involved in defense, reproduction, electron transport and ATP synthesis, and cytokine synthesis and metabolism (Yang and Bielawski 2000). Obtaining two closely related sister species enables researchers to calculate the Ka/Ks ratios in each domain to determine the type of selection experienced by the gene locus in question over the evolutionary history of the taxa. Positive selection could be a sign of a rapidly evolving gene domain, and in a reproductive protein, that might indicate a role in species-specific mating barriers.

Reproductive barriers in *Petunia*

Subspecies of *Petunia axillaris* and *P. integrifolia* have partially overlapping distribution and appear to have diverged through sympatric speciation (Stehmann *et al*. 2009). The subspecies of these two taxa form two distinct sister clades that exhibit marked difference in morphology and anatomy, especially with respect to floral traits (Stehmann and Semir 1997). *P. integrifolia* (subspecies *integrifolia* and *inflata*) produces small, purple

flowers that are pollinated during the day by bees (mainly *Leioproctus* subgen. *Hexantheda*) while the sister species, *P. axillaris* (subspecies *axillaris*, *parodii*, and *subandina*), has a large, white, UV-reflecting floral tube and the flowers are pollinated at night by hawkmoths (*Manduca sexta*, *Manduca contracts*, and *Manduca diffissa*) (Ando *et al.* 2001, Hoballah *et al.* 2005, Stehmann *et al*. 2009). The incipient speciation event between these two species is thought to be caused by the shift in flower color from purple-red to white, which has independently transitioned more than once in *Petunia,* by a loss-of-function mutation in *AN2* (Widmer *et al.* 2009). Altering the flower color dramatically decreases gene flow between populations by inducing a pollinator shift from predominantly bee to hawkmoth. Even though pollinator isolation between the two sister species is very strong, it cannot account for the complete absence of hybrids in nature (Dell'Olivo *et al*. 2011).

Some reports in the literature report complete gametic isolation between *P. integrifolia* and *P. axillaris*, while other researchers describe unilateral incongruity between the two species. In unilateral incongruity (also called unilateral incompatibility by some writers), an interspecific cross yields full seed set in one direction, but the reciprocal cross fails completely. In many, but not all, cases of unilateral incongruity, the SI X SC rule operates: the self-incompatible species rejects pollen from a self-compatible sister species, but the reciprocal cross is successful (Cruz-Garcia *et al*. 2003, de Nettancourt 2001).

In the accessions used by Sink (1984), manual pollinations in either direction were unsuccessful. However, Ando *et al.* (2001) report unilateral incongruity between the two species in accordance with the SI X SC rule. *P. integrifolia* is an obligate outbreeder, with strong self-incompatibility (SI), while *P. axillaris* is largely self-compatible (SC) (Stehmann

and Semir 1997). In the accession used by Ando *et al.* (2001), fruit development was observed when *P. axillaris* pistils were manually pollinated with *P. integrifolia* pollen, but *P. axillaris* pollen failed to fertilize *P. integrifolia* pistils because such pollinations did not yield any capsule formation.

The self-incompatibility response has been implicated as the main agent of pollination failure in interspecific crosses that follow the SI X SC rule, especially in tomato and *Nicotiana*. For example, Cruz-Garcia *et al*. (2003) have shown that S-RNases were causally involved in the rejection of pollen from self-compatible *Nicotiana* species on pistils of self-incompatible species. Others have suggested that in *Petunia*, the pollen of a SC species, like *P. axillaris,* fails to recruit optimal support in the pistils of a SI sister species, *P. integrifolia* (Rieseberg and Willis 2007). Mather and Edwardes (1943) suggested that "genes discouraging foreign pollen occur in both species" of *Petunia*. Although the underlying mechanisms are poorly understood, it is clear that more than one cellular-biochemical mechanism could generate postmating pre-fertilization gametic barriers between solanaceous taxa.

Molecular mechanisms underlying interspecific reproductive barriers

There is evidence that pistil tissues exert selective pressure on pollen genotypes for optimal pollen performance and this phenomenon is known as pollen competition (Hormazo and Herrero 1996, Ottaviano *et al*. 1980, Spira *et al*. 1992). Maternal tissues exert strong control on the ability of pollen to navigate the stigma and style to reach the ovary ensuring only pollen with optimal fitness is able to successfully fertilize, while the pistil genotypes are

11

likewise under selective pressure for optimal fitness (Hiscock and Allen 2008). The outcome of this intense pollen-pistil crosstalk is manifested as pollen competition, sexual selection, and unilateral or bilateral incongruity (McClure 2009). Many pollen-specific and pistil-specific genes have been described (Cheung *et al.* 1993, Goldman *et al.* 1992, Lind *et al.* 1994, Wang *et al.* 1993, Wu *et al.* 1993) and their gene products have been shown to interact in specific ways, but the link, if any, between these gene systems and prezygotic barriers remains elusive.

Species within the Solanaceae family are often found in close proximity to one another, yet they are morphologically and genetically distinct (Ando *et al.* 2001). It has been hypothesized that a functional mismatch between the male and female partner would create genetically-isolated subpopulations and thereby drive sympatric speciation (Rieseberg and Willis 2007). Such divergences in mate-recognition could also reinforce, at the genetic level, barriers to gene flow that result from ecological adaptation, such as those driven by pollinator isolation. The adaptive value of genetic isolation between sibling species is that it would forestall stray hybridizations that would otherwise waste reproductive resources by generating interspecific hybrids (Coyne and Orr 1998). Most interspecific hybrids have lower fitness than their parents because of maladaptive intermediate traits (Rieseberg and Willis 2007). Intermediate floral traits, for example, would not have the specificity of either parental pollinator due to intermediate flower color, size, bloom time, or scent.

Part II. Role of the TTS Proteins within the Pistil

TTS protein in the extracellular matrix of the pistil

The TTSP orthologs described so far—from *Nicotiana* species by Cheung *et al.* (2000) and from *Petunia hybrida* by Twomey (2012)—are localized in an area of the pistil called the extracellular matrix (ECM), a continuum of extracellular surfaces within the transmitting tissue (Cheung *et al.* 1993, Cheung *et al.* 1995, Wang *et al*. 1993). The ECM of the transmitting tissue (TT) is thought to deliver nutrition, guidance, structural support, and protection to the growing pollen tubes and it forms the microenvironment in which pollen tubes interact with maternal tissues (de Nettancourt 2001, Herrero and Dickinson 1979, Heslop-Harrison 1987, Lord and Sanders 1992, Sanders and Lord 1989).

Pollen tubes grown *in vitro* in a culture medium without any supplements from the ECM exhibit random directionality, lower growth rates, and their terminal length is substantially less compared to *in vivo* pollen tubes (Cheung *et al.* 1995, Wu *et al.* 2000). Purified *Nicotiana* TTSPs have been shown to enhance pollen tube growth and attract pollen tubes from a distance in a sugar-free medium (Cheung *et al.* 1995, Wu *et al.* 2000). Transgenic *N. tabacum* plants with antisense suppression of TTSP, resulting in greatly reduced levels of the protein, had diminished pollen tube growth and reduced female fertility (Cheung *et al.* 1995). The adhesive nature of the TTSP combined with the physical and biochemical interactions observed between the TTSP and pollen tubes has led to the speculation that TTSPs serve as both a nutrient source and surface adhesive for growing pollen tubes (Cheung 1995, Cheung 1996, Cheung *et al.* 1995, Twomey, Master of Science thesis, 2012, Wu *et al.* 1995, Wu *et al.* 2000).

13

TTSPs and their orthologs contain a conserved C-terminal *Ole* e 1 domain that interacts with pollen proteins and is believed to contribute to their ability to interact with pollen tubes (Cheung *et al.* 1995). Growth enhancement and attraction of pollen tubes may also be due to the fact that more than 50% of the molecular mass of the known TTSPs comes from carbohydrates linked to the protein backbone by O-glycosylation. TTSPs may be deglycosylated by hydrolases associated with pollen tube tips (Cheung and Wu 1999, Wu *et al.* 1995). These researchers propose that TTSPs adhere to growing pollen tube tips and enzymes bound to the pollen plasma membrane or cell wall deglycosylate the proteins (Wu *et al.* 1995); the oligosaccharides released may serve as a nutritional resource.

Defining characteristics of the TTS protein

Cheung *et al.* (1993) and Chen *et al.* (1993) cloned and characterized three TTSP sequences: TTS-1 (accession Z16403.1) and TTS-2 (accession Z1604.1) from *Nicotiana tabacum*, and *Na*TTS (accession X70441.1) from *N. alata*. Twomey (2012) cloned and characterized a TTSP ortholog from *Petunia hybrida*: *Ph*PRP1 (accession FJ719032.1) and identified a TTSP homolog, AGP31 (accession Q9FZA2) from *Arabidopsis thaliana* (Liu and Mehdy 2007).

The three *Nicotiana* TTSPs (TTS-1, TTS-2, and *Na*TTS) are described as transmitting tissue arabinogalactan proteins. *N. tabacum* is an allopolyploid species (its chromosomes are derived from two different species, one of which is known to be *N. sylvestris*), and two homologous genes (TTS-1 and TTS-2) are expressed in its pistil tissue. In contrast, its diploid relatives *N. sylvestris* and *N. tomentosiformis,* have a single *TTSP* gene (Cheung *et al.* 1993). The putative TTSP ortholog from *C. annuum*, *Ca*PRP1, was isolated from five-day-old roots

of *Capsicum annuum* (Mang *et al.* 2004). *AGP31*, a presumed TTSP ortholog from *A. thaliana,* is expressed in seedling roots and leaves, and also in pistils (Liu and Mehdy 2007).

Based on denaturing SDS-PAGE, the apparent molecular weight of *Ph*PRP1 ranges from 55 kDa to 100 kDa and has a predicted molecular mass of 27.4 kDa (Twomey, Master of Science thesis, 2012). *Ph*PRP1 contains 18 potential O-glycosylation sites with only one N-glycosylation site (Twomey, Master of Science thesis, 2012). In comparison, the molecular weight of TTSPs from *Nicotiana* ranges from 45-105 kDa but enzymatic deglycosylation reduces the mass to about 30 kDa (Wang *et al.* 1993). Chemical deglycosylation of the 30 kDa *Na*TTS reduces the mass to 27 kDa, the value predicted from the amino acid sequence (Wang *et al.* 1993).

Based on multiple alignments of the different TTSP sequences and their orthologs, I have characterized five distinct domains in TTSPs (Figure 3): Signal Domain I (residues 1-22), HVI (residues 23-40), CVI (residues 41-62), HVII (residues 63-109), and CVII (110-257). Signal Domain I is the most conserved of all domains followed by the extremely diverse HVI. CVII is a Pro-rich domain and downstream the larger HVII contains several Lys-Pro-Pro (KPP) motifs that vary in position and number among the TTSP orthologs from various species. *Ph*PRP1 has nine interspersed KPP repeats and the HVII contains seven of these nine KPP repeats. The remaining two KPP repeats are found in the C-terminal region of CVII.

Analysis of the overall amino acid composition of *Ph*PRP1 reveals that the most abundant amino acids are lysine (18%), proline (11.7%), and leucine (9%) with serine (8.6%), valine (8.2%), and threonine (7.4%) being the next most abundant amino acids

15

(Table 4). The C-terminal region, defined as starting at residue 121, shows strong similarity to the *Ole-e-1* superfamily of proteins. The C-terminal region has areas that are hydrophobic and hydrophilic and also has six cysteines that are conserved in all known TTSPs and their orthologs. *Ph*PRP1 displays about 71% amino acid identity with *Na*TTS from *N. alata* and TTS-1 from *N. tabacum* (Twomey, Master of Science thesis, 2012).

An antibody (donated by Dr. Bruce McClure, University of Missouri) was designed based on a 12-amino acid region of *Na*TTS and it was shown to cross-react with *Ph*PRP1 (region boxed in Figure 3). This outcome was predicted given that the epitope is almost perfectly conserved, not only among the *Nicotiana* and *Petunia* proteins, but also in a putative ortholog, *Ca*PRP1, from five-day-old roots reported from *C. annuum* (Mang *et al.* 2004). The antibody has been used to probe pistil extracts from snapdragons (*Antirrhinum majus*, Plantaginaceae), bindweed (*Calystegia sepium,* Convolvulaceae), and pear (*Pyrus communis*, Rosaceae). These taxa are phylogenetically distant from the Solanaceae family, and it is therefore unsurprising that their pistil extracts failed to show any cross-reactivity with the anti-TTSP antibody. TTSP orthologs are presumably profoundly diverged from those detected in the Solanaceae subfamilies, even in the relatively conserved C-terminal region where the antibody epitope is located. It is also possible that TTSP orthologs do not exist in these other plant families if these families do not rely on the TTSP as a gametic barrier.

Part III. Research Objectives and Significance

Research Objectives

My thesis research began with a survey of the Solanaceae to determine which subfamilies express TTSP homologs in their pistils. I separated proteins from pistil extracts derived from these taxa by SDS-PAGE, transferred the proteins to nitrocellulose membranes and hybridized the blot with an anti-TTS antibody directed against a strongly-conserved epitope in the C-terminal region (boxed in Figure 3). With this survey, I hoped to determine which subfamilies in the Solanaceae express TTSP homologs. Of the subfamilies in which TTSP homologs are not detectable, what is their phylogenetic relationship to those subfamilies from which TTSP orthologs are known? Given the demonstrated importance of TTSP in pollen tube growth in *Nicotiana*, and the presence of highly similar orthologs in *P. hybrida*, my prediction was that TTSP homologs would be found in all subfamilies of the Solanaceae.

Another research goal was to determine whether the *Petunia TTSP* orthologs are transmitting tissue-specific. TTSP nucleotide sequences were obtained from the NCBI (Wheeler *et al.* 2002) database using BLAST (Altschul *et al.* 1990) and Expressed Sequence Tag (EST) database (Boguski *et al.* 1993). Sequences with a high sequence similarity and query coverage were located on BLAST (Altschul *et al.* 1990) and EST (Boguski *et al.* 1993), but the sequences were not only from transmitting tissues. The homologous sequences I found in these databases were from corollas, roots, leaves and floral tubes, to name a few. Is it possible *TTSP* orthologs exist in other plant tissues? A plant gene may be expressed in

17

different tissues types and in different developmental contexts; the gene product may play similar roles in these tissues or could acquire quite different functions as a result of alternative splicing or differential post-translational modification.

An alternative possibility is the existence of multiple paralogous gene copies in the genome. Solanaceous genomes may harbor paralogs of *TTSP* that exhibit sequence similarity because of common descent from an ancestral form that underwent gene duplications, with the different paralogous copies encoding distinctly different functions in different tissue types. Could some of the gene copies be pseudogenes that are evolving rapidly simply because they are not under functional constraint? I hypothesized that I would find the protein expressed in tissues other than the pistil, and an important goal of my thesis research was to survey a number of *Petunia* tissues at a variety of developmental stages, from one month-old seedlings to one year-old plants.

Finally, a key goal of the study was to obtain full-length sequences encoding *TTSP* orthologs from *P. integrifolia* and *P. axillaris* in order to compare sequences between closely related sister species. The need to identify plant mate-recognition molecules that act in a species-specific manner is paramount (Sanchez *et al.* 2004, Swanson and Vacquier 2002). By identifying all reproductive barriers between closely related species, we can understand what limits gene introgression and hybridization (Lowry *et al.* 2008). Knowing the genes that are responsible for reproductive isolation could in turn provide insight into how these barriers evolved in the first place.

18

Considering other model systems, *Drosophila* in particular, it appears that reproductive tissues express highly divergent genes compared to non-reproductive tissues (Clark *et al*. 2006, Swanson and Vacquier 2002). One hallmark of proteins involved in species-specific mating interactions is that they have a greater rate of amino acid substitutions than other proteins and thus evolve rapidly (Swanson and Vacquier 2002). Rapid evolution of reproductive proteins is indicative of sexual selection between taxa (Coyne and Orr 1998). In this case, sexual selection would be occurring in the transmitting tissue, with selection of conspecific pollen tubes over heterospecific pollen tubes. Gametic incongruity between closely related species could reinforce speciation by preventing interspecific hybrids and cementing species boundaries (Coyne and Orr 1998). Therefore, determining whether *Petunia TTSP* orthologs are gametic isolating barriers and thus interspecific isolating proteins becomes more important. I hypothesized that sequences encoding TTSP orthologs in the two sister species, *P. integrifolia* and *P. axillaris,* would be functionally constrained in the Signal Domain I and CVI-II, but selective pressures would cause divergence in either HVI or HVII.

Significance

The purpose of this research was to increase knowledge of a potential interspecific reproductive barrier that reinforces isolation of sympatric populations of closely related species in the Solanaceae. Members of Solanaceae are vastly important in agriculture, horticulture, floriculture, and alternative and holistic medicine (Hawkes 1999). Some consider it the third most economically important crop family. I chose to focus on two sister species of *Petunia* because these taxa have well-differentiated pollinator syndromes (Gerats

and Vandenbussche 2005, Knapp 2010, Venail *et al.* 2010). These two species are found in sympatry in South America, yet rarely interbreed (Lorenz-Lemke *et al.* 2006). If by chance, heterospecific pollen lands on a stigma, what barriers prevent fertilization? Nature has backup plans, as can be seen from the redundancy of multiple reproductive barriers that prevent hybrids. My research focused on *TTSP* and its orthologs, which I have identified as potential 'speciation' genes in the Solanaceae: that is, evolutionary change in these genes may generate a post-pollination prezygotic barrier that results in reduced pollen performance in heterospecific pistils. This research should contribute to our understanding of gametic incongruity, reinforcement, postmating reproducing barriers, and plant speciation, in addition to furthering our knowledge of pollen-pistil interactions.

METHODS

To survey the Solanaceae family and vegetative tissues for the TTSP, proteins were extracted and separated by SDS-PAGE. The proteins were then transferred to a nitrocellulose membrane, and the blot was probed with an anti-TTSP antibody directed against a strongly-conserved epitope (NNTKKTLVEQGT) in the C-terminal region (boxed in Figure 3). In order to characterize cDNA, total RNA was extracted and first-strand cDNA was synthesized using reverse-transcription. PCR amplified the target gene and was subsequently cloned into an entry vector. Sequencing was performed at a commercial facility. Bioinformatic and phylogenetic analyses were performed using several key programs that deduced the amino acid sequence from the nucleotide sequence (ExPASy), aligned sequences and phylogenetic

20

reconstructions (MEGA5), created sequence alignment images (ClustalW2) and calculated

Ka and Ks ratios of pairwise comparisons (DnaSP).

Protein analysis

Plant Material and Growth Conditions

Plants were grown from seed in the research greenhouse at Western Washington

University. These plants included: *P. integrifolia*, *P. axillaris*, *Datura stramonium*, *D. metel*,

D. wrightii, *Browallia americana*, *Nicotiana alata*, *N. langsdorfii*, *Capsicum annuum,*

S*olanum lycopersicum* and *Ipomoea alba*. All equipment was sterilized in 70% alcohol for

10 minutes and rinsed with distilled water. The soil was sterilized by heating to 180°F for 1

hour. Greenhouse conditions had temperatures ranging from 15°-30°C with at least a 16 hour

light period and watered once a day for a 10 minute time period. Additional adult plants were

grown in the Biology growth room at Western Washington University. These plants

included: *P. integrifolia*, *P. axillaris*, *P. hybrida* (variation Tidal Wave Silver), *C. annuum*,

and *Schizanthus pinnatus*. These plants were grown under the research greenhouse conditions

described above.

Pollinations

Artificial cross pollinations were conducted on unpollinated flowers. Flowers were

emasculated 2-3 days prior to anthesis and pistils required 1-3 days to reach maturity. When

the pistil matured and the stigma was receptive, pollen was applied liberally from freshly

dehisced anthers. The flower pedicel was tagged after pollination with the date and species

crossed. Additional pollinations were conducted on buds and flowers in order to determine if

there is a correlation between pistil length and fertilization success. *P. integrifolia* pollinations had bud and flower lengths ranging from 1.5-2.5 cm and 2.7-3.3 cm, respectively. *P. axillaris* pollinations had bud and flower lengths ranging from 1.9-5.5 cm and 5.9-6.6 cm, respectively.

Protein extraction

Plant tissues (root, leaf, corolla, and pistil) were harvested, weighed and homogenized with Tris-buffered saline (100 mM NaCl in 40 mM Tris-HCl, pH 7.5) in a chilled mortar to obtain a specific concentration (fresh weight of tissue/buffer volume). The concentration for vegetative tissues (root, leaf, and corolla) was 0.0225 mg/μL and reproductive tissue (mature pistils excluding ovaries) was 0.0075 mg/μL. Homogenate was centrifuged for 2 minutes to remove debris and supernatant was frozen at -20°C or used immediately. Samples were prepared by combining 40 μL of supernatant and 10 μL of 4x+ loading dye (50 mM Tris-HCl pH 6.8, 2% SDS, 10% glycerol, 1% β-mercaptoethanol, 12.5 mM EDTA, and 0.02% bromophenol blue) and boiled for 5 minutes at 80°C.

SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis)

Denaturing SDS-PAGE gels were cast from 1.5 mm plates that were washed with 70% ethanol (v/v) and rinsed with nanopure water. For SDS-PAGE, an 8% running gel (ddH$_2$0, 4x Tris 1M pH 8.8, 40% acrylamide, 10% SDS, 10% AP (ammonium persulfate), and TEMED (tetramethylethylenediamine)) was poured and allowed to polymerize completely, and then the 4% stacking gel (ddH$_2$0, 4x Tris 1M pH 8.8, 40% acrylamide, 10% SDS, TEMED, and 10% AP) was poured on top of the running gel.

Denaturing polyacrylamide gels separate proteins based on the molecular weight and two molecular weight size markers helped determine protein molecular weight (kDa). MagicMark XP (Invitrogen, Carlsbad, CA) captured the chemiluminescence detection and allowed the user to observe the molecular size marker alongside the protein samples as viewed in the AlphaInnotech FluorChem Imager. The Invitrogen pre-stained protein ladder (Novex Sharp Protein Standard, Carlsbad, CA) allowed the user to see the molecular size marker while the gel was running. Denaturing polyacrylamide gels were run at 80V (about 2.5-3 hours) in 1x running buffer diluted from 10x running buffer (30.2 g Tris base, 144 g glycine, 10 g SDS and brought to a final volume of 1 L with nanopure water).

Protein immunoblotting

Proteins were transferred to a nitrocellulose membrane (0.45 μm) at a constant voltage of 100V for 1.5 hours in 4°C. Working 1x transfer buffer was made from 10x transfer buffer (30.2 g Tris base, 144 g glycine, 1 g SDS), methanol (cf is 20%) and filled to 1 L with H_2O. The nitrocellulose membrane was blocked 1 hour in PBST buffer (0.1% Tween-20 in phosphate-buffered saline) containing 4% nonfat milk powder and incubated overnight in the primary antibody dilution solution (PBST buffer containing 2% nonfat milk powder and a 1:1,000 dilution of a rabbit antibody). The purpose being that the milk powder proteins bind to non-specific proteins leaving only the target protein for the antibody. This reduces noise on the membrane and eliminates false positives. The rabbit antibody is anti-TTS that was designed from a 12-amino acid epitope in *Na*TTS that is also strongly conserved in TTS-1, TTS-2, and *Ph*PRP1. The nitrocellulose membrane was washed for 10 minutes with PBST buffer 3 times to remove any unbound primary antibody. The

23

nitrocellulose blot was probed by incubating in the secondary antibody dilution solution

(PBST buffer containing 2% nonfat milk powder and a 1:10,000 dilution of a horseradish

peroxidase-linked secondary antibody against rabbit IgG) for 1 hour. The secondary antibody

is linked to a reporter enzyme that produces a signal when bound to the primary antibody.

The membrane was washed to remove unbound probes with PBST buffer 3 times for 10

minutes. To produce a signal, the blots were either exposed to chemiluminescent or

chromagenic substrates. For chemiluminescent signals, the blot was exposed for 5 minutes to

a stabilized hydrogen peroxide and chemiluminescent substrate that will luminesce when

exposed to the reporter enzyme (SuperSignal West Pico chemiluminescent substrate,

ThermoScientific). An AlphaInnotech FluorChem Imager captured and recorded the

chemiluminescent signals. For chromagenic signals, the blot was exposed to the CN/DAB

substrate kit (VWR International, PA) until bands appeared and washed with nanopure water

to stop the reaction.

Characterization of cDNA sequences

RNA extraction

Exactly 150 mg fresh weight of pistils was extracted from *P. integrifolia* and *P.

axillaris* from open flowers and 900 mg fresh weight from 1 month old leaves was extracted

from *P. integrifolia*. Tissues were flash frozen in liquid nitrogen and ground to a fine powder.

RNA was extracted using the TōTALLY RNA Total RNA Isolation kit (Ambion, Austin,

Texas) and followed according to the manufacturer's protocol. RNA extractions used 1.6 mL

of Denaturation Solution (guanidium thiocyanate, Ambion) that was added to ground powder

24

followed by a 2-3 minute centrifuge step to remove debris. The lysate volume was measured and referred to as the starting volume. One starting volume of Phenol:Chloroform:IAA was added to lysate, shaken vigorously for 1 minute, incubated on ice for 5 minutes, and centrifuged for 5 minutes at 10,000 x g. The upper aqueous phase was transferred to a new tube and Phenol:Chloroform:IAA steps were repeated until the interface between the two aqueous layers was clear. Next, 1/10 volume of Sodium acetate was added to the lysate and mixed for 10 seconds. One starting volume of Acid-Phenol:ChCl$_2$ was added, shaken vigorously for 1 minute, incubated on ice for 5 minutes, and centrifuged for 5 minutes at 10,000 x g. The new volume of lysate was measured and RNA was recovered by precipitation with one equal volume isopropyl alcohol and placed in -20°C for 1 hour. One *P. axillaris* RNA sample was incubated in the -20°C freezer for 5 days and the RNA yield was doubled in comparison to the *P. axillaris* sample that was only incubated for 1 hour. The lysate was centrifuged at 10-12,000 x g for 15 minutes and the supernatant was removed. The optional 70% ethanol (v/v) wash was added in order remove residual salts from the pellet.

The pellet was resuspended in DEPC-water and quantified using the NanoDrop ND-1000 spectrophotometer. The extracts obtained the following RNA yields: *P. integrifolia* pistil was 256.7 ng/µL at a 260/280 ratio of 1.96, *P. integrifolia* young leaves was 181.5 ng/µL at a 260/280 ratio of 1.93, and *P. axillaris* pistil was 878.7 ng/µL at a 260/280 ratio of 2.13. The RNA quality was visualized by denaturing agarose gel electrophoresis (nanopure H$_2$0, 10x MOPS buffer (Ambion, Austin, Texas), 37% formaldehyde, and 1 µg/mL ethidium bromide (Sigma Aldrich, UK)). The protocol provided by Ambion (Austin, Texas) was followed except that ethidium bromide was added to the gel rather than individual samples.

Reverse-transcription

Total RNA served as a template using the ThermoScript RT-PCR System (Invitrogen, Carlsbad, CA) which relies on an avian reverse transcriptase to extend the $Oligo(dT)_{20}$ reverse primer (5' TTT TTT TTT TTT TTT TTT TT 3'). The total RNA was denatured before reverse-transcription at 65°C for 5 minutes that contained 50 μM $Oligo(dT)_{20}$ reverse primer, 1-4.3 μg of RNA, 10 mM dNTPs, and brought to a final volume of 12 μL with DEPC-treated water. The master reaction mix was placed on ice. The cDNA master reaction mix was created using 5x cDNA synthesis buffer (250 mM Tris acetate, pH 8.4, 375 mM potassium acetate, and 40 mM magnesium acetate), 0.1 M DTT, 40 U of RNaseOUT, DEPC-treated water, and 15 U of ThermoScript RT. The cDNA master reaction mix (8 μL) was added to the master reaction mix. First-strand cDNA was synthesized using reverse transcription in a Gene Amp PCR system 2400 (Perkin Elmer) for 30 minutes at 50°C followed by 5 minutes at 85°C. Lastly, 2 U of *E. coli* RNase H was added to hydrolyze RNA and incubated at 37°C for 20 minutes. The cDNA synthesis reaction was stored at -20°C.

PCR amplification of *TTSP* homolog cDNA

All gene-specific primers are listed in Table 1 and were designed based on the full-length cDNA sequence, *Ph*PRP1, from *P. hybrida* (accession number: FJ719032) (Twomey, Master of Science thesis, 2012). To amplify the entire gene, 2PhTTS2 (5' GTT CAG CAC AAT TAG TAC TTA GCA A 3') was used along with the $Oligo(dT)_{20}$ reverse (R) primer. Approximately 25-50 ng of single-stranded cDNA was used for PCR amplifications using 0.2 μM of 2PhTTS2 F primer and 0.2 μM of $Oligo(dT)_{20}$ reverse (R) primer. PCR reactions were set up using the Platinum *Taq* DNA Polymerase kit (Invitrogen, Carlsbad, CA) which

required 1X PCR Buffer (200 mM Tris-HCl, pH 8.4, and 500 mM KCl), 0.2 mM dNTP

mixture, 1.5 mM $MgCl_2$, 1 U of Platinum *Taq* DNA Polymerase, and DEPC-treated water to

the final volume of 50.0 µL. The reaction mixture was denatured at 94°C for 2 minutes and

then cycled as follows 40 times: 94°C for 30 seconds, annealing at 53°C for 30 seconds, and

extension at 72°C for 1 minute. The final extension step was held at 72°C for 10 minutes.

The actin F primer (5' ACA GGT ATT GTG TTG GAC TC 3') and actin R primer (5'

CTG TAC TTT CTC TCT GGT GG 3') were used as a positive control for all reactions to

ensure successful PCR and to check the cDNA viability. As a negative control, no template

was added to the reaction. Amplicons were quantified on a 1.5% (w/v) agarose gel containing

1µg/mL ethidium bromide (Sigma Aldrich, UK) and sized with a 1-kb ladder (Bionexus HI-

LO size marker, Invitrogen, Carlsbad, CA). Gel electrophoresis was performed in 1X TAE

buffer (40 mM Tris-HCl, 1 mM EDTA and 0.1% (v/v) glacial acetic acid, pH 8) for 1.5-2

hours at 85V. Gels were visualized under UV light and images were captured using the

ULTRA LŪM Electronic UV Transilluminator and a Canon Powershot A640. PCR amplified

the full-length gene and the 768 bp or 780 bp band was extracted and purified using the

QIAquick Gel Extraction kit (Qiagen, Valencia, CA).

Using the NanoDrop ND-1000 spectrophotometer, the yield and concentration of the

gel extraction was measured and used for PCR. Additional PCR of gel-purified amplicon was

required because bands were extremely smeared indicating non-specific amplifications. To

amplify *TTSP*, both primers were gene-specific and designed within protein coding region.

PCR amplifications were set up with 25-50 ng of gel purified amplicon and Promega PCR

Master Mix kit (Madison, WI) which required 12.5 µL of 2X PCR Master Mix (50 U/mL

Taq polymerase, pH 8.5, 400 μM dNTPs, and 3 mM MgCl$_2$), 0.2 μM 2PhTTS2 F primer, 0.2 μM 3PhTTS R primer (5' GGC ACC TTS RTT GAG GCT TCG 3'), and molecular grade water to the final volume to 50.0 μL. The reaction mixture was denatured at 94°C for 2 minutes and then cycled as follows 40 times: 94°C for 30 seconds, annealing at 61°C for 30 seconds, and extension at 72°C for 1 minute. The final extension step was held at 72°C for 10 minutes. Amplicons were quantified using 1.5% (w/v) agarose gel electrophoresis and sized with a 1-kb ladder (Bionexus HI-LO size marker, Invitrogen, Carlsbad, CA). Bands were extracted and purified using QIAquick Gel Extraction kit (Qiagen, Valencia, CA) and used for subsequent cloning.

PCR cloning into the pENTR 5'-TOPO TA vector

Fresh PCR product (1-4 μL) of *Pi*PRP1, *Pa*PRP1, and *Pi*PRP2 amplicons were cloned separately into pENTR 5'-TOPO TA vector (Invitrogen, Carlsbad, CA) and mixed with salt solution and brought to a final volume with nuclease free water. Cloning reactions were incubated 30 minutes at room temperature and placed on ice. TOPO cloning reactions were transformed into DH5α *E. coli* and incubated on ice for 30 minutes. Cells were heat shocked for 30 seconds at 42°C and then immediately quenched on ice. Room temperature S.O.C. medium (2% Bacto Tryptone (w/v), 0.5% Yeast Extract (w/v), NaCl 8.6 mM, KCl 2.5 mM, MgSO4 20 mM, and Glucose 20 mM) was added to cells and incubated at 37°C for 1 hour in a shaker set at 200 rpm.

Using selective LB media containing kanamycin (50 μg/mL), transformation reactions were plated and incubated at 37°C for 12-16 hours. Putative clones were streaked on patch plates and incubated at 37°C for 12-16 hours. Cells were grown in a selective LB

liquid medium containing kanamycin (50 µg/mL) and incubated at 37°C for 12-16 hours in a

shaker set at 200 rpm. Cells were pelleted at 8000 rpm for 3 minutes using SL-1500 rotor.

Supernatant was decanted and plasmid DNA was purified using the QIAprep Spin Miniprep

kit (Valencia, CA). The manufacturer's protocol was followed and DNA was eluted with

DEPC-treated water and stored at -20°C.

Plasmids were assayed for gene insert using M13 F (5' TGT AAA ACG ACG GCC

AGT 3') and M13 R (5' CAG GAA ACA GCT ATG AC 3') primers. PCR reactions were set

up using 25-50 ng of plasmid DNA and followed manufacturer's protocol (Promega,

Madison, WI). The reaction mixture was denatured at 94°C for 2 minutes and then cycled as

follows 40 times: 94°C for 30 seconds, annealing at 56°C for 30 seconds, and extension at

72°C for 1.5 minutes. The final extension step was held at 72°C for 10 minutes. Amplicons

were sized against a 1-kb ladder (Invitrogen, Carlsbad, CA) using 1.5% (w/v) agarose gel

electrophoresis. Plasmid DNA was sent in for sequencing at Nevada Genomics (Reno, NV)

along with vector primers (M13 F and M13 R) and gene-specific primers (2PhTTS F and

3PhTTS R).

Bioinformatics and Phylogenetic Analysis

Alignment of multiple sequences

The ExPASy program (Gasteiger *et al.* 2003) was used to convert the cladogram

results from Nevada Genomics (Reno, NV). Nucleotide and deduced amino acid sequences

were aligned using the Molecular Evolution Genetic Analysis package version 5 (MEGA5,

Tamura *et al.* 2011) and sequence alignment images were created using ClustalW2 (Larkin *et*

al. 2007). Nucleotide sequences were confirmed using two different plasmids and four different primers, two gene-specific (2PhTTS F and 3PhTTS R) and two vector primers (M13 F and M13 R), to ensure sequence accuracy. Additional TTSP orthologous cDNA sequences from previous research were obtained from the NCBI (Wheeler *et al.* 2002) using BLAST (Altschul *et al.* 1990) searches and aligned with MEGA5 (Tamura *et al.* 2011).

Evolutionary selection

Pairwise comparisons were conducted using the DNA Sequence Polymorphism software (DnaSP, Rozas *et al.* 2003) which calculated the synonymous (Ka) and non-synonymous (Ks) substitution ratios of cDNA sequences (Hurst 2002). The Ka/Ks ratio was calculated to determine if selection was acting on *TTSP*. If the normalized number of non-synonymous (Ka) substitutions is significantly greater than the number of synonymous (Ks) substitutions, then there would be evidence for positive selection since the Ka/Ks ratio is above one. If the normalized number of non-synonymous (Ka) substitutions is significantly less than the number of synonymous (Ks) substitutions, there would be evidence for purifying selection since the Ka/Ks ratio is below one. The Ka/Ks ratio will be analyzed: over the entire length of the sequences, HVI-II, Signal Domain I, and CVI-II.

Phylogenetic analysis

Phylogenetic reconstructions were constructed by the maximum-likelihood (Felsenstein 1981) and maximum parsimony (Fitch 1977) methods using the MEGA5 package (Tamura *et al.* 2011). Both phylogenetic trees were constructed with 1,000 replicates for bootstrap analysis. The bootstrapping calculates the reliability of clustering patterns

(Felsenstein 1985, Higgs and Attwood 2005). In the maximum-likelihood phylogenetic tree, I identified and included a previously-reported sequence that is a likely TTSP ortholog, *Ca*PRP1 (accession AY533017.1), from *Capsicum annuum* (Mang *et al.* 2004), established TTSP orthologs (*Na*TTS, TTS-1, TTS-2, EST717029, *Ph*PRP1, *Pa*PRP1, and *Pi*PRP1), and a homolog from *Arabidopsis thaliana*, AGP31 (Liu and Mehdy 2007). In the maximum parsimony phylogenetic tree, I added TTSP paralogs, 120 kDa and PELPIII, from taxa in Solanaceae with known and putative TTSP orthologs to the phylogenetic tree to determine if AGP31 is a putative ortholog. The output trees were compared with the summary Solanaceae phylogenetic tree based upon the Martins and Barkman (2005) and Paape *et al.* (2008) Solanaceae phylogenetic trees. Martins and Barkman (2005) produced a Solanaceae phylogenetic reconstruction using the nuclear gene *SAMT,* a gene that is essential for plant survival and reproduction because it is critical for methyl salicylate synthesis. Paape *et al.* (2008) produced a Bayesian consensus species phylogeny using sequence data from two chloroplast genes. Comparing the pattern of variation between the two trees should enable us to distinguish whether *TTSP* phylogeny mirrors speciation patterns. In contrast to the topology reported for highly polymorphic genes such as the S-RNases, whose diversification appears to have preceded the phylogenetic divergence of solanaceous taxa, the *TTSP* phylogeny is expected to be consistent with speciation patterns already established for Solanaceae (Singh *et al.* 1991). It also provides evidence for the determination of true orthologs, putative orthologs, paralogs, and homologs.

RESULTS

Pollinations between *P. axillaris* and *P. integrifolia*

According to the literature, *P. axillaris* and *P. integrifolia* are unilaterally

incongruent, meaning cross pollinations only result in seed set in one direction while the

other cross pollination direction fails to set seed (Ando *et al.* 2001). In the Ando *et al.* study,

P. integrifolia fertilized *P. axillaris* but *P. axillaris* did not fertilize *P. integrifolia* (2001). My

artificial crosses do not indicate unilateral incongruency between these species but showed

complete incongruency instead (Table 2b). In both directions, cross pollinations between

these sister species failed to produce seed on emasculated, mature flowers. However,

artificial cross pollinations have high success rates if buds (*P. integrifolia* range 1.5-2.5 cm

and *P. axillaris* range 1.9-5.5 cm) are used (Table 2a). Seeds per capsule were averaged from

at least six artificial pollinations. In buds, *P. integrifolia* crossed with *P. axillaris* produced

27.6 seeds per capsule and *P. axillaris* crossed with *P. integrifolia* produced 129.3 seeds per

capsule. When mature emasculated flowers (*P. integrifolia* range 2.7-3.3 cm and *P. axillaris*

range 5.9-6.6 cm) were pollinated, the fertilization rate greatly decreased to 0.0 seeds per

capsule for both cross pollinations. The artificial pollinations of fully-open mature

emasculated flowers resulted in zero seed set indicating a strong and complete reproductive

barrier between the two species.

Immunodetection of TTSP orthologs in reproductive tissues

TTSPs in pistil extracts were analyzed using SDS-PAGE and protein immunoblotting

techniques. Crude extracts made from equivalent tissue fresh weights were resolved via SDS-

PAGE, the proteins transferred to a nitrocellulose blot, and the protein blot probed with an

32

anti-TTSP antibody. Five of the six generally-accepted subfamilies of Solanaceae were surveyed for the TTS protein: Petunioideae, Nicotianoideae, Solanoideae, Cestroideae, and Schizanthoideae. The sixth subfamily, Schwenckioideae, was not surveyed due to the difficulty in obtaining seeds. Also, this basal subfamily is more distant from Petunioideae than Cestroideae. Based on the accepted phylogenetic relationships, it is reasonable to assume that if a signal was obtained from Cestroideae, then extracts from the Schwenckioideae would also react with the antibody. A seventh group, Goetzoideae, was not surveyed because it is not universally accepted as a subfamily of the Solanaceae.

On the immunoblots of pistil extracts (Figures 4 and 5a), the anti-TTSP antibodies cross-reacted with proteins from: Petunioideae (*P. hybrida*-data not shown, *P. integrifolia,* and *P. axillaris*)*,* Nicotianoideae (*N. alata* and *N. langsdorfii*-data not shown)*,* Solanoideae (*Datura stramonium*, *C. annuum,* and S*olanum lycopersicum*-data not shown)*.* The pistil protein immunoblots did not show positive signals from Schizanthoideae (*Schizanthus pinnatus*), Cestroideae (*Browallia americana*) or the outgroup *Ipomoea alba*, which belongs to Convolvulaceae. In all cases, the anti-TTSP antibodies cross-reacted with proteins that are heterogenous in molecular mass, presumably due to variability in glycosylation and formed a diffuse smear in the gel lanes. This is reminiscent of the molecular mass heterogeneity reported for the *Nicotiana* TTSPs (Wu *et al*. 1995). As seen in Figure 4 and Figure 5a, the approximate molecular mass of the cross-reacting proteins for each species is as follows: *P. hybrida*, 55 to 100 kDa; *P. integrifolia*, 33 to 120 kDa; *P. axillaris*, 55 to 100 kDa; *N. alata*, 45 to 110 kDa; *Datura stramonium*, 55 to 90 kDa; and *Capsicum annuum*, 80 to 260 kDa.

33

Characterizing TTSP orthologs from *P. integrifolia* and *P. axillaris* pistils

RT-PCR was used to isolate and clone a *TTSP* ortholog, *Pi*PRP1, from *P. integrifolia* and a *TTSP* ortholog, *Pa*PRP1, from *P. axillaris*. *Pi*PRP1 is a 771 bp incomplete cDNA clone (Figure 6), and *Pa*PRP1 is a 759 bp incomplete cDNA clone (Figure 7). The 5' coding and non-coding region of mRNA of these sequences was obtained using the vector-specific M13 primer (Table 1). Both mRNA sequences have a nine amino acid or 27 bp 5'-untranslated region that is conserved (Figure 6 and 7). The initiation codon (ATG) starts at site 28-30 for both *Pi*PRP1 and *Pa*PRP1. Sequence alignments show that *Pi*PRP1 is an incomplete clone encoding a 256 amino acid polypeptide, and *Pa*PRP1 is an incomplete cDNA clone encoding a 253 amino acid polypeptide. A comparison of their coding sequences with the known TTSP orthologs shows that *Pi*PRP1 and *Pa*PRP1 are missing the last three amino acids (CKK) at the C-terminal end. The absence of the last three amino acids resulted from constraints in the design of the primers used for reverse transcription and PCR. I found that oligo (dT) primers and primers that included a portion of the 3' coding and noncoding sequence produced high background during PCR and made the cloning and sequencing of the amplicons difficult. The only solution was to design a gene-specific reverse primer that lacked the terminal three amino acids. The three missing amino acids (CKK) are completely conserved among all the published TTSP orthologs and also in *Ph*PRP1 from *P. hybrida* (Twomey, Master of Science thesis, 2012). Therefore, it is reasonable to assume that 'CKK' are also the last three amino acids in the polypeptide encoded by *Pi*PRP1 and by *Pa*PRP1.

The ExPASy PeptideMass program (Gasteiger *et al.* 2003) predicts a molecular mass of 27.5 kDa for *Pi*PRP1 and a molecular mass of 27.1 kDa for *Pa*PRP1. Comparing *Pi*PRP1 and *Pa*PRP1 cDNAs, both nucleotide sequences are highly similar and display 97.7% similarity, but there are a total of 18 nucleotide substitutions between them (Figure 9a). The variant residues include a 12 nucleotide indel (insertion/deletion) in *Pi*PRP1 that is not found in *Pa*PRP1 or *Ph*PRP1 but is found in all other published TTSP orthologs, namely *Na*TTS, TTS-1, TTS-2 and *Ca*PRP1 (Figure 3). It is interesting that the 12 nucleotide indel codes for 'VKPP' in HVII. KPP motifs are abundant in the proline-rich HVII and are a signature feature of TTSPs, setting these proteins apart from other pistil-expressed arabinogalactan proteins (AGPs). Because this specific 'VKPP' motif is found in all TTSP orthologs except *Pa*PRP1 from *P. axillaris* and *Ph*PRP1 from *P. hybrida*, I infer that the tetrapeptide was deleted in *Pa*PRP1 as *P. axillaris* diverged from the ancestral group. *P. hybrida* is a synthetic species, produced in cultivation by crossing *P. axillaris* and *P. integrifolia* (Sink 1984). *Ph*PRP1 most likely represents the gene copy inherited by this amphidiploid from its *P. axillaris* parent.

When the 12 nucleotide indel is not included, the nucleotide identity between *Pa*PRP1 is 98.9%. Apart from the four amino acid indel, there are four amino acid substitutions between *Pi*PRP1 and *Pa*PRP1, generating a 1% difference between the two orthologs from substitutions alone. Of the six nucleotide substitutions, four are non-synonymous and two are synonymous, resulting in four amino acid changes overall (Figure 9a and 9b). The nucleotide polymorphisms occur only in HVII, or more specifically, within a 44 amino acid span out of the 253 and 257 amino acids encoded by *Pa*PRP1 and *Pi*PRP1

from *P. axillaris* and *P. integrifolia*, respectively (Figure 9b). The 5' untranslated region of these proteins is identical with zero nucleotide substitutions.

*Pa*PRP1 and *Pi*PRP1 are both highly glycosylated, judging from the substantial molecular mass heterogeneity they exhibit on protein immunoblots. *Pi*PRP1 has ten lysine-proline-proline (KPP) repeats (Figure 6) and *Pa*PRP1 has eight KPP repeats (Figure 7) over the entire span of the polypeptide. However, most of the KPP motifs are located within HVII and one or two located in of in the C-terminal region. Among the published TTSP orthologs, the number of KPP triplets ranges from a low of 6 in *Ca*PRP1 from *C. annuum* to as many as 11 in TTS-1 from *N. tabacum* (Figure 3). *Pi*PRP1 and *Pa*PRP1 both have the defining TTSP characteristics, including: a highly conserved signal sequence, consisting of MAKAXVL, where X represents either a leucine or phenylalanine; CVI-II, with a highly-conserved 29-residue histidine-rich domain, and a C-terminal region with six cysteines that are perfectly conserved in all TTSP orthologs; and, HV Domains I-II, with multiple KPP triplets in HVII (Figures 6 and 7).

Immunodetection of TTSP orthologs in vegetative tissues

Protein extracts were prepared from vegetative tissue, including leaves and roots from *P. integrifolia* plants at different developmental stages, and these protein extracts were subjected to immunoblot analysis using the anti-TTSP antibody. The anti-TTSP antibody cross-reacted with discrete protein bands in extracts made from leaves and roots of one month-old seedlings and from leaves of two month-old plants (Figure 5b and 5c). Plant age is defined in terms of the days elapsed since the seeds were sown to when the tissue extracts were made. In contrast to the heterogeneity of pistil TTSP orthologs, the cross-reacting bands

in vegetative tissue were sharply delineated and the molecular mass easily estimated. The cross-reacting band from *P. integrifolia* roots migrates at 190 kDa. Two discrete bands are seen in the leaf extracts, located at 70 and 38 kDa (Figure 5b and 5c). For both root and leaf tissues, no signals were obtained from young leaf or young root tissue of one year-old plants (Figure 5c).

Cloning and sequencing of a TTSP homolog expressed in *P. integrifolia* seedling leaves

RT-PCR was used to clone and sequence *TTSP* homologs from two types of *P. integrifolia* tissues: *Pi*PRP1 was cloned from RNA extracted from pistil tissue, and *Pi*PRP2 was amplified from RNA extracted from seedling leaves. *Pi*PRP2 is a 780 bp full-length cDNA clone and is essentially identical to *Pi*PRP1 except for one synonymous substitution at site 678 where *Pi*PRP1 has an adenine and *Pi*PRP2 has a guanine (Figure 6 and 8). Like *Pi*PRP1, *Pi*PRP2 has a 12 nucleotide indel that is not present in *Pa*PRP1 and *Ph*PRP1, the TTSP orthologs from *P. axillaris* and *P. hybrida*, respectively. The one nucleotide polymorphism between the two sequences can be attributed to allelism, which is common within and among populations of a species. The polymorphism is a silent or synonymous substitution.

Compared to the high concentration of *Pi*PRP1 in pistil extracts, *Pi*PRP2 is not abundant in young leaves. To extract the RNA from seedling leaves, I used six times the fresh tissue weight (0.9 g) as I used to isolate *Pi*PRP1 from pistil tissue. Similarly, in conducting the protein immunoblot analysis, I made leaf and root extracts with a protein concentration a three times greater (0.0225 µg/mL) relative to the protein concentration of the pistil extracts (Figure 5a and 5b). Despite these efforts to scale up, the TTSP homologs

37

from seedling tissues produced faint bands, some of which were at the detectability limits of the highly sensitive chemiluminescent immunodetection protocols I used.

Alignment of putative TTSP orthologs expressed in reproductive and vegetative tissues

In an attempt to find potential TTSP homologs in EST databases, I used the *PaPRP1* cDNA sequence in a BLAST search (Altschul *et al.* 1990) of various solanaceous EST databases. From the meaningful matches I obtained (e value: 0.0-8e-121), I chose only sequences that were at least 200 bp in length. Meaningful homologies were found with these EST clones: flower EST from *P. hybrida* (CV295550.1), corolla tube EST from *P. axillaris* (FN015061.1), root EST from *P. hybrida* (FN003773.1 and FN006765.1). I generated multiple alignments of these sequences, *Ph*PRP1 from *P. hybrida* characterized by Megan Twomey (FN006765.1), and all three of my cDNA clones, namely, *Pi*PRP1 and *Pi*PRP2 from *P. integrifolia* and *Pa*PRP1 from *P. axillaris* (Figure 10). The alignments show that all of these sequences display strong sequence similarity, apart from the indels already noted between *Pi*PRP1 and *Pa*PRP1. To the extent that this information is available (most of the EST sequences are incomplete), the domain architecture characteristic of the published TTSP orthologs is perfectly preserved in these clones.

Alignment of TTSP orthologs

I used the multiple sequence alignment of all known full-length TTSP orthologs to define the domain architecture of this subfamily of pistil-expressed arabinogalactan proteins (Figure 3). The alignment reveals that some regions of the protein vary little from one taxon to another, with sequence variation more common in other parts. All of the sequences have many structural features in common. All contain essentially the same signal sequence

38

(MAKAXVL, where X represents phenylalanine or leucine) in the Signal Domain I, along with two conserved domains and two hypervariable domains. The conserved domains include a histidine-rich domain located between the hypervariable domains, and a C-terminal region with six cysteines that are perfectly conserved in all known TTSP orthologs and also in the Ole e 1 superfamily of arabinogalactan proteins. Signal Domain I and CVI-II show little variability even among very divergent taxa. However, there are many non-synonymous substitutions in HVI, and HVII contains many indels (insertions or deletions), many non-synonymous substitutions, and shows marked variation in the number and location of KPP triplets.

Analysis of amino acid compositions in the eight TTSP orthologs indicates high levels of K (lysine) and P (proline) at 11.32% and 16.93%, respectively (Table 3). Within individual proteins, these values varied from 10.89%-13.08% for lysine and 16.54%-21.15% for proline (Table 4). The next most prevalent amino acids are serine (8.11%), leucine (7.96%), valine (7.81%), threonine (6.81%), and alanine (6.26%) (Table 4).

Selection occurring in Signal Domain I, CV Domains I-II, and HV Domains I-II

Ka/Ks ratios were calculated for pairwise comparisons of the eight TTSP orthologs using the DnaSP program. The results support the hypothesis that TTSP orthologs display positive selection in HVI (residues 63-126) and HVII (residues 196-441) and purifying selection in Signal Domain I (residues 1-62), CVI (residues 127-195), and CVII (442-852), as shown in Tables 5-9. For this analysis, I compared eight TTSP mRNA sequences from three Solanaceae subfamilies representing six different species. The results show that between most pairwise comparisons of TTSP orthologs, even those from closely related

species, there is clear evidence of positive selection (Tables 6 and 8). The selection for sequence variation is confined to the two hypervariable domains (HVI and HVII) in the N-terminal half of the protein. In a total of 28 pairwise comparisons, 21 out of the 28 showed positive selection in either HVI or HVII (Tables 6 and 8). Ka/Ks ratios for 18 of the pairwise comparisons indicate positive selection in HVI (Table 6) and three in HVII (Table 8). Overall, the hypervariable domains had Ka/Ks ratio values ranging from 0.0000-1.9012. An especially striking finding is that in any pairwise comparison, if the HVI displays positive selection, the HVII lacks it, and vice versa. In all of the pairwise comparisons (28 per domain, 84 total), the conserved domains (Signal Domain I and CVI-II) showed purifying selection and had Ka/Ks ratio values ranging from 0-0.6560 (Tables 5, 7, and 9). In pairwise comparisons of the full-length sequence (residues 1-852), Ka/Ks values ranged from 0.000-0.6757 (Table 10).

Phylogenetic analysis of TTSPs and their known or putative orthologs

A summary phylogenetic tree of the Solanaceae was created based upon the Martins and Barkman (2005) Solanaceae phylogenetic reconstruction that used the nuclear gene *SAMT* and the Paape *et al.* (2008) Bayesian consensus species phylogeny using sequence data from two chloroplast genes (Figure 11). The purpose of the summary tree was to simplify the Solanaceae phylogeny between subfamilies and taxa surveyed in this research. The divergence time estimates (in millions of years) are indicated by arrows (Paape *et al.* 2008). The maximum-likelihood phylogenetic tree reconstruction consisted of nine TTSP orthologs: *Na*TTS, TTS-1, TTS-2, *Ca*PRP1, EST717029, *Pi*PRP1, *Pi*PRP2, *Pa*PRP1, and

40

*Ph*PRP1. A homolog that is also a putative ortholog, AGP31, was included to root the tree.

The purpose of this phylogenetic tree was to test whether TTSP phylogeny matched well-

established Solanaceae phylogeny. If the branching is consistent with speciation patterns,

then it is possible that *TTSP* diverges after speciation in order to prevent hybridizations

between recently diverged species.

The maximum-likelihood and maximum parsimony trees I generated both show three

distinct clades representing the three subfamilies Solanaceae I surveyed: Petunioideae,

Nicotianoideae, and Solanoideae. Within the Nicotianoideae clade, it appears that *Na*TTS

and TTS-1 are more closely related than TTS-2 and TTS-1 (Figure 12 and 13). This might

seem counterintuitive, considering that the two genes, TTS-1 and TTS-2, were isolated from

the same species, *N. tabacum* (Cheung *et al.* 2000). However, *N. tabacum* is an

amphidiploid, generated in cultivation from the hybridization of two putative *Nicotiana*

species, *N. sylvestris* and *N. tomentosiformis* (Bland *et al.* 1985). TTS-1 is presumably

inherited from the one *Nicotiana* species and TTS-2 from the other parent species (Cheung *et*

al. 2000). In light of the amphidiploid history of *N. tabacum*, it is not unexpected that TTS-1

and *Na*TTS are more closely related than are TTS-1 and TTS-2.

TTSP phylogeny shows Petunioideae in a separate clade, diverged from

Nicotianoideae and Solanoideae, which cluster as sister clades (Figure 12). The maximum-

likelihood phylogenetic reconstruction is consistent with subfamily clustering pattern of the

Solanaceae summary tree (Figure 11). Additionally, *Pi*PRP1 and *Pi*PRP2 cluster together,

and this elucidates whether these genes are orthologs or paralogs (Figure 12 and 13). If they

were paralogs, *Pi*PRP1 would be connected to *Pa*PRP1 and *Ph*PRP1 and *Pi*PRP2 would be

41

more distantly diverged. Also, *Pa*PRP1 and *Ph*PRP1 are consistently branched together indicating that *P. hybrida* acquired *Ph*PRP1 from *P. axillaris*. If *Ph*PRP1 had a 'hybrid' *TTSP* from both *P. integrifolia* and *P. axillaris*, *Ph*PRP1 would be branched equally with *Pi*PRP1 and *Pa*PRP1. This is not the case, so we can assume *Ph*PRP1was acquired from *P. axillaris*.

DISCUSSION

Transmitting tissue-specific proteins (TTSPs) are expressed in the pistils of *Nicotiana* species and are crucial for optimal pollen tube growth (reviewed by Cheung *et al.* 2000). One goal of my thesis research was to discover whether TTSP orthologs are widespread in the Solanaceae, occurring in all or most of the subfamilies. Using immunodetection protocols, I found evidence that TTSP homologs are expressed in the pistils of three subfamilies: Petunioideae, Solanoideae, and Nicotianoideae. The anti-TTSP antibody did not cross-react with pistil extracts from *Browallia americana*, a member of Cestroideae, or *Schizanthus pinnatus*, which is a member of the basal subfamily, Schizanthoideae.

I cloned and sequenced pistil cDNA clones from two Petunioideae species, *P. axillaris* and *P. integrifolia*. Sequence similarity, perfect preservation of domain architecture, and biochemical characteristics (such as immunoreactivity) provide strong evidence that these cDNA encode TTSP orthologs in *Petunia*. Immunodetection protocols revealed that the protein orthologs are also expressed in vegetative tissues of *Petunia*. I cloned a cDNA from *P. integrifolia* seedling leaves that is almost indistinguishable from the cDNA expressed in pistil tissues. Based on multiple alignments of the available TTSP orthologs, I identified

42

distinct hypervariable and conserved domains and then tested the strength of natural selection

on these domains. The phylogeny of TTSP orthologs is congruent with published

phylogenies inferred from both chloroplast and nuclear genes.

Gametic isolation between *P. axillaris* and *P. integrifolia*

The working model that serves as the scaffold of my thesis research is that molecular

divergence in TTSP orthologs reinforces speciation between sister groups of the Solanaceae.

To investigate this possibility, I focused on two species of the Petunioideae, *P. axillaris* and

P. integrifolia, which are considered sister species (Sink 1984). *P. axillaris* and *P.

integrifolia* are found in sympatry in nature but fail to produce hybrids because of multiple

prezygotic reproductive barriers (Ando *et al*. 2001). Pollinator-driven isolation is clearly a

large contributor to these prezygotic reproductive barriers. Their pollinator syndromes are

dramatically different: *P. integrifolia* has strongly-scented purple flowers that are pollinated

during the day by bees, while *P. axillaris* has long, white, UV-reflecting corolla tubes and the

flowers are pollinated mainly at night by hawkmoths (Stehmann *et al*. 2009). However, some

interspecific pollen transfer probably does occur in nature because bees have been seen

visiting *P. axillaris* in the daytime and hawkmoths have been on *P. integrifolia* flowers at

night. Despite the lack of complete pollinator fidelity, hybrids are rarely found in the native

habitats (Dell'Olivo *et al*. 2011). These premating reproductive barriers appear to be

reinforced by gametic postmating barriers to gene flow, but as described next, there is some

confusion in the literature as to the strength and directionality of the barriers to interspecific

hybridization between these two species of *Petunia*.

Some authors describe complete gametic isolation between *P. integrifolia* and *P. axillaris* (Sink 1984), whereas Ando *et al.* (2001) describe unidirectional cross-compatibility between the two species. In my accessions of the two species, I found evidence of strong bilateral incongruity, with complete failure to set seed in an interspecific cross (Table 2b). The discrepancy in the primary literature may stem at least in part from differences in the metrics employed to determine fertility. For example, Ando *et al.* (2001) used the rate of capsule set as a measure of fertility, whereas I calculated rate of seed set per capsule. I observed that occasionally the ovary enlarges and matures into a capsule following an artificial cross pollination, while containing few to no seeds within the capsule. Measuring total seed number per capsule was therefore more appropriate for my inquiry into the interspecific breeding behavior of the two species.

There are at least three different subspecies of *P. axillaris* (subspecies *axillaris*, *parodii*, and *subandina*) and at least two of *P. integrifolia* (subspecies *integrifolia* and *inflata*) (Ando *et al.* 2001). Thus, genetic divergence among the subspecies, compounded by allelic variation at the population level, could also account for the disparity in the strength and directionality of the interspecific barriers reported in the literature. The *P. axillaris* and *P. integrifolia* accessions used in my research were sourced from Annie's Heirloom Seeds (Clarksville, Michigan) and the origins of these seeds are unknown. In future studies, it would be ideal to obtain seeds of sister species from the native range and compare TTSP cDNA sequences from sympatric and allopatric populations. It is my hypothesis that TTSP will have diverged much more in sympatric populations compared to allopatric populations

44

due to greater selective pressure for reinforcement of gametic isolation between sister species.

Ando *et al.* (2001) made artificial crosses between *P. axillaris* and *P. integrifolia* accessions from syntopic and allotopic populations in Brazil and Uruguay with varied results. *P. axillaris* crossed successfully with *P. integrifolia* with a capsule set rate of 11.0-66.7%, but the reciprocal pollination was largely unsuccessful with *P. integrifolia* as the pistillate parent and *P. axillaris* the pollen donor, and the rate of capsule set was 0.0-11.0% (Ando *et al.* 2001). The bilateral incongruity between my accessions of *P. axillaris* and *P. integrifolia,* and the unilateral incongruity observed by Ando *et al.* (2001), may be explained by genotypic differences among these accessions. The genotypes surveyed by Ando *et al.* may have come from populations separated over a greater geographic range, with a consequent weakening of the interspecific gametic barrier. However, the *Petunia* accessions used in this study have complete incongruity and gene flow between them is blocked completely. The remainder of my thesis focuses on studying TTSP orthologs in these two *Petunia* species and examining sequence divergence between them. Evidence of positive selection in critical domains of the *P. axillaris* and *P. integrifolia* TTSP orthologs may support the model that sequence divergence in the protein could generate and/or reinforce gametic barriers between the two species.

TTSP orthologs in the Solanaceae

If TTSP orthologs are potential drivers of gametic isolation between Solanaceous taxa, the proteins should be ubiquitous or near-ubiquitous in the family. To test this

hypothesis, I chose representative species from five of the widely-recognized subfamilies of the Solanaceae (Petunioideae, Solanoideae, Cestroideae, and Schizanthoideae, and Nicotianoideae). The immunological approach made use of an antibody targeted against an epitope (boxed in Figure 3) that is largely conserved in the known TTSP orthologs. Based on cross-reactivity with the antibody, TTSP orthologs were found to be expressed in pistils of the Petunioideae, Nicotianoideae, and Solanoideae. However, no cross-reactivity was detected in pistil extracts made from *Browallia americana* (Cestroideae) or *Schizanthus pinnatus* (Schizanthoideae). It is possible that TTSP orthologs are absent in these subfamilies, and the gene is a synapomorphy of the other three subfamilies that evolved in their most recent common ancestor. But it is also possible, and perhaps more likely, that the epitope targeted by the antibody has diverged so much between the three sister clades and the two more distantly related ones (Cestroideae and Schizanthoideae) that the antibody fails to recognize TTSP orthologs from the latter. Although the epitope is located in one of the conserved domains (boxed in Figure 3), the sequence conservation is less than perfect even among the three sister clades (Nicotianoideae, Solanoideae, and Petunioideae). For example, the anti-TTSP antibody recognizes NNTKKTLVEQGK in *Na*TTS, and two of these residues are divergent in *Ca*PRP1 from *Capsicum annuum* (Solanoideae) subfamily (Figure 3). Given that the published molecular phylogenies cluster the Nicotianoideae and Solanoideae in the same clade, the more distantly related subfamilies--including Schizanthoideae or Cestroideae--are even more likely to display sequence variation in this region. As might be expected, the antibody failed to cross-react with pistils extracts from the outgroup species,

Ipomoea alba, which belongs to the Convolvulaceae, a sister family to the Solanaceae (Figure 4).

In future studies, it may be worthwhile to screen the Solanaceae (and perhaps the Convolvulaceae as well) for TTSP homologs using an antibody to the highly conserved histidine-rich region of the polypeptide (see Figure 3). Alternatively, it might be possible to isolate cDNA encoding a *TTSP* ortholog from pistils of these taxa using the signal sequence (MAKAXV) to design a forward primer. The signal sequence is extremely conserved among all known TTSP and their solanaceous orthologs, and in conducting a BLAST search using this sequence, I did not obtain any hits that were not known TTSP orthologs.

TTSP orthologs in Petunioideae pistils

Using the immunological approach, I was able to identify TTSP homologs in pistil extracts of Petunioideae (*P. hybrida*, *P. axillaris* and *P. integrifolia*), as well as from members of the Solanoideae (*C. annuum*, *Datura stramonium*) and Nicotianoideae (*N. langsdorfii* and *N. alata*). As reported for the *Nicotiana* species (Cheung *et al*. 2000) the cross-reacting pistil proteins on the immunoblots appeared as heterogenous smears rather than discrete bands. Interestingly, the homologs from different species each have their own unique pattern of mobility on SDS-PAGE gels. I reduced the protein concentration loaded to determine if the smearing might be an artifact of overloading, but the smearing pattern persisted.

Wu *et al.* (1995) have shown for the *Nicotiana* TTSPs that the heterogeneity in molecular mass is largely a consequence of variability in glycosylation of the protein

47

backbone, with the result that a variety of glycoforms are secreted into the extracellular matrix. My immunoblot survey of solanaceous TTSP homologs shows that the heterogeneity in glycosylation is species-specific, with a wide array of glycoforms being secreted in pistils of different taxa (see Figure 4).

TTSP orthologs in vegetative tissue of *P. integrifolia*

TTSPs from *Nicotiana* species were described by Cheung *et al.* (2000) as specific to the transmitting tissue of the pistil (hence, the name: transmitting tissue-specific protein or TTSP). However, my BLAST searches (Boguski *et al.* 1993) of solanaceous EST libraries returned many good matches with cDNA prepared from vegetative tissue. These included cDNA sequences with 75-100% query coverage and 75-100% maximum identity from root, corolla tube, and leaf EST libraries. I chose to investigate the tissue specificity further, and found evidence supporting the hypothesis that TTSP homologs are not unique to the transmitting tissues of pistils but are expressed in other plant tissues as well.

I prepared immunoblots with total protein extracted from leaves and roots of *P. integrifolia* plants representing four different developmental stages: one month-old seedlings, two month-old seedlings, four month-old plants and one year-old plants. As shown in Figure 5b and 5c, protein extracts from seedling tissue cross-react with the anti-TTSP antibody but extracts from four-month and year-old plants do not. It is interesting that TTSP homologs were not detected in vegetative tissue from four month or one year-old plants--not even in the immature leaves at the shoot apex and root tissue from the terminal apex of the root system. Based on these results, it appears that the expression of TTSP homologs in vegetative tissues is temporally regulated.

48

Seedling leaves from *P. integrifolia* have two glycoforms of *Pi*PRP1 that migrate at

70 and 38 kDa; seedling roots appear to have a single glycoform that migrates at 190 kDa

(Figure 5b and 5c). This relative uniformity in glycosylation of TTSP homologs in vegetative

tissue contrasts with the substantial heterogeneity in the glycosylation of the protein

expressed in pistil tissue. Taken together, my results suggest that TTSP homologs are

glycosylated in unique ways in different tissue types and the glycosylation pattern in the

pistil can vary tremendously even between closely related taxa (for instance, compare the

molecular mass heterogeneity of *Pa*PRP1 and *Pi*PRP1 in Figure 5a).

AGPs, such as the TTSPs, are known to be hydroxylated at specific prolines by prolyl

4-hydroxylases (P4H) in the endoplasmic reticulum, with up to 12 putative *P4H* genes

reported in *Arabidopsis thaliana* (Velasquez *et al.* 2011). The resulting hydroxyprolines are

substrates for sequential glycosyltransferases of the GT37 subfamily, which in *Arabidopsis*,

consists of at least ten members. The tissue- and taxon-specific variation in glycosylation of

TTSPs and their orthologs may result from allelic diversity in the activity and specificity of

these glycosyltransferases. The tissue-specific patterns in glycosylation may have functional

significance. For example, the abundance and variety in glycosylation of the pistil TTSP

glycoform may reflect the role of the glycan side chains as a nutritional substrate for growing

pollen tubes (Cheung *et al.* 1996, Cheung *et al.* 2001), with the highly branched side chains

furnishing many more start points for hydrolytic cleavage by exoglucanases associated with

pollen tube tips (Sakurai 1998).

A critical question is whether the TTSP homologs detected in seedling tissues are

encoded by paralogs of the gene expressed in the pistil, or whether the pistil and vegetative

tissue glycoforms are both encoded by a single gene. I cloned and sequenced a cDNA, *Pi*PRP2, from seedling leaves and compared that sequence to the pistil-expressed cDNA, *Pi*PRP1 (Figure 9a). The leaf and pistil cDNAs are identical except for a single nucleotide that represents a synonymous substitution in CVII. Thus, *Pi*PRP1 and *Pi*PRP2 encode an identical polypeptide backbone in the pistil and in seedling leaves, respectively (Figure 9b). Further studies are necessary to definitively establish that *Pi*PRP1 and *Pi*PRP1 represent the coding region of a single gene in the genome of *P. integrifolia*. However, this seems likely given these observations: the *P. integrifolia* polypeptides expressed in leaf and pistil tissue are identical (this study); only one copy of the presumed TTSP ortholog (*AGP31*) is found in the *Arabidopsis thaliana* genome (Liu and Mehdy 2006); two distinctly different (91% amino acid identity) copies of TTSP (TTS-1 and TTS-2) are expressed in pistils of *Nicotiana tabacum*, a known amphitetraploid. I propose that there is a single gene copy of *TTSP* orthologs in the haploid genome of solanaceous taxa, and that the gene is developmentally regulated such that it is highly expressed in seedling tissues and in pistils of mature plants. Further, I propose that the gene product is differentially glycosylated in different tissue types so as to optimize the glycosylation pattern to the unique microenvironment in which it must function.

The concentration of TTSP orthologs in vegetative tissue of *Petunia* seems lower, on a fresh weight basis, in contrast to their abundance in pistil tissue. It is possible that in roots and leaves, the protein is expressed in a small subset of cell types, which would reduce their abundance in a whole tissue extract. The cellular-level role of TTSPs and their orthologs remains a mystery, but the one common denominator—rapid growth of specific cell types—

suggests a role in mediating high rates of membrane extension. In pistils, TTSPs are thought

to facilitate tip expansion of pollen tubes; in the young roots and leaves of seedlings, these

proteins may promote rapid expansion of differentiating sieve tubes and xylem conducting

elements and fibers.

Structural features of the *Petunia* TTSP orthologs

I cloned cDNA encoding a TTSP homolog from *P. axillaris* (*Pa*PRP1) and from *P.

integrifolia* (*Pi*PRP1). Sequence alignment of these two clones and the published TTSP

sequences enabled me to divide the entire sequence into discrete domains: Signal Domain I,

CVI-II and HVI-II (Figure 3 and Figure 9b). Within these domains, there are specific

segments that are worth noting. The putative signal peptide (Signal Domain I) is followed by

a short hypervariable domain (HVI). HVI is succeeded by a highly conserved histidine-rich

domain (CVI), and followed by a longer hypervariable domain (HVII) that is rich in proline.

Finally, there are six completely-conserved cysteine residues in the C-terminal end (CVII),

which has sequence similarity with the PAC domain (for PRP and AGP, containing Cys)

identified as a signature characteristic of many "nonclassical" AGPs (Baldwin *et al*. 2001).

AGPs with PAC domains include two other classes of pistil-expressed proteins, the 120 kDa

glycoprotein and PELPIII, both described from *Nicotiana* species.

Two structural features set the TTSPs and their orthologs apart from other AGPs,

including the 120kDa glycoprotein and PELPIII. TTSPs and their orthologs have a highly

conserved histidine-rich region and a proline-rich region (HVII) that contains multiple KPP

triplets. Solanaceous TTSPs have serine, threonine, alanine or valine preceding the KPP

motifs, while other cell wall proline-rich proteins (PRPs) have valine, proline, glutamic acid, tyrosine, lysine and histidine preceding these motifs (Cheung *et al.* 1993). I find this to be true of not only the *P. axillaris* and *P. integrifolia* orthologs which I have characterized, but also *Ca*PRP1 from *Capsicum annuum* (Mang *et al.* 2004) and AGP31 from *Arabidospsis thaliana* (Liu and Mehdy 2006).

It is striking that no two TTSP orthologs are identical in terms of the number and location of the KPP triplets in their HVII. The two most similar proteins, *Pa*PRP1 and *Pi*PRP1, differ mainly because of a four amino acid indel that, remarkably, codes for a VKPP motif (Figure 9b). This indel is part of a direct repeat sequence and its nucleotide sequence, GTTAAACCACCA, is found twice upstream of the indel site in *Pi*PRP1 and *Pi*PRP2 and once upstream of the indel site in *Pa*PRP1 (direct repeat sequences are highlighted in bold in Figures 6-8). The presence of the short direct repeats may make this region prone to strand slippage during DNA replication (Gore *et al.* 2006), increasing the odds of mutational insertions and deletions of the KPP motif. Nucleotide mispairing caused by direct repeats is also likely to interfere with PCR-based amplification, and I can attest to the fact that clones encoding *TTSP* orthologs are difficult PCR templates!

The KPP motif likely serves as a recognition signal for prolyl 4-hydroxylases that hydroxylate one or both of the proline residues, converting them into substrates for subsequent O-glycosylation by ER-localized glycosyl transferases (Velasquez *et al*. 2011). All the known TTSP orthologs have a unique pattern of KPP motifs that may generate species-specific glycosylation patterns. The unique carbohydrate signature of each pistil TTSP may be a key determinant of its species-specific interactions with conspecific pollen

tubes. In this model, it is apparent that mutations in the KPP motifs, yielding changes in the carbohydrate signature, could create gametic barriers between emerging species. As I discuss next, speciation events may place genes encoding TTSP orthologs under selective pressure for sequence divergence in the domains responsible for species-specific pollen-pistil interactions. I propose that the HVI-II domains of TTSP orthologs may be good candidates to be among the key determinants of species-specific interactions between the pistil extracellular matrix and the pollen tubes growing within it.

TTSP ortholog domains as substrates for natural selection

To examine the hypothesis that TTSPs and their orthologs evolve rapidly, with some domains under positive selection and others under evolutionary constraint, I computed the Ka/Ks ratios for each domain using pairwise comparisons of representative full-length TTSP orthologs (28 pairwise comparisons per domain). The conserved domains of the protein are clearly under strong functional constraints. In all 28 pairwise comparisons for each domain, Signal Domain I and CVI-II, Ka/Ks ratios that are significantly under 1.00 indicating purifying selection on those domains (Tables 5, 7, and 9). The hypervariable domains exhibited Ka/Ks ratios in excess of one (Tables 6 and 8). As adaptively evolving sites, these hypervariable domains would be likely candidates for regions of the protein that interact with a pollen tube component.

These findings support my hypothesis that the HVI and HVII of TTSP orthologs are the prime determinants of species-specific interactions. I propose that these domains are exposed to selective pressure when an ancestral population diverges into two sympatric

53

lineages. With sufficient sequence divergence in either the HVI or HVII, mating interactions between the splitting populations would fail, due to a gametic barrier between them. It has been proposed that gametic proteins can be involved in fertilization through species-specific mate recognition, so reproductive boundaries between species are further defined by diverging gametic proteins involved in mate recognition (Zigler *et al.* 2005). A comparison of TTSP orthologs from the two sister species, *P. axillaris* and *P. integrifolia*, is particularly instructive. The four residue changes between *Pi*PRP1 and *Pa*PRP1 are concentrated in one distinct area, HVII, spaced across 44 amino acids in the polypeptide. These four non-synonymous substitutions, along with the two synonymous substitutions, could be sufficient to generate gametic incongruence, which in turn would reinforce isolation of the two lineages in sympatry.

After a speciation event and throughout the existence of the species, however, these multidomain proteins must come under strong purifying selection. Polymorphisms within populations that lead to non-synonymous substitutions are predicted to be deleterious and are therefore likely to be rapidly eliminated from the population. Although I sequenced a number of independent *Pi*PRP1 cDNA clones that were prepared from RNA pooled from multiple individuals in my accession of *P. integrifolia* (an outbreeder, since the species is self-incompatible), I did not come across any amino acid polymorphisms. The evolutionary history of TTSP orthologs, and presumably all reproductive proteins that participate in generating species boundaries, may be distinctive compared to other genes that are under functional constraints. These reproductive genes must evolve in spurts, with rapid

evolutionary change during speciation events, followed by a period of stasis that could last as long as the species does.

It is striking that in pairwise comparisons, positive selection was never found simultaneously in both hypervariable domains (Table 11). Either the HVI or HVII, but not both, displayed the signature of positive selection between any pair of TTSP orthologs that were compared. One explanation for positive selection only occurring in either hypervariable domain but never both could be that the model used to calculate the type of selection seems to underestimate the evolutionary events occurring between taxa in HVII. Indels are insertions/deletions that are extremely prominent in HVII yet gaps are not counted in the Ka/Ks calculation. Even though HVII has only two pairwise comparisons with positive selection, the alignment displays possible positive selection of indels in that domain but is undetectable using the Ka/Ks model (Figure 3). If indels were weighted, HVII would likely indicate positive selection between most pairwise comparisons. Future studies should aim to calculate selection weighting indels equally or more heavily than synonymous and non-synonymous substitutions.

Another explanation for HVI or HVII indicating positive selection between pairwise comparisons could be that adaptive mutations rapidly reach saturation as soon as the gametic barrier is erected between the incipient species, with further change being maladaptive and subject to purifying selection. In other words, once the protein changes in one hypervariable domain, it has become different enough to not be recognized by the recently diverged sister species and further change is not only unnecessary but deleterious. Even when I compared TTSP orthologs of more distantly related taxa, for example *C. annuum* compared with *P.*

axillaris, only one of the two hypervariable domains showed positive selection, with relatively little sequence variation in the other domain. *Pi*PRP1 and *Pa*PRP1 displayed positive selection in HVII (Table 8), while all the other pairwise comparisons showed positive selection in HVI (Table 6). Another explanation is attributed to the method of calculating

Curiously, I did not find evidence of positive selection in HVI or HVII when I compared *Ph*PRP1 from *P. hybrida* with either *Pa*PRP1 or *Pi*PRP1 (Table 6 and 8). The explanation probably lies in the amphidiploid origins of *P. hybrida*, with *P. axillaris* and *P. integrifolia* as parent species (Wijsman 1982). There are few possibilities for the differences between *Ph*PRP1 and the parental TTSP orthologs. The parental TTSP orthologs have recombined in *P. hybrida*, making *Ph*PRP1 a 'hybrid' gene. *Ph*PRP1 has six residues that are different compared to *Pi*PRP1 because it does not contain the four amino acid indel, so at first glance it appears to have been inherited from the *P. axillaris* parent (Figure 9b). However, there are still four other residues that are divergent between *Ph*PRP1 and *Pa*PRP1. Another explanation for the residue changes between *Pa*PRP1 and *Ph*PRP1 could be described by the breeding history of *P. hybrida*. When crosses between the parental species were successful, hybrid plants were chosen by their ability to breed. *P. integrifolia* is self-incompatible and *P. axillaris* is self-compatible, and through several selections of self-compatible hybrids, the *TTSP* gene from *P. axillaris* was chosen predominantly over the *P. integrifolia* gene copy. The non-synonymous substitutions occurring between *Pa*PRP1 and *Ph*PRP1 may be attributed to *P. axillaris* subspecies progenitor used to create *P. hybrida*. The *P. axillaris* subspecies or accession used in this study may be different than the *P.*

56

*hybrid*a progenitor and either *TTSP* could have undergone more divergence due to the speciation process. It is my hypothesis that *TTSP* genes will diverge rapidly when incipient species are found sympatrically. When speciation occurs because of allopatric populations, *TTSP* will not be under strong selective pressures to reinforce isolation. Future studies should compare different subspecies and accessions from the same species, perhaps from *P. axillaris*.

Interestingly, when *P. hybrida* is used as the pollen parent, it readily sets seed on both *P. axillaris* and *P. integrifolia* pistils. In other words, *P. hybrida* pollen exhibits complete congruency with *P. integrifolia* and *P. axillaris* pistils with about the same seeds per capsule that would be expected from compatible crosses within the same species (Table 2B). Perhaps *P. hybrida* pollen carries a broad-specificity allele of the hypothetical pollen tube cognate that interacts with multiple closely related TTSP orthologs, thereby enabling pollen tubes to grow successfully in pistils of the ancestral species, *P. integrifolia* and *P. axillaris*. The distinctly lower fertility of *P. hybrida* pollen in *P. integrifolia* pistils (see Table 2B) suggests a greater mismatch between this *P. hybrida* pollen tube cognate and *Pi*PRP1. That *P. hybrida* pollen sets any seed at all when crossed with *P. integrifolia* pistils could be explained by the much shorter pistils of the latter (average length of *P. integrifolia* pistils is 9mm, compared to 28mm for the *P. hybrida* accessions used in this study). The reciprocal pollinations are dramatically different, with very low seed set when *P. hybrida* pistils are pollinated with either *P. integrifolia* or *P. axillaris* pollen (Table 2B). Perhaps the homogenization of domain structure in the 'hybrid' TTSP ortholog expressed in *P. hybrida* prevents the protein from interacting successfully with either *P. axillaris* or *P. integrifolia* pollen.

57

Phylogenetic analysis of *TTSP orthologs*

The maximum-likelihood phylogenetic reconstruction was created consisting of TTSP orthologs (*Na*TTS, TTS-1, TTS-2, *Pi*PRP1, *Pi*PRP2, *Pa*PRP1, *Ph*PRP1, *Ca*PRP1, and EST717029) and a putative ortholog (AGP31) from *Arabidopsis thaliana* (Figure 12). The phylogenetic tree indicates that *N. alata* and *N. tabacum* form a distinct clade, and *C. annuum* and *Solanum tuberosum* form another separate clade. The *Petunia* species are the most diverged and most basal of the solanaceous TTSP orthologs, with the *P. axillaris* ortholog clustered closer to the *P. hybrida* ortholog than to the *P. integrifolia* ortholog. The close concordance with the published phylogeny of the solanaceous subfamilies suggests that the evolutionary history of these pistil-expressed genes is informative for reconstructing species-level phylogenies. The discovery of genes playing a similar role in pollen-pistil interactions in the more problematic taxa might therefore be useful in phylogenetic studies.

A maximum parsimony phylogenetic reconstruction consisting of putative and known TTSP orthologs (*Na*TTS, TTS-1, TTS-2, *Pi*PRP1, *Pi*PRP2, *Pa*PRP1, *Ph*PRP1, EST717029 and *Ca*PRP1) and presumed paralogs (120 kDa and PELPIII) produced a tree that grouped all the putative TTPS orthologs and clearly separated the presumed paralogs (Figure 13). It is interesting to note that the putative ortholog, AGP31, from *Arabidopsis thaliana* (At1g29280) was clustered between the known TTSP orthologs and paralogs from solanaceous taxa. This may indicate that fewer evolutionary steps have occurred from TTSP orthologs and AGP31 than between TTSP orthologs and their solanaceous paralogs. The Solanaceae are in a completely different taxonomic order (Solanales) than *Arabidopsis thaliana* (Brassicales), which offers strong evidence that AGP31 is a true TTSP ortholog.

58

AGP31 is expressed in etiolated seedlings and in differentiating vascular tissue of seedlings, as well as being strongly expressed throughout the pistil (Liu and Mehdy 2006). While this pattern of expression may have been an obstacle to accepting the AGP31 as a TTSP ortholog, it is now clear that TTSP orthologs are expressed in non-reproductive tissues in the Solanaceaae as well. As mentioned earlier, I have demonstrated the expression of *Pi*PRP2 in seedling leaves of *P. integrifolia*, and also uncovered the presence of extremely homologous sequences in EST databases of root and leaf tissue of other solanaceous species.

The maximum-likelihood phylogenetic tree clearly follows the summary cladogram constructed by Martins and Barkman (2005) and Paape *et al*. (2008) using nuclear and chloroplast genes, respectively. However, the maximum parsimony cladogram does not as Solanoideae branches earliest and is shown to be more basal than Petunioideae and Nicotianoideae. If *TTSP* acts as a reinforcement gene, then *TTSP* phylogeny will mirror speciation events between closely related taxa; however, highly divergent taxa may have *TTSP* convergence and then evolutionary events may be masked. I propose that *TTSP* phylogeny is most meaningful when elucidating closely related taxa divergence, but it is not advisable to use *TTSP* for phylogenetic reconstructions from highly divergent taxa.

That the phylogeny of *TTSP* orthologs maps to the species phylogeny is further indication that these genes are good candidates for the much sought after 'speciation genes' (Riesberg and Willis 1997). The distinctive pattern of evolution of TTSP orthologs—with some domains under selective pressures to diverge following incipient speciation—offers a fascinating model for understanding the forces that drive plant speciation. Sequence divergence in such reproductive proteins may even serve as a footprint of sympatric

59

speciation. I would predict that sequence divergence in the variable domains of TTSP orthologs correlates closely in sympatric species, but poorly or not at all in taxa known to have evolved through allopatric isolation. In addition to elucidating the evolutionary dynamics of TTSPs and their orthologs, understanding their cellular role—in both reproductive and vegetative tissues--is also an urgent priority in my view.

Figure 1. *Petunia axillaris.* The photographs illustrate floral structure and whole plant morphology of the accessions used in this study.

Figure 2. *Petunia integrifolia*. The photographs illustrate floral structure and whole plant morphology of the accessions used in this study.

```
NaTTS    MAKALVLFQLSVLLLSSFTVVLSEEEGIGGW-FVDKHHDHLSPAQAPKPHKGHH-HPKHS  58
TTS-1    MAKALVLFQLSVLLLSSFTVVLSQEEDIGGW-FTTKHHDHLSPAQAPKPHKGNH-HPKHS  58
TTS-2    MAKALVLFQLSVLLLSSFTVVLSQEEGIGGW-FIDKHHGHFPPAEAPKAHKGYH-HPKYS  58
CaPRP1   MAKALVLFQLSVLFLSSFAV-LSHGDH---WYSLDKNVDHLPPAQAPKPHKGHHPPPKNS  56
PhPRP1   MAKAFVLFHLSVLLLSSFTV-LSHGEGLMGW-SLTKHEDHLPPAQAPKPHKGHH-HPKHS  57
PaPRP1   MAKAFVLFHLSVLLLSSFTV-LSHGEGLMGW-SLTKHEDHLPPAQAPKPHKGHH-HPKHS  57
PiPRP1   MAKAFVLFHLSVLLLSSFTV-LSHGEGLMGW-SLTKHEDHLPPAQAPKPHKGHH-HPKHS  57
PiPRP2   MAKAFVLFHLSVLLLSSFTV-LSHGEGLMGW-SLTKHEDHLPPAQAPKPHKGHH-HPKHS  57
         ****:***:****:****:* **:  :    *    *: .*:.*:****.*** *  ** *

NaTTS    PAPSPT--------KPPTYSPSKPPVKPVKPPTKPPTYSPSKPPAKAP------------   98
TTS-1    PAPSP--------TKPPTYSPSKPPVKPP---VKPPTKAPTYSPSKPP----AKPPVKPP  103
TTS-2    PAPSPSYTPSKPPTKPPTYSPSKPPAKPP---VKPPTKPPTYSPSKPP----AKPPVKPP  111
CaPRP1   PAPSPIDTP-----TPPPAKSPSPPPAKPPTP---PPAKPPSPPPSKPPTKPPAKSPSPPP  109
PhPRP1   PAPSPA-------TPPPAYSPSKPPVKPPTPSVKPPAKPPVKPPTPS-----VKPP----  101
PaPRP1   PAPSPA-------TPPPAYSPSKPPVKPPTPSVRPPAKPPVKPPSPS-----VKPP----  101
PiPRP1   PAPSPA-------TPPPAYSPSKPPVKPPTPSVKPPAKPPVKPPTPS-----VKPPVKPP  105
PiPRP2   PAPSPA-------TPPPAYSPSKPPVKPPTPSVKPPAKPPVKPPTPS-----VKPPVKPP  105
         *****         **: *** **.** ** **:  .*  .*:  .

NaTTS    -----VKPPTPSPYPAP-APITRKPVAVRGLVYCKPCKFRGVKTLNQASPLLGAVVELVCN  153
TTS-1    ------TPTPSPY-PAPAPITRKPVAVRGLVYCKPCKFRGVKTLNQASPLLGAVVKLVCN  156
TTS-2    -------TPSPY---PAPITRKPVAVRGLVYCKPCKFRGVKTLNQASPLLGAVVKLVCN  160
CaPRP1   AKPPTKPPTPSPYYP-----SRKPVITRGLVYCKPCKYRGIETLYRAKPLEGAVKLVCK  164
PhPRP1   TPS-VKPPTPSPYYP-----SRKPVAVRGLVYCKPCKYRGVETLNLATPLQGAIVKLACN  155
PaPRP1   TPS-VKPPTPSPYYP-----SRKPAAVRGLVYCKPCKYRGVETLNLATPLQGAIVKLACN  155
PiPRP1   TPS-VKPPTPSPYYP-----SRNPVAVRGLVYCKPCKYRGVETLNLATPLQGAIVKLACN  159
PiPRP2   TPS-VKPPTPSPYYP-----SRNPVAVRGLVYCKPCKYRGVETLNLATPLQGAIVKLACN  159
             *****       :*:*. **********:**::** *.** **:*:*.*:

NaTTS    NTKKTLVEQGKTDKNGFFWIMPKFLSSAAYHKCKVFLVSSNNTYCDVPTDYNGGKSGALL  213
TTS-1    NTKKTLVEQGKTDKNGFFWIMPKFLSSAAYHKCKVFLVSSNNTYCDVPTDYNGGKSGALL  216
TTS-2    NTKKTLVEQGKTDKNGFFWIMPKLLSSGAYHKCKVFLVSSNNTYCDVPTDYNGGKSGALL  220
CaPRP1   NSKKTLVEQGKTDKNGYFWIMPKLLTSGAYHKCKVFLVSSNNSYCNVPTNFNGGKSGALL  224
PhPRP1   NTKKTLVEQGTTDKNGFFLILPKMLSSGAYHKCKVFLVSSKNTHCDVPTNFNGGKSGALL  215
PaPRP1   NTKKTLVEQGTTDKNGFFLILPKMLSSGAYHKCKVFLVSSKNTHCDVPTNFNGGKSGALL  215
PiPRP1   NTKKTLVEQGTTDKNGFFLILPKMLSSGAYHKCKVFLVSSKNTHCDVPTNFNGGKSGALL  219
PiPRP2   NTKKTLVEQGTTDKNGFFLILPKMLSSGAYHKCKVFLVSSKNTHCDVPTNFNGGKSGALL  219
         *:*********:*****:* *:**:*:* .*************:*: :*:***::********

NaTTS    KYTPLPKPP---AASSLPVKLPTFDVFTVGPFGFEPSKKVPCKK  254
TTS-1    KYTPLPKPP--AATS-LPVKLPTFDVFTVGPFGFEPSKKVPCKK  257
TTS-2    KYTPLPKPP--AASS-LPVKPPTYDVFTVGPFGFEPSKKVPCKK  261
CaPRP1   KYTPPSKPTPSAITK--------FDVFTVGPFGFEPSKKVPCKK  260
PhPRP1   KYTPLPKPP---ATSHLPVKPPTFDVFTVGPFGFEASSKVPCKK  256
PaPRP1   KYTPLPKPP---ATSHLPVKPPTFDVFTVGPFGFEASTKVPCKK  256
PiPRP1   KYTPLPKPP---ATSHLPVKPPTFDVFTVGPFGFEASTKVPCKK  260
PiPRP2   KYTPLPKPP---ATSHLPVKPPTFDVFTVGPFGFEASTKVPCKK  260
         **** .**.     :.        :**********.*.******
```

Figure 3. Clustal W2 image of MEGA5 protein alignment of TTS orthologs from: *N. alata* (*Na*TTS), *N. tabacum* (TTS-1 and TTS-2), *Capsicum annuum* (*Ca*PRP1), *P. hybrida* (*Ph*PRP1), *P. axillaris* (*Pa*PRP1), and *P. integrifolia* (*Pi*PRP1). Identical amino acids among all sequences are indicated by (*); (:) indicates one nucleotide difference between the sequences; (.) indicates two or more nucleotide differences between the sequences; ( ) indicates a gap either by an insertion or deletion from one of the sequences. The numbers at the end of the row indicate the amino acid positions. Domains are indicated by: Signal Domain I (residues 1-22), HVI is boxed in dashes (residues 23-40), CVI (residues 41-62), HVII is boxed in dots (residues 63-109), and CVII (residues 110-261). Anti-TTSP epitope is boxed with a solid line.

Figure 4. SDS-PAGE and protein blot analysis of select taxa from five Solanaceae subfamilies and 1 taxa from Convolvulaceae. Protein extractions were made from unpollinated mature pistils at a constant tissue fresh weight per volume of extraction buffer. Lane M: Magic Marker XP size marker; lane 1: *P. integrifolia*; lane 2: *Nicotiana alata*; lane 3: *Datura stramonium*; lane 4: *Browallia americana*; lane 5: *Capsicum annuum*; lane 6: *Schizanthus pinnatus*; lane 7: *Ipomoea alba*.
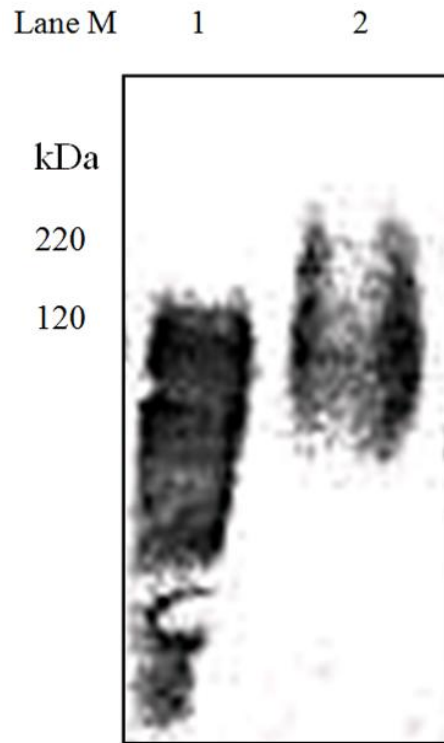
Figure 5a. SDS-PAGE and protein blot analysis of extracts made from *Petunia integrifolia* and *Petunia axillaris* pistils. Protein extractions were from unpollinated pistils at a constant tissue fresh weight per volume of extraction buffer. Lane 1: *P. integrifolia*; lane 2: *P. axillaris*.
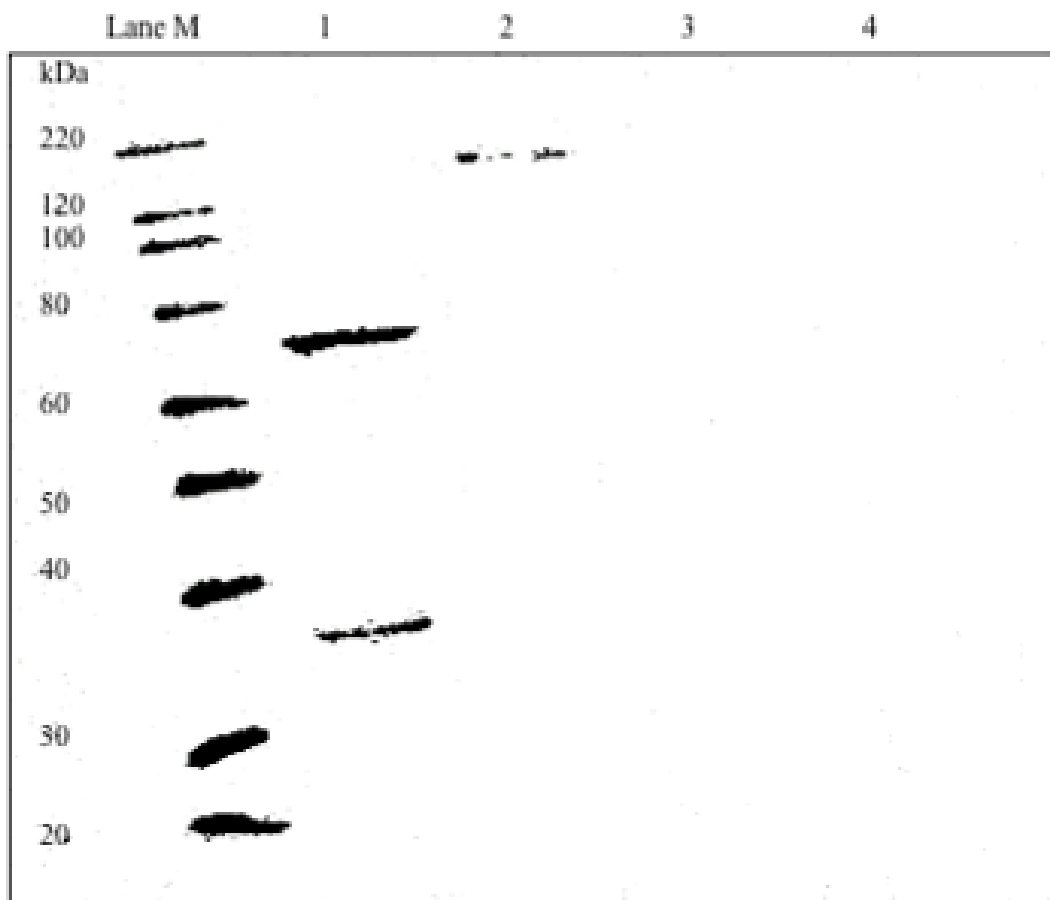
Figure 5b. SDS-PAGE analysis of vegetative tissues from one month-old seedlings of *P. integrifolia* and *Browallia americana*. Vegetative tissue extracts equivalent to 12 mg of fresh tissue were loaded in each lane. Lane M: Magic Marker XP size marker; lane 1: seedling leaf from *P. integrifolia*; lane 2: seedling root from *P. integrifolia*; lane 4: seedling leaf from *B. americana*; lane 4: seedling root from *B. americana.*
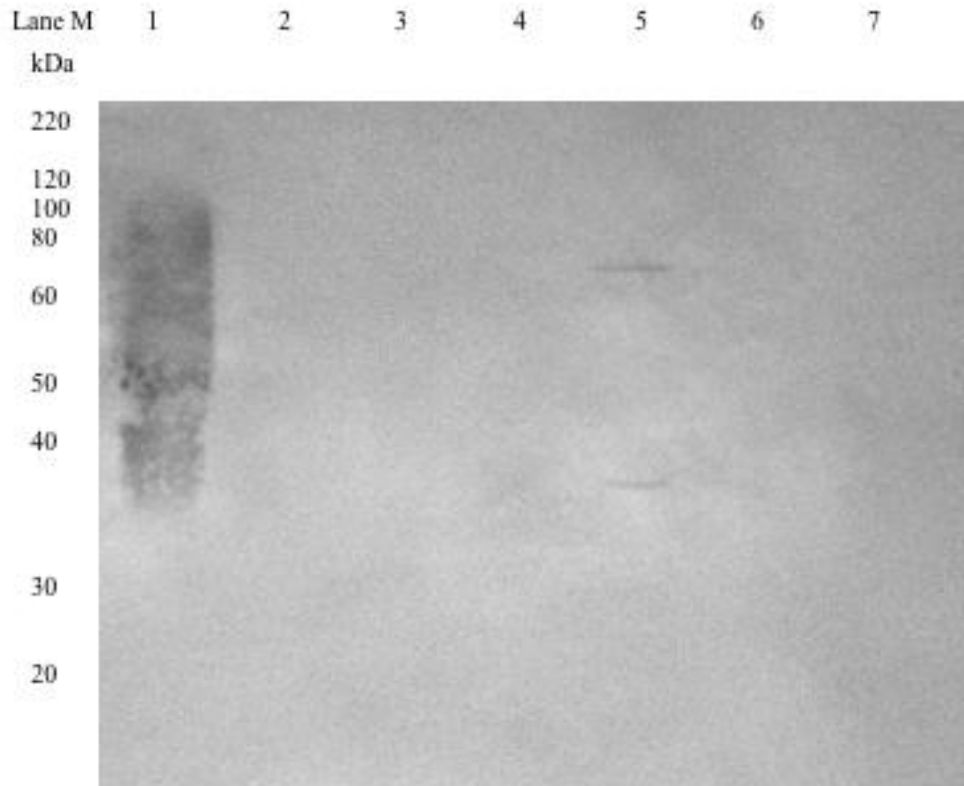
Figure 5c. Chromogenic image of SDS-PAGE analysis of *P. integrifolia* extracts prepared from reproductive tissue (pistil) and vegetative tissue (root, leaf, and corolla) of two month-old seedlings and one year old plants. The extracts loaded on the gel contained the equivalent of 3 mg pistil tissue and 12 mg of the vegetative tissue. Lane M: marker lane; lane 1: pistils from two month-old plant; lane 2: young corolla from one year old plant; lane 3: young corolla from two month-old seedling; lane 4: young leaves from one year old plant; lane 5: leaves from two month-old plant; lane 6: root tips from one year old plant; lane 7: root tips from two month-old plant.

```
gttcagcacaattagtacttagcaataatggcaaaggcctttgttcttttttcatctttca
                                  M  A  K  A  F  V  L  F  H  L  S

gttttattactcagctcattcacagttcttagccatggtgaagggttaatgggttggtca
 V  L  L  L  S  S  F  T  V  L  S  H  G  E  G  L  M  G  W  S

ttgaccaaacatgaagaccaccttccaccagctcaagccccaaagcctcacaagggccac
 L  T  K  H  E  D  H  L  P  P  A  Q  A  P  K  P  H  K  G  H

caccatccaaaacattccccagccccttcaccagcaaccccaccaccagcttatagccca
 H  H  P  K  H  S  P  A  P  S  P  A  T  P  P  P  A  Y  S  P

tcaaaaccaccagttaaaccacctaccccctcagttaaaccaccagctaagccaccagtt
 S  K  P  P  V  K  P  P  T  P  S  **V  K  P  P**  A  K  P  P  V

aaaccacctaccccatcagttaaaccaccagttaaaccacctaccccatcagttaagcca
 K  P  P  T  P  S  **V  K  P  P**  V  K  P  P  T  P  S  **V  K  P**

**ccaaca**ccgtcaccttattacccttctaggaatcctgtagctgttcgtggccttgtttac
 **P  T**  P  S  P  Y  Y  P  S  R  N  P  V  A  V  R  G  L  V  Y

tgcaaaccttgcaagtatagaggggttgaaactttaaacctggctaccccactccaggga
 C  K  P  C  K  Y  R  G  V  E  T  L  N  L  A  T  P  L  Q  G

gcgatagtgaaactagcgtgcaacaacacaaagaagacactagttgaacagggcacaaca
 A  I  V  K  L  A  C  N  N  T  K  K  T  L  V  E  Q  G  T  T

gacaagaatggattcttcttgatcttgcccaaaatgttgtcctcaggggcctaccacaaa
 D  K  N  G  F  F  L  I  L  P  K  M  L  S  S  G  A  Y  H  K

tgcaaggtgttcttagtctcatcgaagaatactcactgcgatgtcccaacaaatttcaat
 C  K  V  F  L  V  S  S  K  N  T  H  C  D  V  P  T  N  F  N

ggtggaaaatctggtgctctcttaaaatacaccccacttcccaaaccaccagcgactagt
 G  G  K  S  G  A  L  L  K  Y  T  P  L  P  K  P  P  A  T  S

catctccctgttaaacccccaacatttgatgtattcactgtggggcctttggattcgaa
 H  L  P  V  K  P  P  T  F  D  V  F  T  V  G  P  F  G  F  E

gcctcaaccaaggtgccttgcaaaaaatag
 A  S  T  K  V  P  C  K  K  -
```

Figure 6. Nucleotide and deduced amino acid sequences of the *Pi*PRP1 cDNA from *Petunia integrifolia* pistil.

```
gttcagcacaattagtacttagcaataatggcaaaggcctttgttcttttttcatctttca
                             M  A  K  A  F  V  L  F  H  L  S

gttttattactcagctcattcacagttcttagccatggtgaagggttaatgggttggtca
 V  L  L  L  S  S  F  T  V  L  S  H  G  E  G  L  M  G  W  S

ttgaccaaacatgaagaccaccttccaccagctcaagccccaaagcctcacaagggccac
 L  T  K  H  E  D  H  L  P  P  A  Q  A  P  K  P  H  K  G  H

caccatccaaaacattccccagccccttcaccagcaaccccaccaccagcttatagccca
 H  H  P  K  H  S  P  A  P  S  P  A  T  P  P  P  A  Y  S  P

tcaaaaccaccagttaaaccacctacccccctcagttagaccaccagctaagccaccagtt
 S  K  P  P  V  K  P  P  T  P  S  V  R  P  P  A  K  P  P  V

aaaccacctagcccatcagttaaaccacctaccccatca**gttaagccacca**acaccgtca
 K  P  P  S  P  S  V  K  P  P  T  P  S  **V  K  P  P**  T  P  S

ccttattacccttctaggaaacctgcagctgttcgtggccttgtttactgcaaaccttgc
 P  Y  Y  P  S  R  K  P  A  A  V  R  G  L  V  Y  C  K  P  C

aagtatagagggggttgaaactttaaacctggctaccccactccagggagcgatagtgaaa
 K  Y  R  G  V  E  T  L  N  L  A  T  P  L  Q  G  A  I  V  K

ctagcgtgcaacaacacaaagaagacactagttgaacagggcacaacagacaagaatgga
 L  A  C  N  N  T  K  K  T  L  V  E  Q  G  T  T  D  K  N  G

ttcttcttgatcttgcccaaaatgttgtcctcaggggcctaccacaaatgcaaggtgttc
 F  F  L  I  L  P  K  M  L  S  S  G  A  Y  H  K  C  K  V  F

ttagtctcatcaaagaatactcactgcgatgtcccaacaaatttcaatggtggaaaatct
 L  V  S  S  K  N  T  H  C  D  V  P  T  N  F  N  G  G  K  S

ggtgctctcttaaaatacaccccacttcccaaaccaccagcgactagtcatctccctgtt
 G  A  L  L  K  Y  T  P  L  P  K  P  P  A  T  S  H  L  P  V

aaaccccccaacatttgatgtattcactgtggggcctttggattcgaagcctcaaccaag
 K  P  P  T  F  D  V  F  T  V  G  P  F  G  F  E  A  S  T  K

gtgccttgcaaaaaatag
 V  P  C  K  K  -
```

Figure 7. Nucleotide and deduced amino acid sequences of the *Pa*PRP1 cDNA from *Petunia axillaris* pistil.

69

```
gttcagcacaattagtacttagcaataatggcaaaggcctttgttcttttttcatctttca
                                M   A   K   A   F   V   L   F   H   L   S

gttttattactcagctcattcacagttcttagccatggtgaagggttaatgggttggtca
 V   L   L   L   S   S   F   T   V   L   S   H   G   E   G   L   M   G   W   S

ttgaccaaacatgaagaccaccttccaccagctcaagccccaaagcctcacaagggccac
 L   T   K   H   E   D   H   L   P   P   A   Q   A   P   K   P   H   K   G   H

caccatccaaaacattccccagccccttcaccagcaaccccaccaccagcttatagccca
 H   H   P   K   H   S   P   A   P   S   P   A   T   P   P   P   A   Y   S   P

tcaaaaccaccagttaaaccacctaccccctca**gttaaaccacca**gctaagccaccagtt
 S   K   P   P   V   K   P   P   T   P   S   **V   K   P   P**   A   K   P   P   V

aaaccacctaccccatca**gttaaaccacca**gttaaaccacctaccccatca**gttaagcca**
 K   P   P   T   P   S   **V   K   P   P**   V   K   P   P   T   P   S   **V   K   P**

**cca**acaccgtcaccttattacccttctaggaatcctgtagctgttcgtggccttgtttac
 **P**   T   P   S   P   Y   Y   P   S   R   N   P   V   A   V   R   G   L   V   Y

tgcaaaccttgcaagtatagagggggttgaaactttaaacctggctaccccactccaggga
 C   K   P   C   K   Y   R   G   V   E   T   L   N   L   A   T   P   L   Q   G

gcgatagtgaaactagcgtgcaacaacacaaagaagacactagttgaacagggcacaaca
 A   I   V   K   L   A   C   N   N   T   K   K   T   L   V   E   Q   G   T   T

gacaagaatggattcttcttgatcttgcccaaaatgttgtcctcaggggcctaccacaaa
 D   K   N   G   F   F   L   I   L   P   K   M   L   S   S   G   A   Y   H   K

tgcaaggtgttcttagtctcatcgaagaatactcactgcgatgtcccaacaaatttcaat
 C   K   V   F   L   V   S   S   K   N   T   H   C   D   V   P   T   N   F   N

ggtggaaaatctggtgctctcttaaaatacacccccacttcccaagccaccagcgactagt
 G   G   K   S   G   A   L   L   K   Y   T   P   L   P   K   P   P   A   T   S

catctccctgttaaaccccccaacatttgatgtattcactgtggggcctttggattcgaa
 H   L   P   V   K   P   P   T   F   D   V   F   T   V   G   P   F   G   F   E

gcctcaaccaaggtgccttgcaaaaaatag
 A   S   T   K   V   P   C   K   K   -
```

Figure 8. Nucleotide and deduced amino acid sequences of the *Pi*PRP2 cDNA from *P. integrifolia* young leaf.

.

70

```
PiPRP1  ATGGCAAAGGCCTTTGTTCTTTTTCATCTTTCAGTTTTATTACTCAGCTCATTCACAGTT 60
PiPRP2  ATGGCAAAGGCCTTTGTTCTTTTTCATCTTTCAGTTTTATTACTCAGCTCATTCACAGTT 60
PhPRP1  ATGGCAAAGGCCTTTGTTCTTTTTCATCTTTCAGTTTTATTACTCAGCTCATTCACAGTT 60
PaPRP1  ATGGCAAAGGCCTTTGTTCTTTTTCATCTTTCAGTTTTATTACTCAGCTCATTCACAGTT 60
        ************************************************************

PiPRP1  CTTAGCCATGGTGAAGGGTTAATGGGTTGGTCATTGACCAAACATGAAGACCACCTTCCA 120
PiPRP2  CTTAGCCATGGTGAAGGGTTAATGGGTTGGTCATTGACCAAACATGAAGACCACCTTCCA 120
PhPRP1  CTTAGCCATGGTGAAGGGTTAATGGGTTGGTCATTGACCAAACATGAAGACCACCTTCCA 120
PaPRP1  CTTAGCCATGGTGAAGGGTTAATGGGTTGGTCATTGACCAAACATGAAGACCACCTTCCA 120
        ************************************************************

PiPRP1  CCAGCTCAAGCCCCAAAGCCTCACAAGGGCCACCACCATCCAAAACATTCCCCAGCCCCT 180
PiPRP2  CCAGCTCAAGCCCCAAAGCCTCACAAGGGCCACCACCATCCAAAACATTCCCCAGCCCCT 180
PhPRP1  CCAGCTCAAGCCCCAAAGCCTCACAAGGGCCACCACCATCCAAAACATTCCCCAGCCCCT 180
PaPRP1  CCAGCTCAAGCCCCAAAGCCTCACAAGGGCCACCACCATCCAAAACATTCCCCAGCCCCT 180
        ************************************************************

PiPRP1  TCACCAGCAACCCCACCACCAGCTTATAGCCCATCAAAACCACCAGTTAAACCACCTACC 240
PiPRP2  TCACCAGCAACCCCACCACCAGCTTATAGCCCATCAAAACCACCAGTTAAACCACCTACC 240
PhPRP1  TCACCAGCAACCCCACCACCAGCTTATAGCCCATCAAAACCACCAGTTAAACCACCTACC 240
PaPRP1  TCACCAGCAACCCCACCACCAGCTTATAGCCCATCAAAACCACCAGTTAAACCACCTACC 240
        ************************************************************

PiPRP1  CCCTCAGTTAAACCACCAGCTAAGCCACCAGTTAAACCACCTACCCCATCAGTTAAACCA 300
PiPRP2  CCCTCAGTTAAACCACCAGCTAAGCCACCAGTTAAACCACCTACCCCATCAGTTAAACCA 300
PhPRP1  CCCTCAGTTAAACCACCAGCTAAGCCACCAGTTAAACCACCTACCCCATCA--------- 291
PaPRP1  CCCTCAGTTAGACCACCAGCTAAGCCACCAGTTAAACCACCTAGCCCATCA--------- 291
        ********** ************************* ********     *******

PiPRP1  CCAGTTAAACCACCTACCCCATCAGTTAAGCCACCAACACCGTCACCTTATTACCCTTCT 360
PiPRP2  CCAGTTAAACCACCTACCCCATCAGTTAAGCCACCAACACCGTCACCTTATTACCCTTCT 360
PhPRP1  ---GTTAAACCACCTACCCCATCAGTTAAGCCACCAACACCGTCACCTTATTACCCTTCT 348
PaPRP1  ---GTTAAACCACCTACCCCATCAGTTAAGCCACCAACACCGTCACCTTATTACCCTTCT 348
           *********************************************************

PiPRP1  AGGAATCCTGTAGCTGTTCGTGGCCTTGTTTACTGCAAACCTTGCAAGTATAGAGGGGTT 420
PiPRP2  AGGAATCCTGTAGCTGTTCGTGGCCTTGTTTACTGCAAACCTTGCAAGTATAGAGGGGTT 420
PhPRP1  AGGAAACCTGTAGCTGTTCGTGGCCTTGTTTACTGCAAACCTTGCAAGTATAGAGGGGTT 408
PaPRP1  AGGAAACCTGCAGCTGTTCGTGGCCTTGTTTACTGCAAACCTTGCAAGTATAGAGGGGTT 408
        ***** ****  ************************************************

PiPRP1  GAAACTTTAAACCTGGCTACCCCACTCCAGGGAGCGATAGTGAAACTAGCGTGCAACAAC 480
PiPRP2  GAAACTTTAAACCTGGCTACCCCACTCCAGGGAGCGATAGTGAAACTAGCGTGCAACAAC 480
PhPRP1  GAAACTTTAAACCTGGCTACCCCACTCCAGGGAGCGATAGTGAAACTAGCGTGCAACAAC 468
PaPRP1  GAAACTTTAAACCTGGCTACCCCACTCCAGGGAGCGATAGTGAAACTAGCGTGCAACAAC 468
        ************************************************************

PiPRP1  ACAAAGAAGACACTAGTTGAACAGGGCACAACAGACAAGAATGGATTCTTCTTGATCTTG 540
PiPRP2  ACAAAGAAGACACTAGTTGAACAGGGCACAACAGACAAGAATGGATTCTTCTTGATCTTG 540
PhPRP1  ACAAAGAAGACACTAGTTGAACAGGGCACAACAGACAAGAATGGATTCTTCTTGATCTTG 528
PaPRP1  ACAAAGAAGACACTAGTTGAACAGGGCACAACAGACAAGAATGGATTCTTCTTGATCTTG 528
        ************************************************************

PiPRP1  CCCAAAATGTTGTCCTCAGGGGCCTACCACAAATGCAAGGTGTTCTTAGTCTCATCGAAG 600
PiPRP2  CCCAAAATGTTGTCCTCAGGGGCCTACCACAAATGCAAGGTGTTCTTAGTCTCATCGAAG 600
PhPRP1  CCCAAAATGTTGTCCTCAGGGGCCTACCACAAATGCAAGGTGTTCTTAGTCTCATCAAAG 588
PaPRP1  CCCAAAATGTTGTCCTCAGGGGCCTACCACAAATGCAAGGTGTTCTTAGTCTCATCAAAG 588
        ******************************************************** ***

PiPRP1  AATACTCACTGCGATGTCCCAACAAATTTCAATGGTGGAAAATCTGGTGCTCTCTTAAAA 660
PiPRP2  AATACTCACTGCGATGTCCCAACAAATTTCAATGGTGGAAAATCTGGTGCTCTCTTAAAA 660
PhPRP1  AATACTCACTGCGATGTCCCAACAAATTTCAATGGTGGAAAATCTGGTGCTCTCTTAAAA 648
PaPRP1  AATACTCACTGCGATGTCCCAACAAATTTCAATGGTGGAAAATCTGGTGCTCTCTTAAAA 648
        ************************************************************

PiPRP1  TACACCCCACTTCCCAAACCACCAGCCGACTAGTCATCTCCCTGTTAAACCCCCAACATTT 720
PiPRP2  TACACCCCACTTCCCAAGCCACCAGCCGACTAGTCATCTCCCTGTTAAACCCCCAACATTT 720
PhPRP1  TACACCCCACTTCCCAAACCACCAGCCGACTAGTCATCTCCCTGTTAAACCCCCAACATTT 708
PaPRP1  TACACCCCACTTCCCAAACCACCAGCCGACTAGTCATCTCCCTGTTAAACCCCCAACATTT 708
        ***************** ******************************************

PiPRP1  GATGTATTCACTGTGGGGCCTTTTGGATTCGAAGCCTCAACCAAGGTGCCTTGCAAAAAA 780
PiPRP2  GATGTATTCACTGTGGGGCCTTTTGGATTCGAAGCCTCAACCAAGGTGCCTTGCAAAAAA 780
PhPRP1  GATGTATTCACTGTGGGGCCTTTTGGATTCGAAGCCTCAAGCAAGGTGCCTTGCAAAAAA 768
PaPRP1  GATGTATTCACTGTGGGGCCTTTTGGATTCGAAGCCTCAACCAAGGTGCCTTGCAAAAAA 768
        **************************************** ******************

PiPRP1  TAG 783
PiPRP2  TAG 783
PhPRP1  TAG 771
PaPRP1  TAG 771
        ***
```

Figure 9a. Clustal W2 image of MEGA5 alignment of *Petunia* TTS nucleotide sequences from *P. hybrida* (*Ph*PRP1), *P. axillaris* (*Pa*PRP1), and *P. integrifolia* (*Pi*PRP1). Identical nucleotides among all sequences are indicated by (*); (:) indicates one nucleotide difference between the sequences; (.) indicates two or more nucleotide differences between the sequences; ( ) indicates a gap either by an insertion or deletion from one of the sequences. The numbers at the end of the row indicate the amino acid positions.

```
PiPRP1        MAKAFVLFHLSVLLLSSFTVLSHGEGLMGWSLTKHEDHLPPAQAPKPHKGHHHPKHSPAP 60
PiPRP2        MAKAFVLFHLSVLLLSSFTVLSHGEGLMGWSLTKHEDHLPPAQAPKPHKGHHHPKHSPAP 60
PhPRP1        MAKAFVLFHLSVLLLSSFTVLSHGEGLMGWSLTKHEDHLPPAQAPKPHKGHHHPKHSPAP 60
PaPRP1        MAKAFVLFHLSVLLLSSFTVLSHGEGLMGWSLTKHEDHLPPAQAPKPHKGHHHPKHSPAP 60
              ************************************************************

PiPRP1        SPATPPPAYSPSKPPVKPPTPSVKPPAKPPVKPPTPSVKPPVKPPTPSVKPPTPSPYYPS 120
PiPRP2        SPATPPPAYSPSKPPVKPPTPSVKPPAKPPVKPPTPSVKPPVKPPTPSVKPPTPSPYYPS 120
PhPRP1        SPATPPPAYSPSKPPVKPPTPSVKPPAKPPVKPPTPS----VKPPTPSVKPPTPSPYYPS 116
PaPRP1        SPATPPPAYSPSKPPVKPPTPSVRPPAKPPVKPPSPS----VKPPTPSVKPPTPSPYYPS 116
              ***********************:**********:**    *******************

PiPRP1        RNPVAVRGLVYCKPCKYRGVETLNLATPLQGAIVKLACNNTKKTLVEQGTTDKNGFFLIL 180
PiPRP2        RNPVAVRGLVYCKPCKYRGVETLNLATPLQGAIVKLACNNTKKTLVEQGTTDKNGFFLIL 180
PhPRP1        RKPVAVRGLVYCKPCKYRGVETLNLATPLQGAIVKLACNNTKKTLVEQGTTDKNGFFLIL 176
PaPRP1        RKPAAVRGLVYCKPCKYRGVETLNLATPLQGAIVKLACNNTKKTLVEQGTTDKNGFFLIL 176
              *:*.********************************************************

PiPRP1        PKMLSSGAYHKCKVFLVSSKNTHCDVPTNFNGGKSGALLKYTPLPKPPATSHLPVKPPTF 240
PiPRP2        PKMLSSGAYHKCKVFLVSSKNTHCDVPTNFNGGKSGALLKYTPLPKPPATSHLPVKPPTF 240
PhPRP1        PKMLSSGAYHKCKVFLVSSKNTHCDVPTNFNGGKSGALLKYTPLPKPPATSHLPVKPPTF 236
PaPRP1        PKMLSSGAYHKCKVFLVSSKNTHCDVPTNFNGGKSGALLKYTPLPKPPATSHLPVKPPTF 236
              ************************************************************

PiPRP1        DVFTVGPFGFEASTKVPCKK 260
PiPRP2        DVFTVGPFGFEASTKVPCKK 260
PhPRP1        DVFTVGPFGFEASSKVPCKK 256
PaPRP1        DVFTVGPFGFEASTKVPCKK 256
              *************:******
```

Figure 9b. Clustal W2 image of MEGA5 alignment of *Petunia* TTS protein sequences from *P. hybrida* (*Ph*PRP1), *P. axillaris* (*Pa*PRP1), and *P. integrifolia* (*Pi*PRP1 and *Pi*PRP2). Identical amino acids among all sequences are indicated by (*); (:) indicates one nucleotide difference between the sequences; (.) indicates two or more nucleotide differences between the sequences; ( ) indicates a gap either by an insertion or deletion from one of the sequences. The numbers at the end of the row indicate the amino acid positions.

```
PiPRP1      MAKAFVLFHLSVLLLSSFTVLSHGEGLMGWSLTKHEDHLPPAQAPKPHKGHHHPKHSPAP 60
PiPRP2      MAKAFVLFHLSVLLLSSFTVLSHGEGLMGWSLTKHEDHLPPAQAPKPHKGHHHPKHSPAP 60
PhPRP1      MAKAFVLFHLSVLLLSSFTVLSHGEGLMGWSLTKHEDHLPPAQAPKPHKGHHHPKHSPAP 60
FN006765.1  MAKAFVLFHLSVLLLSSFTVLSHGEGLMGWSLTKHEDHLPPAQAPKPHKGHHHPKHSPAP 60
FN003773.1  MAKAFVLFHLSVLLLSSFTVLSHGEGLMGWSLTKHEDHLPPAQAPKPHKGHHHPKHSPAP 60
CV295550.1  MAKAFVLFHLSVLLLSSFTVLSHGEGLMGWSLTKHEDHLPPAQAPKPHKGHHHPKHSPAP 60
PaPRP1      MAKAFVLFHLSVLLLSSFTVLSHGEGLMGWSLTKHEDHLPPAQAPKPHKGHHHPKHSPAP 60
FN015061.1  MAKAFVLFHLSVLLLSSFTVLSHGEGLMGWSLTKHEDHLPPAQAPKPHKGHHHPKHSPAP 60
            ************************************************************

PiPRP1      SPATPPPAYSPSKPPVKPPTPSVKPPAKPPVKPPTPSVKPPVKPPTPSVKPPTPSPYYPS 120
PiPRP2      SPATPPPAYSPSKPPVKPPTPSVKPPAKPPVKPPTPSVKPPVKPPTPSVKPPTPSPYYPS 120
PhPRP1      SPATPPPAYSPSKPPVKPPTPSVKPPAKPPVKPPTPS----VKPPTPSVKPPTPSPYYPS 116
FN006765.1  SPATPPPAYSPSKPPVKPPTPSVKPPAKPPVKPPSPS----VKPPTPSVKPPTPSPYYPS 116
FN003773.1  SPATPPPAYSPSKPPVKPPT-SVKPPAKPPVKPPSPS----VKPPTPSVKPPTPSPYYPS 115
CV295550.1  SPATPPPAYSPSKPPVKPPTPSVKPPAKPPVKPPSPS----VKPPTPSVKPPTPSPYYPS 116
PaPRP1      SPATPPPAYSPSKPPVKPPTPSVRPPAKPPVKPPSPS----VKPPTPSVKPPTPSPYYPS 116
FN015061.1  SPATPPPAYSPSKPPVKPPTPSVKPPAKPPVKPPSPS----VKPPTPSVKPPTPSPYYPS 116
            *******************.**:**********:**      ******************

PiPRP1      RNPVAVRGLVYCKPCKYRGVETLNLATPLQGAIVKLACNNTKKTLVEQGTTDKNGFFLIL 180
PiPRP2      RNPVAVRGLVYCKPCKYRGVETLNLATPLQGAIVKLACNNTKKTLVEQGTTDKNGFFLIL 180
PhPRP1      RKPVAVRGLVYCKPCKYRGVETLNLATPLQGAIVKLACNNTKKTLVEQGTTDKNGFFLIL 176
FN006765.1  RKPVAVRGLVYCKPCKYRGVETLNLATPLQGAIVKLACNNTKKTLVEQGTTDKNGFFLIL 176
FN003773.1  RKPVAVRGLVYCKPCKYRGVETLNLATPLQGAIVKLACNNTKKTLVEQGTTDKNGFFLIL 175
CV295550.1  RKPVAVRGLVYCKPCKYRGVETLNLATPLQGAIVKLACNNTKKTLVEQGTT--------- 167
PaPRP1      RKPAAVRGLVYCKPCKYRGVETLNLATPLQGAIVKLACNNTKKTLVEQGTTDKNGFFLIL 176
FN015061.1  RKPVAVRGLVYCKPCKYRGVETLNLATPLQGAIVKLACNNTKKTLVEQGTTDKNGFFLIL 176
            *:*.****************************************************

PiPRP1      PKMLSSGAYHKCKVFLVSSKNTHCDVPTNFNGGKSGALLKYTPLPKPPATSHLPVKPPTF 240
PiPRP2      PKMLSSGAYHKCKVFLVSSKNTHCDVPTNFNGGKSGALLKYTPLPKPPATSHLPVKPPTF 240
PhPRP1      PKMLSSGAYHKCKVFLVSSKNTHCDVPTNFNGGKSGALLKYTPLPKPPATSHLPVKPPTF 236
FN006765.1  PKMLSSGAYHKCKVFLVSSKNTHCDVP-TFNG-KSGALL-YTPLSQT-A----------- 221
FN003773.1  P----------------------------------------------------------- 176
CV295550.1  ------------------------------------------------------------
PaPRP1      PKMLSSGAYHKCKVFLVSSKNTHCDVPTNFNGGKSGALLKYTPLPKPPATSHLPVKPPTF 236
FN015061.1  PKMLSSGAYHKCKVFLVSSKNTHCDVT-NFNGGKSGALK-YTPLPKT-ATSHLPVPP--F 231


PiPRP1      DVFTVGPFGFEASTKVPCKK 260
PiPRP2      DVFTVGPFGFEASTKVPCKK 260
PhPRP1      DVFTVGPFGFEASSKVPCKK 256
FN006765.1  --------------------
FN003773.1  --------------------
CV295550.1  --------------------
PaPRP1      DVFTVGPFGFEASTKVPCKK 256
FN015061.1  DV-TVGPF------------ 238
```

Figure 10. Clustal W2 image of MEGA5 alignment from BLAST search using *Pa*PRP1. Homologous sequences expressed in pistils: *P. integrifolia* (*Pi*PRP1 and *Pi*PRP2), *P. hybrida* (*Ph*PRP1, FN006765.1, FN003773.1, and CV295550.1), and *P. axillaris* (*Pa*PRP1 and FN015061.1). Sequences expressed in non-pistil tissues: flowers (CV295550.1), corolla tube (FN015061.1), roots (FN003773.1 and FN006765.1). Sequences obtained in this study: pistil (*Ph*PRP1, *Pi*PRP1, and *Pa*PRP1), and seedling leaves (*Pi*PRP2). Identical amino acids among all sequences are indicated by (*); (:) indicates one nucleotide difference between the sequences; (.) indicates two or more nucleotide differences between the sequences; ( ) indicates a gap either by an insertion or deletion from one of the sequences. The numbers at the end of the row indicate the amino acid positions.
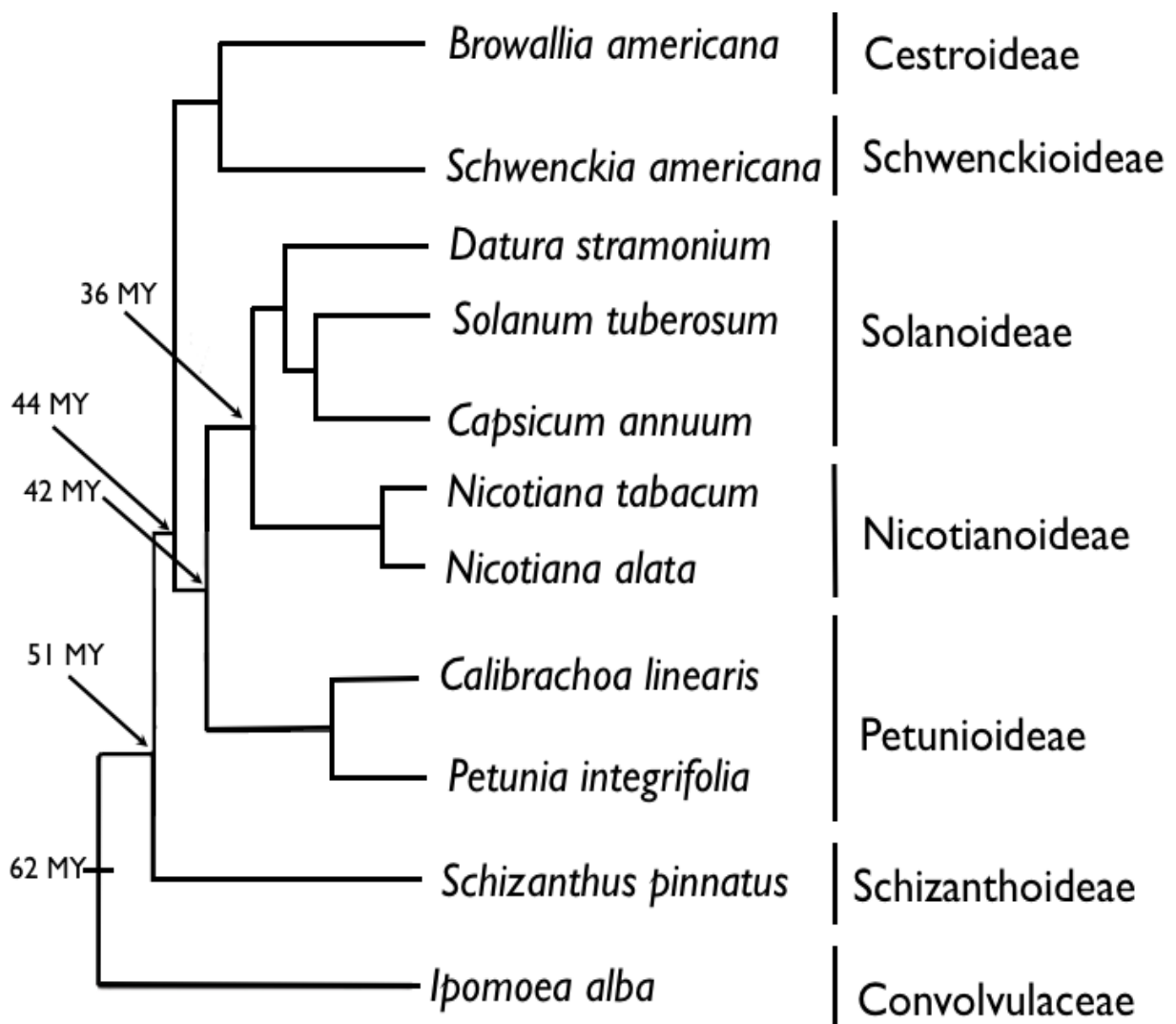
Figure 11. Summary phylogeny of relevant solanaceous taxa re-drawn from phylogenetic trees constructed by Martins and Barkman (2006) and Paape *et al.* (2008). The phylogeny inferred by Martins and Barkman (2006) was based on a nuclear gene encoding a salicylic acid methyltransferase (*SAMT*). Paape *et al.* (2008) constructed a consensus species phylogeny and divergence time estimates of the Solanaceae using Bayesian inference.
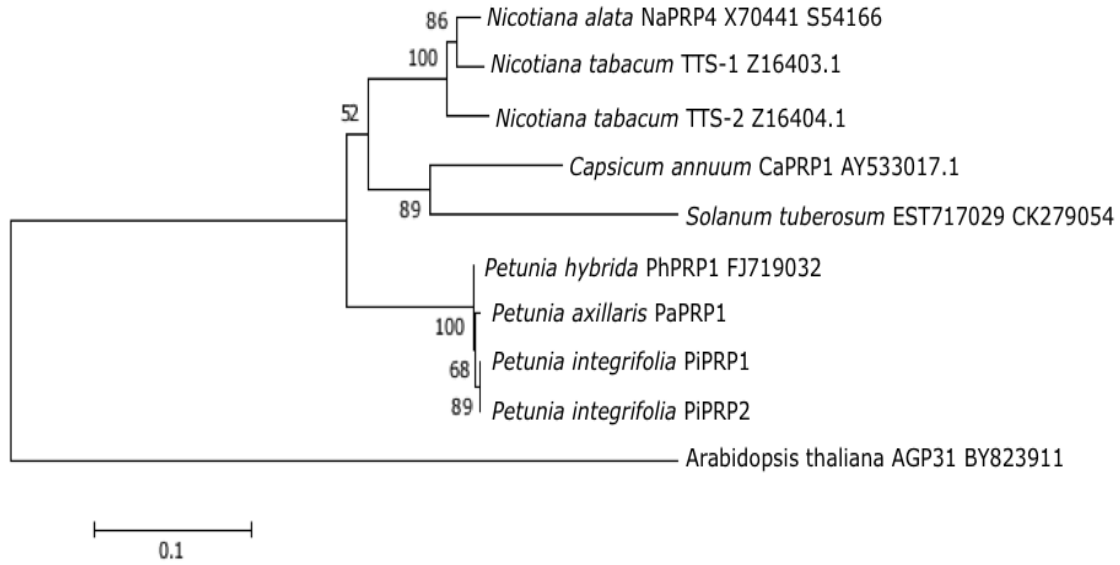
Figure 12. Phylogenetic reconstruction of TTS proteins using the maximum-likelihood statistical method. The Jukes-Cantor model was used with 1,000 bootstrap replications applied in the MEGA5 package. Scale bar represents 10 nucleotide substitutions per 100 positions. Each branch leaf is labeled by species, protein name, and accession number.
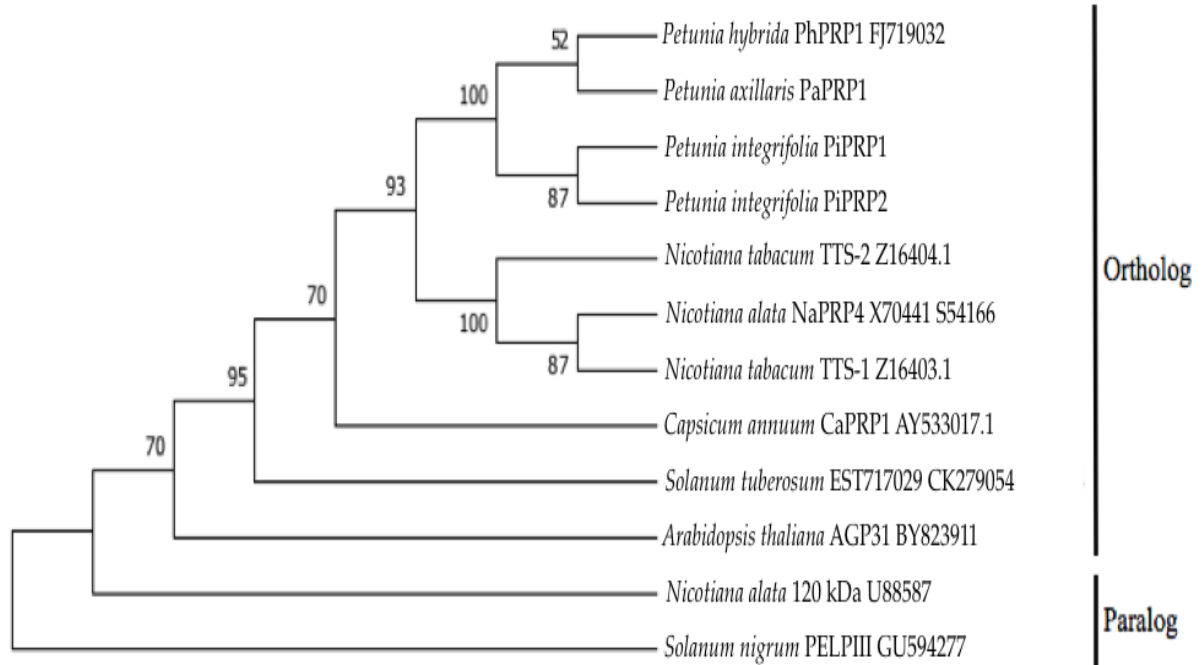
Figure 13. Cladogram consisting of arabinogalactan proteins, TTSP, PELPIII, and 120 kDa glycoprotein, using the maximum parsimony statistical model. The Close-Neighbor-Interchange (CNI) on Random Trees search method was applied using MEGA5 package to construct a maximum parsimony tree with 1,000 bootstrap replications. Each branch leaf is labeled by species, protein name, and accession number.

Table 1. Designed primers used for PCR**.** F refers to forward primer and R refers to reverse primer.

Primer	Sequence	Annealing temp. (°C)	Size (bp)
2PhTTS F	5' GTT CAG CAC AAT TAG TAC TTA GCA A 3'	61.9	25
3PhTTS F	5' CAG TTT TAT TAC TCA GCT CAT TCA CAG TTC 3'	63.3	30
3PhTTS R	5' GGC ACC TTS RTT GAG GCT TCG 3'	65.5	21
4PhTTS F	5' CAC CCA GTT TTA TTA CTC AGC TCA TTC AC 3'	66.6	29
4PhTTS R	5' TGA CGG TGT TGG TGG CTT AAC TGA TG 3'	68.0	26
5PhTTS F	5' CAC CGT TCA GCA CAA TTA GTA CTT AGC 3'	64.6	27
5PhTTS R	5' TCC TAA CTT TCT TTC CCC AAT CAA 3'	59.4	24
Actin F	5' ACA GGT ATT GTG TTG GAC TC 3'	58.4	20
Actin R	5' CTG TAC TTT CTC TCT GGT GG 3'	60.4	20
M13 F	5' TGT AAA ACG ACG GCC AGT 3'	70.0	18
M13 R	5' CAG GAA ACA GCT ATG AC 3'	66.0	17
Oligo(dT)$_{20}$ R	5' TTT TTT TTT TTT TTT TTT TT 3'	50.0	20

Table 2A: Bud pollinations between sister taxa of *Petunia*. Pollinations were conducted on emasculated young buds to determine average seeds per capsule (mean ±standard deviation is shown in the cells below).

	Seeds/capsule	
Pistillate parent	Pollen parent	
	Petunia integrifolia	*Petunia axillaris*
Petunia integrifolia	114.0 ±19.8	43.4 ±32.8
Petunia axillaris	150.8 ±91.1	19.5 ±0.7

Table 2B: Open flower pollinations between sister taxa of *Petunia*. Pollinations were conducted on emasculated open flowers between *P. integrifolia*, *P. axillaris* and *P. hybrida* to determine average seed number per capsule.

	Seeds/capsule		
Pistillate parent	Pollen parent		
	Petunia integrifolia	*Petunia axillaris*	*Petunia hybrida*
Petunia integrifolia	134.0	0.0	71.0
Petunia axillaris	0.0	230.0	257.0
Petunia hybrida	32.8	11.7	302.0

Table 3: Total amino-acid composition of all TTS cDNA sequences using the Datamonkey program. Numbers represent percentage of amino acid in overall cDNA sequence.

Amino-acid composition									
A	6.26	C	2.50	D	2.25	E	2.20	F	4.40
G	5.56	H	3.90	I	1.55	K	11.32	L	7.96
M	1.35	N	3.20	P	16.93	Q	1.75	R	1.65
S	8.11	T	6.81	V	7.81	W	1.05	Y	3.45

Table 4. Amino-acid composition of each TTS cDNA sequence for each species. Numbers represent percentage of amino acid in overall cDNA sequence.

TTS gene	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr	Total
NaTTS	6.69	2.36	2.36	2.36	4.72	5.51	3.54	1.18	12.60	8.27	0.79	2.76	16.54	1.57	1.18	8.27	6.30	9.06	0.79	3.15	254
TTS-1	6.61	2.33	2.33	1.56	4.67	5.06	3.11	1.17	12.84	8.17	0.78	3.11	17.12	1.95	1.17	7.78	7.78	8.56	0.78	3.11	257
TTS-2	6.13	2.30	1.92	1.92	4.21	6.13	2.68	1.53	13.02	7.66	0.77	2.68	18.01	1.53	1.15	8.43	6.51	8.05	0.77	4.60	261
CaPRP1	6.15	2.31	2.31	1.54	3.85	5.00	2.69	1.92	13.08	6.92	0.77	3.46	21.15	1.15	1.54	9.23	5.77	6.54	0.77	3.85	260
PhPRP1	6.25	2.34	1.56	1.95	4.30	5.86	4.67	0.78	11.72	8.98	1.17	2.73	17.97	1.17	1.17	8.59	7.42	8.20	0.39	2.73	256
PjPRP1	6.23	1.95	1.56	1.95	4.28	5.84	4.67	0.78	10.89	8.95	1.17	3.11	18.68	1.17	1.17	8.17	7.78	8.56	0.39	2.72	257
PjPRP2	6.23	1.95	1.56	1.95	4.28	5.84	4.67	0.78	10.89	8.95	1.17	3.11	18.68	1.17	1.17	8.17	7.78	8.56	0.39	2.72	257
PaPRP1	6.30	2.10	1.68	1.68	3.78	5.88	5.04	0.84	10.92	9.24	1.26	2.94	17.65	1.26	1.26	8.82	7.56	8.40	0.42	2.94	238
Average	6.32	2.21	1.91	1.86	4.26	5.64	3.87	1.13	12.01	8.38	0.98	2.99	18.24	1.37	1.23	8.43	7.11	8.24	0.59	3.24	255

Table 5: Pairwise comparisons of synonymous and non-synonymous substitutions of cDNA sequences in Signal Domain I using DnaSP program.

Seq 1	Seq 2	Syn. diff.	Syn. poss.	Ks	Non-syn. Diff.	Non-syn. Poss.	Ka	Ka/Ks
*Na*TTS	TTS-1	0.00	14.00	0.0000	0.00	43.00	0.0000	0.0000
*Na*TTS	TTS-2	0.00	14.00	0.0000	0.00	43.00	0.0000	0.0000
*Na*TTS	*Pi*PRP1	1.00	13.67	0.0770	2.00	43.33	0.0476	0.6182
*Na*TTS	*Pi*PRP2	1.00	13.67	0.0770	2.00	43.33	0.0476	0.6182
*Na*TTS	*Pa*PRP1	1.00	13.67	0.0770	2.00	43.33	0.0476	0.6182
*Na*TTS	*Ph*PRP1	1.00	13.67	0.0770	2.00	43.33	0.0476	0.6182
*Na*TTS	*Ca*PRP1	1.00	13.83	0.0760	2.00	43.17	0.0478	0.6289
TTS-1	TTS-2	0.00	14.00	0.0000	0.00	43.00	0.0000	0.0000
TTS-1	*Pi*PRP1	1.00	13.67	0.0770	2.00	43.33	0.0476	0.6182
TTS-1	*Pi*PRP2	1.00	13.67	0.0770	2.00	43.33	0.0476	0.6182
TTS-1	*Pa*PRP1	1.00	13.67	0.0770	2.00	43.33	0.0476	0.6182
TTS-1	*Ph*PRP1	1.00	13.67	0.0770	2.00	43.33	0.0476	0.6182
TTS-1	*Ca*PRP1	1.00	13.83	0.0760	2.00	43.17	0.0478	0.6289
TTS-2	*Pi*PRP1	1.00	13.67	0.0770	2.00	43.33	0.0476	0.6182
TTS-2	*Pi*PRP2	1.00	13.67	0.0770	2.00	43.33	0.0476	0.6182
TTS-2	*Pa*PRP1	1.00	13.67	0.0770	2.00	43.33	0.0476	0.6182
TTS-2	*Ph*PRP1	1.00	13.67	0.0770	2.00	43.33	0.0476	0.6182
TTS-2	*Ca*PRP1	1.00	13.83	0.0760	2.00	43.17	0.0478	0.6289
*Pi*PRP1	*Pi*PRP2	0.00	13.33	0.0000	0.00	43.67	0.0000	0.0000
*Pi*PRP1	*Pa*PRP1	0.00	13.33	0.0000	0.00	43.67	0.0000	0.0000
*Pi*PRP1	*Ph*PRP1	0.00	13.33	0.0000	0.00	43.67	0.0000	0.0000
*Pi*PRP1	*Ca*PRP1	2.00	13.50	0.1650	4.00	43.50	0.0981	0.5945
*Pi*PRP2	*Pa*PRP1	0.00	13.33	0.0000	0.00	43.67	0.0000	0.0000
*Pi*PRP2	*Ph*PRP1	0.00	13.33	0.0000	0.00	43.67	0.0000	0.0000
*Pi*PRP2	*Ca*PRP1	2.00	13.50	0.1650	4.00	43.50	0.0981	0.5945
*Pa*PRP1	*Ph*PRP1	0.00	13.33	0.0000	0.00	43.67	0.0000	0.0000
*Pa*PRP1	*Ca*PRP1	2.00	13.50	0.1650	4.00	43.50	0.0981	0.5945
*Ph*PRP1	*Ca*PRP1	2.00	13.50	0.1650	4.00	43.50	0.0981	0.5945

Table 6: Pairwise comparisons of synonymous and non-synonymous substitutions of cDNA sequences in Hypervariable Domain I using DnaSP program.

Seq 1	Seq 2	Syn. diff.	Syn. poss.	Ks	Non-syn. Diff.	Non-syn. Poss.	Ka	Ka/Ks
*Na*TTS	TTS-1	1.00	9.00	0.1203	5.00	42.00	0.1296	1.0770
*Na*TTS	TTS-2	0.00	8.50	0.0000	5.00	42.50	0.1280	not-defined
*Na*TTS	*Pi*PRP1	2.00	9.50	0.2471	11.00	41.50	0.3270	1.3234
*Na*TTS	*Pi*PRP2	2.00	9.50	0.2471	11.00	41.50	0.3270	1.3234
*Na*TTS	*Pa*PRP1	2.00	9.50	0.2471	11.00	41.50	0.3270	1.3234
*Na*TTS	*Ph*PRP1	2.00	9.50	0.2471	11.00	41.50	0.3270	1.3234
*Na*TTS	*Ca*PRP1	2.50	9.17	0.3390	12.50	41.83	0.3811	1.1240
TTS-1	TTS-2	1.00	8.83	0.1227	6.00	42.17	0.1578	1.2861
TTS-1	*Pi*PRP1	1.50	9.83	0.1705	9.50	41.17	0.2758	1.6176
TTS-1	*Pi*PRP2	1.50	9.83	0.1705	9.50	41.17	0.2758	1.6176
TTS-1	*Pa*PRP1	1.50	9.83	0.1705	9.50	41.17	0.2758	1.6176
TTS-1	*Ph*PRP1	1.50	9.83	0.1705	9.50	41.17	0.2758	1.6176
TTS-1	*Ca*PRP1	3.17	9.50	0.4408	14.83	41.50	0.4855	1.1014
TTS-2	*Pi*PRP1	1.50	9.33	0.1809	11.50	41.67	0.3441	1.9021
TTS-2	*Pi*PRP2	1.50	9.33	0.1809	11.50	41.67	0.3441	1.9021
TTS-2	*Pa*PRP1	1.50	9.33	0.1809	11.50	41.67	0.3441	1.9021
TTS-2	*Ph*PRP1	1.50	9.33	0.1809	11.50	41.67	0.3441	1.9021
TTS-2	*Ca*PRP1	2.50	9.00	0.3470	12.50	42.00	0.3792	1.0928
*Pi*PRP1	*Pi*PRP2	0.00	10.33	0.0000	0.00	40.67	0.0000	0.0000
*Pi*PRP1	*Pa*PRP1	0.00	10.33	0.0000	0.00	40.67	0.0000	0.0000
*Pi*PRP1	*Ph*PRP1	0.00	10.33	0.0000	0.00	40.67	0.0000	0.0000
*Pi*PRP1	*Ca*PRP1	4.50	10.00	0.6872	8.50	41.00	0.2427	0.3532
*Pi*PRP2	*Pa*PRP1	0.00	10.33	0.0000	0.00	40.67	0.0000	0.0000
*Pi*PRP2	*Ph*PRP1	0.00	10.33	0.0000	0.00	40.67	0.0000	0.0000
*Pi*PRP2	*Ca*PRP1	4.50	10.00	0.6872	8.50	41.00	0.2427	0.3532
*Pa*PRP1	*Ph*PRP1	0.00	10.33	0.0000	0.00	40.67	0.0000	0.0000
*Pa*PRP1	*Ca*PRP1	4.50	10.00	0.6872	8.50	41.00	0.2427	0.35317
*Ph*PRP1	*Ca*PRP1	4.50	10.00	0.6872	8.50	41.00	0.2427	0.3532

Table 7: Pairwise comparisons of synonymous and non-synonymous substitutions of cDNA sequences in Conserved Domain I using DnaSP program.

Seq 1	Seq 2	Syn. diff.	Syn. poss.	Ks	Non-syn. Diff.	Non-syn. Poss.	Ka	Ka/Ks
*Na*TTS	TTS-1	1.00	16.00	0.0653	1.00	50.00	0.0203	0.3109
*Na*TTS	TTS-2	3.00	16.67	0.2058	4.00	49.33	0.0858	0.4169
*Na*TTS	*Pi*PRP1	4.00	16.00	0.3041	0.00	50.00	0.0000	0.0000
*Na*TTS	*Pi*PRP2	4.00	16.00	0.3041	0.00	50.00	0.0000	0.0000
*Na*TTS	*Pa*PRP1	4.00	16.00	0.3041	0.00	50.00	0.0000	0.0000
*Na*TTS	*Ph*PRP1	4.00	16.00	0.3041	0.00	50.00	0.0000	0.0000
*Na*TTS	*Ca*PRP1	4.00	16.33	0.2965	2.00	49.67	0.0414	0.1396
TTS-1	TTS-2	2.00	16.67	0.1308	4.00	49.33	0.0858	0.6560
TTS-1	*Pi*PRP1	5.00	16.00	0.4042	1.00	50.00	0.0203	0.0502
TTS-1	*Pi*PRP2	5.00	16.00	0.4042	1.00	50.00	0.0203	0.0502
TTS-1	*Pa*PRP1	5.00	16.00	0.4042	1.00	50.00	0.0203	0.0502
TTS-1	*Ph*PRP1	5.00	16.00	0.4042	1.00	50.00	0.0203	0.0502
TTS-1	*Ca*PRP1	3.00	16.33	0.2107	3.00	49.67	0.0630	0.2990
TTS-2	*Pi*PRP1	5.00	16.67	0.3831	4.00	49.33	0.0858	0.2240
TTS-2	*Pi*PRP2	5.00	16.67	0.3831	4.00	49.33	0.0858	0.2240
TTS-2	*Pa*PRP1	5.00	16.67	0.3831	4.00	49.33	0.0858	0.2240
TTS-2	*Ph*PRP1	5.00	16.67	0.3831	4.00	49.33	0.0858	0.2240
TTS-2	*Ca*PRP1	3.00	17.00	0.2012	5.00	49.00	0.1097	0.5452
*Pi*PRP1	*Pi*PRP2	0.00	16.00	0.0000	0.00	50.00	0.0000	0.0000
*Pi*PRP1	*Pa*PRP1	0.00	16.00	0.0000	0.00	50.00	0.0000	0.0000
*Pi*PRP1	*Ph*PRP1	0.00	16.00	0.0000	0.00	50.00	0.0000	0.0000
*Pi*PRP1	*Ca*PRP1	4.00	16.33	0.2965	2.00	49.67	0.0414	0.1400
*Pi*PRP2	*Pa*PRP1	0.00	16.00	0.0000	0.00	50.00	0.0000	0.0000
*Pi*PRP2	*Ph*PRP1	0.00	16.00	0.0000	0.00	50.00	0.0000	0.0000
*Pi*PRP2	*Ca*PRP1	4.00	16.33	0.2965	2.00	49.67	0.0414	0.1400
*Pa*PRP1	*Ph*PRP1	0.00	16.00	0.0000	0.00	50.00	0.0000	0.0000
*Pa*PRP1	*Ca*PRP1	4.00	16.33	0.2965	2.00	49.67	0.0414	0.1400
*Ph*PRP1	*Ca*PRP1	4.00	16.33	0.2965	2.00	49.67	0.0414	0.1400

Table 8: Pairwise comparisons of synonymous and non-synonymous substitutions of cDNA sequences in Hypervariable Domain II using DnaSP program.

Seq 1	Seq 2	Syn. diff.	Syn. poss.	Ks	Non-syn. Diff.	Non-syn. Poss.	Ka	Ka/Ks
*Na*TTS	TTS-1	2.00	41.33	0.0500	2.00	105.67	0.1296	0.3840
*Na*TTS	TTS-2	8.00	41.33	0.2239	2.00	105.67	0.1280	0.0857
*Na*TTS	*Pi*PRP1	12.00	42.33	0.3561	15.00	104.67	0.3270	0.4465
*Na*TTS	*Pi*PRP2	12.00	42.33	0.3561	15.00	104.67	0.3270	0.4465
*Na*TTS	*Pa*PRP1	13.00	42.25	0.3961	15.00	104.75	0.3270	0.4012
*Na*TTS	*Ph*PRP1	13.00	42.33	0.3950	14.00	104.67	0.3270	0.3729
*Na*TTS	*Ca*PRP1	22.75	42.67	0.9308	28.25	104.33	0.3811	0.3609
TTS-1	TTS-2	8.00	41.33	0.2239	2.00	105.67	0.1578	0.0857
TTS-1	*Pi*PRP1	12.00	42.33	0.3561	16.00	104.67	0.2758	0.4799
TTS-1	*Pi*PRP2	12.00	42.33	0.3561	16.00	104.67	0.2758	0.4799
TTS-1	*Pa*PRP1	13.00	42.25	0.3961	16.00	104.75	0.2758	0.4312
TTS-1	*Ph*PRP1	13.00	42.33	0.3950	15.00	104.67	0.2758	0.4025
TTS-1	*Ca*PRP1	21.75	42.67	0.8538	28.25	104.33	0.4855	0.3934
TTS-2	*Pi*PRP1	15.00	42.33	0.4796	16.00	104.67	0.3441	0.3563
TTS-2	*Pi*PRP2	15.00	42.33	0.4796	16.00	104.67	0.3441	0.3563
TTS-2	*Pa*PRP1	16.00	42.25	0.5273	16.00	104.75	0.3441	0.3239
TTS-2	*Ph*PRP1	16.00	42.33	0.5258	15.00	104.67	0.3441	0.3024
TTS-2	*Ca*PRP1	16.75	42.67	0.5559	26.25	104.33	0.3792	0.5514
*Pi*PRP1	*Pi*PRP2	0.00	43.33	0.0000	0.00	103.67	0.0000	0.0000
*Pi*PRP1	*Pa*PRP1	1.00	43.25	0.0235	4.00	103.75	0.0000	1.6850
*Pi*PRP1	*Ph*PRP1	1.00	43.33	0.0234	1.00	103.67	0.0097	0.4145
*Pi*PRP1	*Ca*PRP1	19.00	43.67	0.6509	27.00	103.33	0.3212	0.4935
*Pi*PRP2	*Pa*PRP1	1.00	43.25	0.0235	4.00	103.75	0.0396	1.6851
*Pi*PRP2	*Ph*PRP1	1.00	43.33	0.0234	1.00	103.67	0.0097	0.4145
*Pi*PRP2	*Ca*PRP1	19.00	43.67	0.6509	27.00	103.33	0.3212	0.4935
*Pa*PRP1	*Ph*PRP1	0.00	43.25	0.0000	3.00	103.75	0.0295	not-defined
*Pa*PRP1	*Ca*PRP1	20.50	43.58	0.7399	27.50	103.42	0.3284	0.4438
*Ph*PRP1	*Ca*PRP1	20.00	43.67	0.7075	26.00	103.33	0.3065	0.4332

Table 9: Pairwise comparisons of synonymous and non-synonymous substitutions of cDNA sequences in Conserved Domain II using DnaSP program.

Seq 1	Seq 2	Syn. diff.	Syn. poss.	Ks	Non-syn. Diff.	Non-syn. Poss.	Ka	Ka/Ks
*Na*TTS	TTS-1	8.00	95.50	0.0888	1.00	297.50	0.0034	0.0383
*Na*TTS	TTS-2	10.00	96.17	0.1119	5.00	296.83	0.0170	0.1519
*Na*TTS	*Pi*PRP1	27.50	96.33	0.3593	22.50	296.67	0.0800	0.2227
*Na*TTS	*Pi*PRP2	26.50	96.33	0.3427	22.50	296.67	0.0800	0.2334
*Na*TTS	*Pa*PRP1	26.50	96.33	0.3427	22.50	296.67	0.0800	0.2334
*Na*TTS	*Ph*PRP1	26.00	96.00	0.3360	23.00	297.00	0.0817	0.2432
*Na*TTS	*Ca*PRP1	32.83	95.83	0.4577	34.17	297.17	0.0125	0.0273
TTS-1	TTS-2	10.00	96.17	0.1119	4.00	296.83	0.0136	0.1215
TTS-1	*Pi*PRP1	26.50	96.33	0.3427	21.50	296.67	0.0762	0.2223
TTS-1	*Pi*PRP2	25.50	96.33	0.3265	21.50	296.67	0.0762	0.2334
TTS-1	*Pa*PRP1	25.50	96.33	0.3265	21.50	296.67	0.0762	0.2334
TTS-1	*Ph*PRP1	27.00	96.00	0.3525	22.00	297.00	0.0780	0.2213
TTS-1	*Ca*PRP1	30.83	95.83	0.4203	33.17	297.17	0.1208	0.2874
TTS-2	*Pi*PRP1	25.00	97.00	0.3158	20.00	296.00	0.0708	0.2242
TTS-2	*Pi*PRP2	24.00	97.00	0.3002	20.00	296.00	0.0708	0.2358
TTS-2	*Pa*PRP1	24.00	97.00	0.3002	20.00	296.00	0.0708	0.2358
TTS-2	*Ph*PRP1	23.50	96.67	0.2938	20.50	296.33	0.0726	0.2471
TTS-2	*Ca*PRP1	28.00	96.50	0.3669	33.00	296.50	0.1205	0.3284
*Pi*PRP1	*Pi*PRP2	1.00	97.17	0.0104	0.00	295.83	0.0000	0.0000
*Pi*PRP1	*Pa*PRP1	1.00	97.17	0.0104	0.00	295.83	0.0000	0.0000
*Pi*PRP1	*Ph*PRP1	3.00	96.83	0.0316	1.00	296.17	0.0034	0.1076
*Pi*PRP1	*Ca*PRP1	35.33	96.67	0.5011	39.67	296.33	0.1474	0.2942
*Pi*PRP2	*Pa*PRP1	2.00	97.17	0.0209	0.00	295.83	0.0000	0.0000
*Pi*PRP2	*Ph*PRP1	4.00	96.83	0.0425	1.00	296.17	0.0034	0.0800
*Pi*PRP2	*Ca*PRP1	34.33	96.67	0.4812	39.67	296.33	0.1474	0.3063
*Pa*PRP1	*Ph*PRP1	2.00	96.83	0.0209	1.00	296.17	0.0034	0.1627
*Pa*PRP1	*Ca*PRP1	34.33	96.67	0.4812	39.67	296.33	0.1474	0.3063
*Ph*PRP1	*Ca*PRP1	33.83	96.33	0.4737	40.17	296.67	0.1493	0.3152

Table 10: Pairwise comparisons of synonymous and non-synonymous substitutions of entire cDNA sequences using DnaSP program.

Seq 1	Seq 2	Syn. diff.	Syn. poss.	Ks	Non-syn. Diff.	Non-syn. Poss.	Ka	Ka/Ks
*Na*TTS	TTS-1	12.00	176.83	0.0711	9.00	540.17	0.0168	0.2360
*Na*TTS	TTS-2	21.00	177.67	0.1286	16.00	539.33	0.0303	0.2356
*Na*TTS	*Pi*PRP1	46.50	178.83	0.3193	50.50	538.17	0.1002	0.3138
*Na*TTS	*Pi*PRP2	45.50	178.83	0.3108	50.50	538.17	0.1002	0.3224
*Na*TTS	*Pa*PRP1	46.50	178.75	0.3195	50.50	538.25	0.1002	0.3136
*Na*TTS	*Ph*PRP1	46.00	178.50	0.3157	50.00	538.50	0.0991	0.3139
*Na*TTS	*Ca*PRP1	63.08	178.83	0.4766	78.92	538.17	0.1632	0.3424
TTS-1	TTS-2	21.00	178.00	0.1284	16.00	539.00	0.0303	0.2360
TTS-1	*Pi*PRP1	46.00	179.17	0.3143	50.00	537.83	0.0993	0.3159
TTS-1	*Pi*PRP2	45.00	179.17	0.3058	50.00	537.83	0.0993	0.3247
TTS-1	*Pa*PRP1	46.00	179.08	0.3145	50.00	537.92	0.0992	0.3154
TTS-1	*Ph*PRP1	47.50	178.83	0.3279	49.50	538.17	0.0981	0.2992
TTS-1	*Ca*PRP1	59.75	179.17	0.4411	81.25	537.83	0.1687	0.3825
TTS-2	*Pi*PRP1	47.50	180.00	0.3252	53.50	537.00	0.1069	0.3287
TTS-2	*Pi*PRP2	46.50	180.00	0.3167	53.50	537.00	0.1069	0.3375
TTS-2	*Pa*PRP1	47.50	179.92	0.3254	53.50	537.08	0.1069	0.3285
TTS-2	*Ph*PRP1	47.00	179.67	0.3217	53.00	537.33	0.1058	0.3289
TTS-2	*Ca*PRP1	51.25	180.00	0.3581	78.75	537.00	0.1632	0.4557
*Pi*PRP1	*Pi*PRP2	1.00	181.17	0.0055	0.00	535.83	0.0000	0.0000
*Pi*PRP1	*Pa*PRP1	2.00	181.08	0.0111	4.00	535.92	0.0075	0.6757
*Pi*PRP1	*Ph*PRP1	4.00	180.83	0.0225	2.00	536.17	0.0037	0.1644
*Pi*PRP1	*Ca*PRP1	64.83	181.17	0.4864	81.17	535.83	0.1692	0.3479
*Pi*PRP2	*Pa*PRP1	3.00	181.08	0.0168	4.00	535.92	0.0075	0.4464
*Pi*PRP2	*Ph*PRP1	5.00	180.83	0.0282	2.00	536.17	0.0037	0.1312
*Pi*PRP2	*Ca*PRP1	63.83	181.17	0.4759	81.17	535.83	0.1692	0.3555
*Pa*PRP1	*Ph*PRP1	2.00	180.75	0.0111	4.00	536.25	0.0075	0.6757
*Pa*PRP1	*Ca*PRP1	65.33	181.08	0.4920	81.67	535.92	0.1703	0.3461
*Ph*PRP1	*Ca*PRP1	64.33	180.83	0.4823	80.67	536.17	0.1679	0.3481

Table 11: Summary of Ka/Ks ratios between pairwise comparisons in domains: Signal I, Hypervariable I, Conserved I (CVI), Hypervariable II, and Conserved II (CVII). Positive selection is highlighted.

Seq 1	Seq 2	Signal I	HVI	CVI	HVII	CVII
*Na*TTS	TTS-1	0.0000	1.0770	0.3109	0.3840	0.0383
*Na*TTS	TTS-2	0.0000	not-defined	0.4169	0.0857	0.1519
*Na*TTS	*Pi*PRP1	0.6182	1.3234	0.0000	0.4465	0.2227
*Na*TTS	*Pi*PRP2	0.6182	1.3234	0.0000	0.4465	0.2334
*Na*TTS	*Pa*PRP1	0.6182	1.3234	0.0000	0.4012	0.2334
*Na*TTS	*Ph*PRP1	0.6182	1.3234	0.0000	0.3729	0.2432
*Na*TTS	*Ca*PRP1	0.6289	1.1240	0.1396	0.3609	0.0273
TTS-1	TTS-2	0.0000	1.2861	0.6560	0.0857	0.1215
TTS-1	*Pi*PRP1	0.6182	1.6176	0.0502	0.4799	0.2223
TTS-1	*Pi*PRP2	0.6182	1.6176	0.0502	0.4799	0.2334
TTS-1	*Pa*PRP1	0.6182	1.6176	0.0502	0.4312	0.2334
TTS-1	*Ph*PRP1	0.6182	1.6176	0.0502	0.4025	0.2213
TTS-1	*Ca*PRP1	0.6289	1.1014	0.2990	0.3934	0.2874
TTS-2	*Pi*PRP1	0.6182	1.9021	0.2240	0.3563	0.2242
TTS-2	*Pi*PRP2	0.6182	1.9021	0.2240	0.3563	0.2358
TTS-2	*Pa*PRP1	0.6182	1.9021	0.2240	0.3239	0.2358
TTS-2	*Ph*PRP1	0.6182	1.9021	0.2240	0.3024	0.2471
TTS-2	*Ca*PRP1	0.6289	1.0928	0.5452	0.5514	0.3284
*Pi*PRP1	*Pi*PRP2	0.0000	0.0000	0.0000	0.0000	0.0000
*Pi*PRP1	*Pa*PRP1	0.0000	0.0000	0.0000	1.6850	0.0000
*Pi*PRP1	*Ph*PRP1	0.0000	0.0000	0.0000	0.4145	0.1076
*Pi*PRP1	*Ca*PRP1	0.5945	0.3532	0.1400	0.4935	0.2942
*Pi*PRP2	*Pa*PRP1	0.0000	0.0000	0.0000	1.6851	0.0000
*Pi*PRP2	*Ph*PRP1	0.0000	0.0000	0.0000	0.4145	0.0800
*Pi*PRP2	*Ca*PRP1	0.5945	0.3532	0.1400	0.4935	0.3063
*Pa*PRP1	*Ph*PRP1	0.0000	0.0000	0.0000	not-defined	0.1627
*Pa*PRP1	*Ca*PRP1	0.5945	0.35317	0.1400	0.4438	0.3063
*Ph*PRP1	*Ca*PRP1	0.5945	0.3532	0.1400	0.4332	0.3152

LITERATURE CITED

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic Local Alignment Search Tool. J Mol Biol 215:403-410.

Ando T, Nomura M, Tsukahara J, Watanabe H, Kokubun H, Tsukamoto T, Hashimoto G, Marchesi E, Kitching IJ. 2001. Reproductive isolation in a native population of *Petunia sensu* Jussieu (Solanaceae). Ann Bot 88:403-413.

Baldwin TC, Domingo C, Schindler T, Seetharaman G, Stacey N, Roberts K. 2001. DcAGP1, a secreted arabinogalactan protein, is related to a family of basic proline-rich proteins. Plant Mol Biol 45:421–435.

Bland MM, Matzinger DF, Levings CS. 1985. Comparison of the mitochondrial genome of Nicotiana tabacum with its progenitor species. TAG Theor Appl Gen 69:535-541.

Boguski MS, Lowe TM, Tolstoshev CM. 1993. dbEST—database for "expressed sequence tags." Nat Genet 4(4):332-333.

Butlin RK. 1987. Speciation by reinforcement. Trends Eco Evol 2:8-13.

Chen C-G, Mau S-L, Clarke A. 1993. Nucleotide sequence and style-specific expression of a novel proline-rich protein gene from *Nicotiana alata*. Plant Mol Biol 21:391-395.

Cheung AY. 1993. Floral extracellular matrix proteins. Flowering Newslet 15:3-13.

Cheung AY, May B, Kawata EE, Gu Q, Wu H-M. 1993. Characterization of cDNAs for stylar transmitting tissue-specific proline-rich proteins in tobacco. Plant J 3:151-160.

Cheung AY. 1995. Pollen-pistil interactions in compatible pollination. Proc Natl Acad Sci USA 92:3077-3080.

Cheung AY, Wang H, Wu H-M. 1995. A floral transmitting tissue specific glycoprotein attacks pollen tubes and stimulates their growth. Cell 82:383-393.

Cheung AY, Zhan X-Y, Wang H, Wu H-M. 1996. Organ-specific and Agamous-regulated expression and glycosylation of a pollen tube growth-promoting protein. Proc Natl Acad Sci 93:3853-3858.

Cheung AY. 1996. Pollen-pistil interactions during pollen-tube growth. Trends Plant Sci 1:45-51.

Cheung AY, Wu H-M. 1999. Arabinogalactan proteins in plant sexual reproduction. Protoplasma 208:87-98.

Cheung AY, Wu H-M, Di Stilio V, Glaven R, Chen C, Wong E, Ogdahl J, Estavillo A. 2000. Pollen-pistil interactions in *Nicotiana tabacum*. Ann Bot 85:29-37.

Christie P, Macnair MR. 1987. The distribution of postmating reproductive isolating genes in populations of the yellow monkey flower, *Mimulus guttatus*. Evol 41(3):571-578.

Clark NL, Aagaard JE, Swanson WJ. 2006. Evolution of reproductive proteins from animals and plants. Reprod Rev 131:11-22.

Coyne JA, Orr HA. 1998. The evolutionary genetics of speciation. Philos Trans R Soc Lond B Biol Sci 353:287–305.

Coyne J, Orr HA. 2004. *Speciation.* Sunderland, MA: Sinauer Associates.

Cruz-Garcia F, Hancock CN, McClure B. 2003. S-RNase complexes and pollen rejection. J Exp Bot 54:123-130.

Dell'Olivo A, Hoballah ME, Gubitz T, Kuhlemeier C. 2011. Isolation barriers between *Petunia axillaris and Petunia integrifolia* (Solanaceae). Evol 65(7):1979-1991.

De Nettancourt D. 2001. Incompatability and incongruity in wild and cultivated plants. Spring-Verlag, Berlin.

Dobzhansky T. 1937. Genetics and the origin of species. Columbia Univ. Press, New York.

Dobzhansky T. 1940. Speciation as a stage in evolutionary divergence. Am Nat 74:312-321.

Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368-376.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783-791.

Fitch WM. 1977. On the problem of discovering the most parsimonious tree. Am Nat 111:223-257.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 31:3784-3788.

Gerats T, Vandenbussche M. 2005. A model system for comparative research: *Petunia*. Trends Plant Sci 10:251-256.

Goldman MHS, Goldberg RB, Mariani C. 1992. Developmental expression of tobacco pistil-specific genes encoding novel extensin-like proteins. Plant Cell 4:1041-1051.

Gore JM, Ran FA, Ornston LN. 2006. Deletion mutations caused by DNA strand slippage in *Acinetobacter baylyi*. Appl Environ Microbiol 72:5239-5245.

Grant V. 1965. Evidence for the selective origin of incompatibility barriers in the leafy-stemmed gilias. Proc Natl Acad Sci USA 54:1567-1571.

Grant V. 1981. Plant speciation. Columbia University Press, New York, USA.

Hawkes JG. 1999. The economic importance of the family Solanaceae. In: Nee M, Symon D, Lester RN, Jessop JP (eds) Solanaceae IV - advances in taxonomy and utilization. Royal Botanic Gardens, Kew, pp 1-8.

Herrero M, Dickinson HG. 1979. Pollen-pistil incompatibility in *Petunia hybrida*: changes in the pistil following compatible and incompatible intraspecific crosses. J Cell Sci 36:1-18.

Herrero M, Dickinson HG. 1980. Pollen tube growth following compatible and incompatible intraspecific pollinations in *Petunia hybrida.* Planta 148:217-221.

Heslop-Harrison J. 1987. Pollen germination and pollen-tube growth. Int Rev Cytol 107:1-78.

Heslop-Harrison Y. 1981. Stigma characteristics and angiosperm taxonomy. Nordic J Bot 1(3):1756-1851.

Heslop-Harrison Y, Shivanna KR. 1977. The receptive surface of the angiosperm stigma. Ann Bot 41:1233-1258.

Higgs PG, Attwood TK. 2005. <u>Bioinformatics and Molecular Evolution</u>. United Kingdom: Blackwell Publishing.

Hiscock SJ, Allen AM. 2008. Diverse cell signaling pathways regulate pollen-stigma interactions: the search for consensus. New Phytol 179:286-317.

Hoballah ME, Stuurman J, Turlings T, Guerin PM, Connétable S, Kuhlemeier C. 2005. The composition and timing of flower odour emission by wild *Petunia axillaris* coincide with the antennal perception and nocturnal activity of the pollinator *Manduca sexta*. Planta 222:141-150.

Hogenboom NG. 1975. Incompatibility and incongruity: two different mechanisms for the non-functioning of intimate partner relationships. Proc R Soc Lond B 188:361-375.

Hopkins R, Rausher MD. 2012. Pollinator-mediated selection on flower color allele drives reinforcement. Science 335:1090-1092.

Hormazo JI, Herrero M. 1996. Dynamics of pollen tube growth under different competition regimes. Sex Plant Reprod 9:153-160.

Hurst LD. 2002. The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet 18(9):486.

Kay KM, Schemske DW. 2008. Natural selection reinforces speciation in a radiation of neotropical rainforest plants. Evol 62:2628-2642.

Knapp S. 2010. On 'various contrivances': pollination, phylogeny and flower form in the Solanaceae. Phil Trans R Soc B 365:449-460.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23(21):2947-2948.

Lee CB, Swatek KN, McClure B. 2008. Pollen proteins bind to the C-terminal domain of *Nicotiana alata* pistil arabinogalactan proteins. J of Biol Chem 283:26965-26972.

Liu C, Mehdy MC. 2007. A nonclassical arabinogalactan protein gene highly expressed in vascular tissues, *AGP31*, is transcriptionally repressed by methyl jasmonic acid in *Arabidopsis*. Plant Phys 45:863-874.

Lind JL, Bacic A, Clarke AE, Anderson MA. 1994. A style-specific hydroxyproline-rich glycoprotein with properties of both extensins and arabinogalactan proteins. Plant J 6:491-502.

Lord EM, Russell SD. 2002. The mechanisms of pollination and fertilization in plants. Ann Rev Cell Dev Biol 18:81-105.

Lord EM, Sanders LC. 1992. Roles for the extracellular matrix in plant development and pollination: a special case of cell movement in plants. Dev Biol 153:16-28.

Lorenz-Lemke AP, Mäder G, Muschner C, Stehmann JR, Bonatto SL, Salzano FM, Freitas, LB. 2006. Diversity and natural hybridization in a highly endemic species of *Petunia* (Solanaceae): a molecular and ecological analysis. Mol Eco 15:4487-4497.

Lowry DB, Modliszewski JL, Wright KM, Wu CA, Willis JH. 2008. The strength and genetic basis of reproductive isolating barriers in flowering plants. Phil Trans R Soc B 363: 3009-3021.

Lubliner N, Singh-Cundy DT, Singh-Cundy A. 2003. Characterization of the pollen growth transition (PGT) in self-incompatible *Petunia inflata*. Sex Plant Reprod 15:243–253.

Mang HG, Lee JH, Park JA, Pyee J, Pai HS, Lee J, Kim WT. 2004. The *CaPRP1* gene encoding a putative proline-rich glycoprotein is highly expressed in rapidly elongating early roots and leaves in hot pepper. Biochimika et Biophysica Acta. 1674:103-108.

Martins TR, Barkman TJ. 2005. Reconstruction of Solanaceae phylogeny using the nuclear gene *SAMT*. Syst Bot 30:433-445.

Mather K, Edwardes PMJ. 1943. Specific differences in *Petunia*. III. Flower colour and genetic isolation. J Genet 45:243-260.

Mayr E. 1942. Systematics and the origin of species. Columbia Univ. Press, New York.

Mayr E. 1963. Animal species and evolution. Harvard University Press, Cambridge, Massachusetts.

McClure B. 2009. Darwin's foundation for investigating self-incompatibility and the progress toward a physiological model for S-RNase-based SI. J Exp Bot 60:1069-1081.

Mulcahy GB, Mulcahy DL. 1982. The two phases of growth of *Petunia hybrida* (Hort. Vilm-Andz.) pollen tubes through compatible styles. J Palynol 18:61-64.

Nasrallah JB. 2002. Recognition and rejection of self in plant reproduction. Science 296:305-308.

Ottaviano E, Sari-Goria M, Mulcahy DL. 1980. Pollen tube growth rates in *Zea mays*: implications for genetic improvement of crops. Science 210(4468):437-8.

94

Paape T, Igic B, Smith SD, Olmstead R, Bohs L, Kohn JR. A 15-myr-old genetic bottleneck. Mol Biol Evol 25(4):655-663.

Rieseberg LH, Blackman BK. 2010. Speciation genes in plants. Ann Bot 106(3):439-455.

Rieseberg LH, Willis, JH. 2007. Plant speciation. Science 317:910-914.

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19(18):2496-2497.

Sakurai N. 1998. Dynamic function and regulation of apoplast in the plant body. J Plant Res 111:133-148.

Sanchez AM, Bosch M, Bots M, Nieuwland R, Mariana C. 2004. Pistil factors controlling pollination. Plant Cell 16:S98-S106.

Sanders LC, Lord EM. 1989. Directed movement of latex particles in the gynoecia of three species of flowering plants. Science 243:1606-1608.

Singh A, Ai Y, Kao TH. 1991. Characterization of ribonuclease activity of three allele associated proteins of *Petunia inflata*. Plant Physiol 96:61-68.

Singh A, Evensen KB, Kao TH. 1992. Ethylene synthesis and floral senescence following compatible and incompatible pollination in *Petunia inflata*. Plant Physiol 99(1):38-45.

Sink KC. 1984. Petunia. In: KC Sink (Ed), *Petunia*: *Monographs on Theoretical and Applied Genetics 9*. Springer-Verlag, Berlin. pp 3-9.

Spira TP, Snow AA, Whigham DF, Leak J. 1992. Flower visitation, pollen deposition, and pollen-tube competition in *Hibiscus Moscheutos* (Malvaceae). Amer J Bot 79:428-433.

Stehmann JR, Semir J. 1997. A new species and new combinations in *Calibrochoa* (Solanaceae). Novon 7:417-419.

Stehmann JR, Lorenz-Lemke AP, Freitas L, Semir J. 2009. The genus *Petunia*. In: T Gerats & J Strommer (eds), *Petunia*. New York, NY: Springer.

Swanson R, Edlund AF, Preuss D. 2004. Species specificity in pollen-pistil interactions. Ann Rev Genetics 38:739-818.

Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. Nature Rev 3:137-144.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum-likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol doi 10: 1093/molbev/msr123.

Twomey M. 2012. Characterization of *Ph*PRP1, an extracellular arabinogalactan protein from *Petunia hybrida* pistils. Master of Science thesis. Western Washington University. Bellingham, WA.

Velasquez SM, Ricardi MM, Dorosz JG, Fernandez PV, Nadra AD, Pol-Fachin L, Egelund J, Gille S, Harholt J, Ciancia M, Verli H, Pauly M, Bacic A, Olsen CE, Ulvskov P, Petersen BL, Somerville C, Lusem ND, Estevez JM. 2011. O-Glycosylated cell wall proteins are essential in root hair growth. Science 332:1401.

Venail J, Dell'Olivo A, Kuhlemeier C. 2010. Speciation genes in the genus *Petunia*. Philos Trans R Soc Lond B Biol Sci 365(1539):461-468.

Wallace AR. 1889. Darwinism: an exposition of the theory of natural selection, with some of its applications. Macmillan, London.

Wang H, Wu H-M, Cheung AY. 1993. Developmental and pollination regulation of the accumulation and glycosylation of a transmitting tissue-specific proline-rich glycoprotein. Plant Cell 5:1639-1650.

Wheeler DL, Church DM, Lash AE, Leipe DD, Madden TL, Pontius JU, Schuler GD, Schriml LM, Tatusova TA, Wagner L, Rapp BA. 2002. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 30(1):13-16.

Widmer A, Lexer C, Cozzolino S. 2009. Evolution of reproductive isolation in plants. Heredity 102:31-38.

Wijsman HJW. 1982. On the interrelationships of certain species of *Petunia*. I. Taxonomic notes on the parental species of *Petunia hybrida*. Acta Botanica Neerlandica 31:477-490.

Wu H-M, Wang H, Cheung AY. 1995. A pollen tube growth stimulatory glycoprotein is deglycosylated by pollen tubes and displays a glycosylation gradient in the flower. Cell 82:393-403.

Wu H-M, Wong E, Ogdahl J, Cheung AY. 2000. A pollen tube growth-promoting arabinogalactan protein from *Nicotiana alata* is similar to the tobacco TTS protein. Plant J 22(2):165-176.

Wu H-M, Zou JT, May B, Gu Q, Cheung AY. 1993. A tobacco gene family for flower cell wall proteins with a proline-rich domain and a cysteine-rich domain. Proc Natl Acad Sci 90:6829-6933.

Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. Trends Eco Evol 15:496-503.

Yuhki N, O'Brien SJ. 1990. DNA variation of the mammalian major histocompatibility complex reflects genomic diversity and population history. Proc Natl Acad Sci 87:836-840.

Yost JM, Kay KM. 2009. The evolution of post-pollination reproductive isolation in *Costus*. Sex Plant Reprod 22:247-255.

Zigler KS, McCartney MA, Levitan DR, Lessios HA. 2005. Sea urchin bindin divergence predicts gametic compatibility. Evol 59:2399-2404.