DETERMINING THE GENES FOR DISTYLY IN *TURNERA*

HASAN HAMAM

A THESIS SUBMITTED TO

THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

GRADUATE PROGRAM IN BIOLOGY

YORK UNIVERSITY

TORONTO, ONTARIO

July 2019

Abstract

Heterostyly promotes outbreeding and reduces pollen wastage. Distylous species have two mating types: L-morph and S-morph plants, which are genetically determined by the S-locus. It was hypothesized that the S-locus carries at least three tightly linked genes constituting a supergene. The S-locus genes responsible for distyly in Turnera had not been determined. Using deletion mapping and chromosome walking, the S- and s-haplotypes of the S-locus were constructed. Three hemizygous genes, TsSPH1, TsYUC6, and TsBAHD occur only on the S-haplotype and appear to determine S-morph characteristics. TsSPH1 is expressed in filaments, TsYUC6 in anthers, and TsBAHD in styles. A long-homostyle mutant (Drh) did not possess TsBAHD while another long-homostyle mutant (Mhomo-H) possessed a substitution mutation in TsBAHD. Short-homostyle mutants did not express TsSPH1. The incompatibility system of Turnera joelii possessed a mutation that appears to be linked to S-locus. A possible mechanism of incompatibility is discussed which involves brassinosteroids and auxins.

Acknowledgements

I am grateful to God and Dr. Joel Shore for taking me on as a graduate student and allowing me to work in his lab for the past several years. Without his continuous supervision, expertise, and encouragement, I could not have completed this work. I would also like to thank my graduate advisor, Dr. Kathi Hudak, whose advice and excellent insight helped me achieve this final product. I would also like to thank the collaborators on the project, Dr. McCubbin and his lab for letting me use some of his data and figures. Also, I would like to thank my lab-mates for entertaining me during lunch breaks and tolerating my numerous questions and queries.

I would like to express my greatest gratitude to my family and friends for always pushing me to my limits, supporting and believing in me. I owe a special thanks to my wonderful parents and siblings, for allowing me to spend a large portion of my life chasing my higher education. Without your unconditional love and support, I would not have had the ability to survive the long hours in the lab, stress, and the freedom to explore things that make me the happiest. To my brother and partner in crime, thank you for taking this journey with me and always supporting me especially when I faced problems in my experiments. To my little sister, your love and support always drove me to do my best in everything I aimed for, thank you for always being by my side.

Lastly, I extend my appreciation to the person who supported and stood beside me throughout the last two crazy years, Zeinab. Thank you for your endless love and support and for pushing me to work when I didn't feel like doing so. For all of your efforts and continuous love and support, words cannot thank you enough.

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Definitions of Terms

SI self-incompatibility

SC self-compatible

***S*-locus** a locus determining the self-incompatibility system or distyly

S-morph a plant with short styles and long stamens

L-morph a plant with long styles and short stamens

Haplotype a series of linked genes

***S*-haplotype** a series of linked genes containing the "dominant allele(s)" or hemizygous genes of the *S*-locus

***s*-haplotype** a series of linked genes containing the "recessive allele(s)" of the *S*-locus and/or the region homologous to the *S*-haplotype

Hemizygous a gene without an allelic counterpart (i.e. a gene that possesses a single copy only on just one chromosome)

Malpighiales one of the largest orders of flowering plants where it contains about 16000 species. It is divided into 32 to 42 families including Passifloraceae.

cM centimorgan, a unit used to measure the distance of genes

Å Angstrom, equivalent to 0.1 nanometer or 10^{-10} m.

kb kilobase pairs

bp base pairs

µL micro-liter

Contributions of Co-authors

Permission was granted by the New Phytologist Journal to use the following figures in this study: figures 3, 5, 7, 8, 17, 18, S2, S3, and S4.

Figure 2 was constructed by Shore, Labonne, Chafe, and McCubbin; Figure 5 Panel for *TsWRKY* was provided by J. Labonne; Figure 5 was constructed by Henning and McCubbin.

1. Introduction

The female and male sexual organs, the pistil and stamen respectively, usually display small and continuous variation in flowering plant populations due to various genetical and environmental factors (Barrett and Shore, 2008). However, populations of some species exhibit a different pattern of variation where hermaphroditic individuals, plants with both male and female reproductive organs, fall into two or three morphologically distinct mating groups or morphs (Barrett and Shore, 2008). These differences can include different stamen height, style length, and several other characteristics (Darwin, 1877; Barrett, 1992). Darwin focused on studying these polymorphisms where he devoted most of his book "*The Different Forms of Flowers on Plants of the Same Species"* to a particular form of polymorphism called heterostyly (Darwin, 1877; Barrett, 2019, in press). Heterostyly has received a lot of interest making it among the best-studied plant adaptations which are heavily featured in ecology and evolutionary biology books (Barrett, 2019).

Heterostyly is found in at least 28 families across the flowering plants (Ganders, 1979; Barrett and Shore, 2008). A defining feature of the heterostylous populations is reciprocal herkogamy, which is the reciprocal arrangement of the sexual organs in the floral morphs (Figure 1). According to Darwin (1877), reciprocal herkogamy increases the plant's fitness by reducing male gamete wastage on incompatible stigmas (Lloyd and Webb, 1992). Another main trait most heterostylous species possess is heteromorphic self-incompatibility (SI hereafter) or intra-morph incompatibility (Kappel *et al.,* 2017). SI is a physiological mechanism that prevents self-fertilization (Ueno *et al.,* 2016). In contrast to reciprocal herkogamy, SI is an effective way

of avoiding inbreeding depression and hence, promoting the female component of fitness (Charlesworth and Charlesworth, 1987; Barrett and Shore, 2008; Charlesworth and Willis, 2009). As a result of the SI system, successful pollination would only occur between stamens and pistils that have similar heights (Darwin, 1877; Barrett and Shore, 2008).

Despite over a century of study, many questions regarding the molecular aspects of heterostyly and its SI system remain unsolved.

1.1 A Brief overview of self-incompatibility (SI) systems in plants

Several different SI systems occur in plants, and they can be classified into either gametophytic or sporophytic SI, where the latter can be further broken down into homomorphic and heteromorphic SI (Stevens and Murray 1982).

1.1.1 Gametophytic SI

The defining characteristic of gametophytic SI, is that pollen captured by the stigma is rejected if there is a match between the haploid SI genotype of the pollen (the male gametophyte) and one of the SI alleles (or *S*-haplotypes), present in the diploid tissues of the style (Newbigin *et al.,* 1993; Matton *et al.,* 1994). If there is no match of alleles, the pollen tubes grow and achieve fertilization. There are two forms of gametophytic SI that have been best studied: The S-RNase system common to Solanaceae, Plantaginaceae, and Rosaceae (Newbigin *et al.,* 1993; Newbigin *et al.,* 2008); and the *S*-glycoprotein based mechanism that is found in *Papaver* (Franklin-Tong and Franklin 2003). The S-RNase system involves pistil-specific glycoproteins that are catalytically active RNases (McClure *et al.,* 1989; Newbigin *et al*., 1993). Since these RNases are thought to act as cytotoxins, their job is to degrade the RNA from the

pollen and pollen tubes derived from the same mating type (e.g. self-pollen) (Kao and Tsukamoto 2004; Goldraij *et al*., 2006). Lee *et al.* (1994) found that when the S-RNase was inhibited in *Petunia inflata*, the plant lost its ability to reject self-pollen. More recently than their style-specific counterparts (i.e. the S-RNases above), the pollen-specific component of the S-RNase system was discovered and found to be an F-box protein (Lai *et al*. 2002; Sijacic *et al*., 2004). These genes (i.e. those encoding the S-RNAse and the F-box proteins) have been shown to be tightly *S*-linked. The F-box protein is often involved in ubiquitinating proteins for degradation (McClure, 2010). In *Petunia*, F-Box proteins were shown to interact with S-RNases and activate their degradation (Kubo *et al*., 2010; Kubo *et al*., 2015). This system is thought to establish the specificity of pollen rejection by acting as S-RNase inhibitors through targeting S-RNase from cross pollen (i.e. non self-pollen) for ubiquitination and subsequent degradation, thus allowing successful growth of pollen and ultimately, fertilization (McClure *et al*., 2011; Chen *et al*., 2012). Further studies are still underway to fully reveal the mechanism of SI**.**

In the genus *Papaver* from the family Papaveraceae (the poppy family), a different gametophytic system occurs (Zhang *et al*., 2009). The proteins are not ribonucleases like those found in *Petunia* (Franklin-Tong *et al*., 1991). They are thought to act as signaling peptides encoded by a gene called *PrsS* (*Papaver rhoeas* stigma S) (Foote *et al*., 1994; Wheeler *et al*., 2009). The pollen-specific counterpart of the *PrsS* gene is known as *PrpS* (*P. rhoeas* pollen S) and it appears to be a novel protein. The PrpS protein is a transmembrane receptor that is associated with the plasma membrane of the pollen tube (Wheeler *et al*., 2009). When a PrsS signaling peptide interacts with a PrpS pollen-tube-specific transmembrane protein from self-pollen, a Ca^{2+} signaling cascade is activated that results in inhibiting the growth of the self-

pollen tube by activating programmed cell death (Zhang *et al*., 2009; Wheeler *et al*., 2009; Wu *et al*., 2011).

While it is evident that there are many aspects of the SI signaling pathway have been identified in poppy, the exact mechanism by which non-self-pollen is distinguished from self-pollen has yet to be determined (Wu *et al*., 2011).

1.1.2 Sporophytic SI

The rejection of pollen in sporophytic SI does not depend on the haploid genotype of the pollen, but on the diploid genotype of the pollen-producing plant, at the *S*-locus. There are two kinds of sporophytic SI: homomorphic and heteromorphic. The terms homo- and heteromorphic refer to the flower morphology of the plant species under study. For example, if the species is characterized by only a single morph (i.e. all flowers are phenotypically similar), sporophytic SI is said to be homomorphic (Stevens and Murray, 1982). In contrast, if the species is represented by more than one morph, as in heterostylous systems, SI is said to be heteromorphic (Stevens and Murray 1982).

1.1.3 Homomorphic sporophytic SI

Despite the differences in the molecular mechanisms (see below), the sporophytic and gametophytic SI systems are inherited in a similar way. Inheritance of SI is controlled by an *S*-locus, a single highly polymorphic locus, that is made up of at least two tightly linked genes which represent the style and stamen components of the self-recognition system (Gibbs, 1986; Nasrallah and Nasrallah, 1989; Newbigin *et al.*, 1993; Kao and Tsukamoto, 2004; Barret and Shore, 2008).

Homomorphic sporophytic SI has been best described in the *Brassicaceae* family (mustard) (Takayama and Isogai, 2005). This incompatibility system relies on the diploid genotype of the pollen-producing plant, and the diploid genotype of the female parent at the *S*-locus. An *S*-linked receptor Kinase (*SRK*) is expressed in the upper epidermal layers of the stigma which has been shown to contribute to the specificity of pollen rejection (Takasaki *et al*. 2000; Nasrallah and Nasrallah 2014a). A cysteine-rich protein known as SRC is expressed in pollen and localized to the pollen coat. It acts as a ligand for the style-specific SRKs (Schopfer *et al*., 1999; Takayama *et al*., 2001). An incompatibility pathway is activated when the pollen-specific SRC and the style-specific SRK share the same *S*-haplotype and bind. The binding activates a complex signaling cascade that leads to self-pollen inhibition (Nasrallah 2011; Takayama and Isogai, 2005). When the ligand and the receptor are of different origins (i.e. non-self), they cannot bind, and thus the signaling pathway is not activated. Even though many aspects of the mechanism have been established, full details of the mechanism have yet to be determined. The process by which SRC can access the membrane where SRKs are bound remains unknown. Furthermore, the precise mechanism by which the two molecules recognize each other and how the downstream signal transduction pathway leads to the inhibition also remains to be fully determined (Nasrallah 2011; Takayama and Isogai, 2005).

Several modifiers of the SI system have been discovered in the Brassicaceae such as the *S*-locus glycoprotein (*SLG*), M-locus protein kinase (MLPK), arm repeat-containing protein 1 (ARC 1) and Exo70A1 (Nasrallah and Nasrallah 2014a). SLG enhances the activity of SRK when self-pollen is present, whereas MLPK and ARC 1 are said to interact with SRK and Exo70A1, a component of the exocyst complex (Kakita *et al*., 2007; Samuel *et al*., 2009; Indriolo *et al*.,

2012). The exocyst complex, which is a target for ubiquitination by ARC 1, is responsible for delivering water and other essential factors for proper pollen hydration and germination on the surface of the stigma (Samuel *et al*., 2009). Thus, once self-pollen SRC is received and bound to its target, SRK along with MLPK activates ARC 1 where the latter ubiquitinates Exo70A1 and targets it for degradation (Nasrallah and Nasrallah, 2014a). Therefore, the pollen grain will not be able to hydrate and germinate due to the loss of the essential factors brought in by the exocyst complex.

1.1.4 Heteromorphic sporophytic SI

To date, heteromorphic or heterostylous SI has been reported in at least 28 angiosperm families (Barrett and Shore 2008). Previous studies pointed out that the heteromorphic and homomorphic systems, are unlikely to be related based on phylogenetic evidence and differences in the timing of the pollen rejection (Charlesworth, 1982; Gibbs, 1986; Lloyd and Webb, 1992). The genes determining SI in heteromorphic species have remained almost entirely unknown (Athanasiou and Shore, 1997), until very recently (see below). This might be due to the evolution of heterostyly having occurred several times independently raising the possibility that different genes are involved in these independent evolutionary lineages (Barrett and Shore, 2008). For that reason, the mechanisms that control the rejection of self-pollen are likely to be different between the heterostylous taxa (Dulberger, 1992), and even the flower morphs that belong to the same heterostylous species might also differ in their mechanism of action (Athanasiou and Shore, 1997; Safavian and Shore, 2010).

Recent molecular studies have allowed advances in our understanding of the genetics of heteromorphic SI. In distyly, genes that control SI are thought to be linked and inherited as if they are determined by a single diallelic locus that shows typical Mendelian patterns of inheritance (Barrett and Shore, 2008).

1.2 Distyly

For most SI systems that occur in plants, the floral morphology of the mating types does not differ. However, in heterostyly, the characteristics of the stamens and styles differ by length and organization between the mating types (Barrett and Shore, 2008). Such differences are seen in distylous species where they have two mating types that differ in the lengths of stamens and styles (Darwin, 1877). Short-styled individuals, also called as S-morph plants, produce flowers with short styles and long stamens (Figure 1a), whereas Long-styled plants, also called L-morph plants, produce flowers with long styles and short stamens (Figure 1b; Ganders, 1979; Barrett and Shore, 2008; Yasui *et al*., 2012; Kappel *et al*., 2017). The stamen filaments of the heterostylous species are often fused to the floral tube of the tubular flowers with different lengths observed between the two morphs (Ganders, 1979). The anthers in the L-morph plants, which are attached to the petal tube (at least in *Primula*), are positioned low compared to their high positioning (towards the opening of the tube) in S-morph plants. Thus, the difference in anther and stigma heights between the morphs is said to exhibit reciprocal herkogamy, which is the spatial separation of the sexual organs, with reverse orientations in the two morphs (Figure 1). Additional morphological features often distinguish the morphs, such as different lengths of stigmatic papillae (cells on the stigma surface that receive the pollen) and differences in pollen size and quantity (Dulberger, 1992). In addition, most

heterostylous species exhibit self-incompatibility (Wedderburn and Richards, 1990). Ganders (1979) found that distylous species are self- and intra-morph incompatible with successful pollinations most often achieved by crossing short and long morphs. With this in mind, distyly and SI will be first described in three genera: *Fagopyrum*, *Linum,* and *Primula* before moving on to discussing distylous species of *Turnera* (the subject of this study) in greater detail.

Figure 1. Photographs of S- and L-morph flowers of distylous *T. subulata.* (a) An S-morph plant with short styles and long stamens. **(b)** An L-morph plant with long styles and short stamens. The multi-branched stigmas are seen protruding through the lower portions of the filaments. Two petals were removed from each flower to allow a clear view of the reproductive organs.

1.2.1 Distyly in *Linum* (Flax Family)

Some species of *Linum* such as *Linum grandiflorum* present distylous characteristics (Dulberger, 1992; McDill *et al*., 2009). It has been suggested that the distyly syndrome in *Linum* is controlled by an *S*-locus supergene (Nicholls, 1985; Dulberger, 1992). As in most distylous species, S-morph plants are *S/s* heterozygous at the *S-locus*, whereas the L-morph plants are *s/s* homozygous recessive (Ushijima *et al*., 2012). The distylous species of *Linum* showed typical ancillary polymorphisms found in other distylous species as well as being SI (Dulberger, 1992).

Recent studies used RNA transcriptomic analysis of *L. grandiflorum* to identify an S-morph specific gene, the *THRUM STYLE SPECIFIC GENE 1, TSS1,* that is exclusively expressed in the styles of S-morph plants (Ushijima *et al*., 2012). Furthermore, several other genes with increased expression in the styles of the S-morph plants have been discovered. These include a gene encoding an aspartyl protease and an MYB transcription factor. The roles of these genes in distyly are still unknown.

1.2.2 Distyly in *Fagopyrum* (Buckwheat)

Fagopyrum esculentum (known as buckwheat) presents the typical characteristics of distyly (Darwin, 1877; Matsui and Nishio, 2004). Sharma and Boyes (1961) proposed that the *S*-locus is a supergene that controls both morphology and the SI system. In a model first proposed for *Primula* species (see below), they postulated that the *S* supergene consists of five genes: *G* controls style length, I^S controls stylar incompatibility, I^P controls pollen incompatibility, *P* controls pollen size, and *A* controls anther height. S-morph plants are heterozygous, *S/s*, whereas the L-morph plants are homozygous, *s/s* (Lewis and Jones, 1992).

Recently, next-generation sequencing allowed Yasui *et al*. (2012) to identify genes specific to the *S*-haplotype. As a result, four *S*-morph specific transcripts were discovered, *SSG1* to *SSG4* (Yasui *et al*., 2012). The authors found an S-morph plant with a branch that produced long-styled flowers that carried deletions of *SSG2* and *SSG3* genes, where the latter were not found in any of the L-morphs plants. This indicated a complete linkage between *SSG3* and the dominant *S*-haplotype. *SSG3* encodes a protein that is related to Early Flowering 3 protein (*ELF3*) from *Arabidopsis thaliana* (Yasui *et al*., 2012). Interestingly, two SC long-homostyle plants from *Fagopyrum* species carried inactivating mutations in *S*-locus *ELF3* which indicated that the gene controls the length of the style and possibly the female incompatibility behavior as well. Exploring a 610 kb region surrounding *ELF3* resulted in discovering *SSG2* and other functionally uncharacterized genes and many repetitive elements (Yasui *et al*., 2012). Further functional analysis of these genes and the surrounding region of *ELF3* may lead to the discovery of additional genes determining distyly in buckwheat.

1.2.3 Distyly and the supergene model in *Primula* (Primroses)

In *Primula*, distyly is also controlled by the *S*-locus with the dominant *S* and recessive *s* alleles (Bateson and Gregory, 1905). S-morph plants are *S/s* heterozygous while L-morph plants are homozygous for the recessive allele *s/s*. As a result of Alfred Ernst's classical genetic studies of *Primula* plants with uncommon traits (i.e. long- and short-homostyles), our understanding of the genetics of distyly advanced (Ernst, 1955; reviewed in Gilmartin, 2010). This breakthrough made *Primula* the most often used species in the study of distyly and the molecular mechanisms that underlie its expression (Barrett and Shore, 2008). Using genetic analysis and

inheritance data of novel morphs (i.e. long- and short-homostyles), the supergene model of inheritance was proposed for distyly (Ernst, 1955). It was hypothesized that three distinct but tightly linked genes were responsible for controlling different aspects of distyly in *Primula* (Ernst, 1955; Barrett and Shore, 2008). These three *S*-locus genes were referred to as *G*, *A*, and *P* (Ernst, 1955; Barrett and Shore, 2008; Cohen, 2010) where the *G* determines style length and female incompatibility, the *A* determines anther position where the dominant allele results in high anthers and the recessive counterpart leads to low positioned anthers, and the *P* controls pollen size and the male incompatibility. Thus, it was proposed that the *S*-locus in *Primula* acts as a supergene containing three very tightly linked genes which all together control an integrated phenotype (Schwander *et al*., 2014). The order of the three genes in *Primula* was thought to be *GPA* where the S-morph is a heterozygote and has a genotype of *GPA/gpa* while the L-morph is a recessive homozygote and has a genotype of *gpa/gpa* (Barrett and Shore, 2008). In addition to the three genes mentioned, other models suggested that the *S*-locus in *Primula* may contain four more loci based on the argument that a single locus was unlikely to control both the morphological and incompatibility phenotypes (Lewis and Jones, 1992; Kurian and Richards, 1997; Richards, 2003).

Recent studies of *Primula vulgaris* and *P. veris* have changed our understanding of heterostyly in *Primula* which was previously based upon classical genetic methods. First, several *S*-locus markers were discovered that helped in constructing a high-resolution genetic map of the *P. vulgaris S* locus region (Li *et al*., 2008; Cocker *et al*., 2015; Li *et al*., 2015). Further, using *S*-locus linked BACs as probes for fluorescence *in situ* hybridization, results showed that the *S*-locus is located next to the centromere of the largest chromosome in *Primula* (Li *et al*., 2015).

On the other hand, analyses of genomic DNA sequences and restriction-associated DNA (RAD) tag genotyping of a large pool of S- and L-morph plants resulted in identifying six markers which suggested that meiotic recombination was suppressed in a large region at or surrounding the *S*-locus (Nowak *et al*., 2015). Genome and transcriptome sequencing, as well as mutation studies of *P. vulgaris* and *P. veris*, allowed assembly of the sequence of the dominant and recessive *S*-locus haplotypes (Li *et al*., 2016; Burrows and McCubbin, 2017). The assemblies of the two haplotypes differed by the addition of a large region of ~280 kb that was specific for the S-morph plants (Li *et al*., 2016). This region contained five genes: a duplicate paralogue of a *GLO1* B-class homoeotic gene *(GLO2)*, *CYP734A50*, a gene encoding a Kelch-Repeat F-box protein (*KFB*), a gene for a Pumilio-like RNA-binding protein (*PUM*), and a gene encoding a protein with a highly conserved C-terminal domain (*CCM*) (Nowak *et al*., 2015; Huu *et al*., 2016).

Several conclusions can be made from the studies done on *P. vulgaris* and *P. veris*. The 280 kb region that is S-morph specific indicated that the *S*-locus is hemizygous rather than a diallelic supergene and that it contains at least five genes. Since the homologous copy of the *S*-locus is absent from the recessive haplotype (i.e. there are no recessive allelic counterparts), it provides an explanation for the suppressed recombination between the two “alleles” and rejects the hypothesis regarding the rise of long and short homostyles from crossing-over between the two haplotypes. Given that the region is hemizygous, there is no “pairing partner” on the recessive haplotype, and thus homologous recombination cannot occur in that region.

1.2.4 Distyly in *Turnera*

The genus *Turnera*, formerly a member of the plant family Turneraceae, is now classified as a member of the Passifloraceae (Shore *et al*., 2006). There are 143 species in this genus, mainly of Neotropical origin, with the majority being distylous and SI (Shore *et al*., 2006; Arbo, 2015). Interestingly, no other members of the Passifloraceae have been identified as distylous indicating that distyly has evolved at least once within the former Turneraceae (Shore *et al*., 2006).

Distyly in *Turnera* also appears to be determined by a single diallelic *S*-locus where S-morph plants are heterozygous *S*/*s* and L-morph plants are homozygous recessive *ss* (Figure 1; Labonne and Shore, 2011; Shore and Barrett, 1985). Like *Linum*, *Fagopyrum*, and *Primula*, the *S*-locus in *Turnera* is thought to be a supergene that controls many aspects of distyly, although classical genetic studies provided evidence for only two genes (Tamari *et al*., 2005; Barrett and Shore, 2008).

Various approaches had been used to study distyly, such as exploring differences in protein expression between the morphs (Athanasiou and Shore, 1997). Two proteins, a polygalacturonase and an α-dioxygenase, both specific to the S-morph, were identified as proteins that are differentially expressed between the styles of the S- and L-morph plants (Athanasiou *et al*. 2003; Khosravi *et al*. 2003, 2004; Tamari and Shore, 2004).

The style-specific polygalacturonase was absent from SC homostylous species suggesting that it might have a role in determining the SI in distylous species of *Turnera* (Khosravi *et al*. 2003; Tamari and Shore 2004). The precise role of polygalacturonase in distyly (if any) remains

unknown (Athanasiou *et al*. 2003; Khosravi *et al*., 2003; Tamari and Shore 2004, 2006; Shore *et al*., 2006; Barret and Shore, 2008).

High levels of α-dioxygenase expression were found in stylar tissues of S-morph plants of *Turnera* species (Khosravi *et al*., 2004). The gene encoding the α-dioxygenase is not linked to the *S*-locus while the polygalacturonase gene is linked to but was not found to reside at the *S*-locus. Both the α-dioxygenase and polygalacturonase are apparently upregulated by the dominant *S*-locus gene(s) in *Turnera* suggesting that they may be involved in downstream aspects of distyly (Athanasiou *et al*. 2003; Khosravi *et al*., 2004; Labonne *et al*., 2010). In a remarkable parallel, a polygalacturonase gene expressed in S-morph styles was recently discovered in buckwheat (Takeshima *et al*., 2019). The gene is not linked to the *S*-locus and its role in distyly is unknown, but it may be involved in SI.

To understand the molecular and genetic basis of distyly in *Turnera*, recent efforts have focused on determining the genes responsible for distyly using diverse methods including proteomics, chromosome walking, deletion mapping, and genomic sequencing. Several *S*-linked molecular markers had been identified and their location with respect to the *S*-locus was mapped (Labonne *et al*., 2009). As a result of the genetic mapping, one of the markers was identified as a retrotransposon, and the other two were a sterol sulfotransferase gene and N-acetyltransferase gene (Labonne *et al*., 2009). These markers provided valuable landmarks for subsequent chromosome walking and X-ray mutant characterization experiments in distylous *T. subulata* (Labonne *et al*., 2009; Labonne and Shore, 2011).

Using genetic material obtained from an S-morph (*Ss*) plant, a bacterial artificial chromosome (BAC) library was constructed and screened for the presence of *S*-locus linked genetic markers to help in isolating relevant BAC clones. The isolated clones were sequenced, and PCR amplicons of the ends of the BAC clones were used as probes in order to extend the chromosome walking (Labonne and Shore, 2011). This identified two BAC clones that carried the *s*-haplotype (Labonne and Shore, 2011). In addition, BAC clones containing parts of the *S*-haplotype were also identified (Labonne, 2011). Continued work has led to the discovery of additional S-haplotype BAC clones and genome scaffolds (Shore *et al*., 2019).

2. Objectives

While distyly has provided a useful model system in ecological genetics and pollination biology, a clear understanding of the genes involved and the molecular basis of SI have yet to be determined in *Turnera*. For this reason, genetic and molecular experiments were used to identify the genes that determine distyly in *Turnera*. Below, the three *S*-locus candidate genes (*TsSPH1, TsYUC6,* and *TsBAHD*) are shown to occur only in S-morph plants. The function of the three candidate genes is exploited in four mutants of *Turnera*.

With respect to plants possessing phenotypic mutations in their styles, it is expected that the presence and/or the expression of the style-specific candidate gene(s) is altered in the mutant plants. On the other hand, for plants possessing phenotypic mutations in their male reproductive organs, it is expected that the presence and/or the expression of the stamen-specific candidate gene(s) are altered in the mutant plants. These hypotheses will be tested using PCR amplification to study the presence/absence of the candidate genes in the mutant plants. In addition, RT-PCR will be used to study the expression of the genes in stamen- and style-specific tissue. If necessary, the genes will be sequenced in search of mutations.

3. Methods

Plants used in this study included a number of species in the genus *Turnera*. Many of the plants had been sampled decades ago and have been in cultivation in the greenhouse at York University. Various mutants have also been propagated via cuttings since they were first identified. Below, a description of the mutants used in this study is provided.

3.1 Description of Mutants

A series of S- and L-morphs as well as mutant long- and short-homostyle plants are shown in Figure 2. Figures 2a, 2d and 2g show normal L-morph, Figures 2c, 2f, and 2i show normal S-morph. Figures 2e and 2h show two long-homostyle mutant plants while figure 2b shows the short-homostyle mutant plant.

3.1.1 Long-homostyle mutants: Drh and Mhomo-H

A self-compatible long-homostyle mutant, Drh, is an autotetraploid that belongs to the *Turnera scabra* species sampled originally from the Dominican Republic (DR) (Figure 2h; Table 1; Tamari *et al*., 2005). The plant first appeared as a seedling growing out of a pot that contained an S-morph plant from the same population. Upon further investigation, it was observed to possess both long styles and long stamens with pollen similar in length to that of an S-morph plant. Thus, the mutant phenotype is the length of the style (Tamari *et al*., 2005).

Another long-homostyle mutant, Mhomo-H, is a product of a cross between two S-morph plants, *Turnera subulata* x *Turnera krapovickasii* (Figure 2e; Table 1; Tamari *et al*., 2005). From the original cross, 16 S- and 4 L-morph plants were obtained of which one S-morph plant produced a shoot bearing flowers with long styles and long stamens (Tamari *et al*., 2005). On

several occasions, cuttings that were taken from the mutant shoot were rooted. Shoots periodically revert to produce flowers with the normal S-morph phenotype (Mhomo-S hereafter), and these were also rooted. Both Mhomo-H and Mhomo-S have similar stamen lengths and pollen size, but the long-homostyle mutant has a long style and is self-compatible (Tamari *et al*., 2005).

Both long-homostyle mutants were used to determine the genes responsible for controlling the length of the style and possibly the female incompatibility system in *Turnera*.

3.1.2 Short-homostyle, TrinSH

A self-incompatible *Turnera subulata* short-homostyle mutant, TrinSH, possesses short styles and short stamens (Figure 2b; Table 1). Thus, the mutant phenotype is the length of the stamen. TrinSH was crossed with an L-morph plant from *Turnera subulata*, E1L, to produce progeny, (LISH hereafter), that can be further tested for mutations. TrinSH and its progeny were used to determine the genes responsible for controlling the length of the stamen.

3.1.3 Self-compatible S-morph mutant, SSC

An S-morph plant of *Turnera joelii* that was initially self-incompatible began producing shoots that were self-compatible (Table 1). This somatic mutant plant, referred to as the SSC mutant, is a genetic mosaic possessing some branches that have the S-morph phenotype and SI, while others are S-morph yet SC. Rooted cuttings of both the SC and the SI shoots have allowed propagation of the mutant and non-mutant forms of this plant.

Since the SSC mutant was only discovered recently, its mutation was characterized using crosses and self-pollinations. The location of the mutation with respect to the S-locus (at or linked to the *S*-locus) was studied using inheritance experiments. In addition, the role of the candidate gene/s, that might be responsible for controlling the incompatibility system, was studied using presence/absence experiments.

Figure 2. Photographs of *T. subulata* and *T. scabra* species with different floral morphology. (a and d) *T. subulata* S-morph; **(b)** *T. subulata* short-homostyle TrinSH; **(c and f)** *T. subulata* L-morph; **(e)** Long-homostyle progeny of Mhomo-H; **(g)** *T. scabra* S-morph; **(h)** *T. scabra* Drh long-homostyle; **(i)** *T. scabra* L-morph. Some of the figures are replicated in order to make a clearer comparison between homostyle mutant plants and the normal L- and S-morph plants.

Table 1. Summary of mutants used in this study.

Name	Mutant phenotype	Compatibility	Description
SSC	S-morph Self-compatible	Self-incompatible changed to Self-compatible	A spontaneous mutation occurred in an S-morph of *Turnera joelii* that was self-incompatible. Mostly upper branches of the plant are self-compatible. The mutation affected pollen so that it can fertilize both, S- and L-morphs instead of longs only.
Drh	Long Homostyle	Self-compatible	A spontaneous mutant of *Turnera scabra* with long styles and long stamens that is self-compatible. The pollen possesses the incompatibility phenotype of short pollen, while the style has the SI of L-morphs.
LH1 X-ray Mutant and other long-styled X-ray mutants	Long Homostyle	LH1 is self-compatible and all other X-ray mutants are self-incompatible	The plant is from *Turnera subulata* species. An X-ray mutation deleted several genes that changed the length of the style from short to long. There are 11 other X-ray mutants which all are L-morphs.
Mhomo-H	Long Homostyle	Self-compatible	A spontaneous somatic mutant long-homostyle from *Turnera subulata x Turnera krapovickasii*. The plant has the same incompatibility phenotype as Drh above. Periodically revertant shoots are produced resulting in nonmutant S-morph flowers and shoots, referred to as Mhomo-S.
Mhomo-S	S-morph revertant	Self-incompatible	An S-morph revertant of the long-homostyle mutant Mhomo-H from *Turnera subulata x Turnera krapovickasii*.
TrinSH	Short Homostyle	Self-incompatible	A short-homostyle of *Turnera subulata* with short styles and long stamens. The SI system is just like that of a normal S-morph. The reduced stamen length is the mutant phenotype.

3.2 DNA extraction

A rapid small-scale DNA extraction protocol, Mini-CTAB, outlined by Labonne *et al*. in 2009 and modified from Doyle and Doyle (1990), was used to isolate genomic DNA. An intermediate-sized flower bud (~ 5mm long) was removed from the tip of the branch and ground in 100 µl of CTAB isolation buffer [2% w/v CTAB (Sigma-Aldrich), 1.4 M NaCl, 0.2% v/v 2-mercaptoethanol, 20 mM EDTA and 100 mM Tris-HCl (pH 8.0)] in porcelain spot plates placed on ice. An additional 400 µL of CTAB buffer was added to each sample before being incubated in 1.5 mL microcentrifuge tube at 60°C for at least 30 min with occasional mixing. A volume of 500 µL of chloroform/isoamyl alcohol (24:1) was added to each sample, vortexed briefly and centrifuged for 5 minutes at 15,000 x *g*. After transferring the supernatant to new microcentrifuge tubes, nucleic acids were precipitated by adding 400 µL of cold isopropanol and centrifuging for 5 minutes at 15,000 x *g.* The supernatant was poured off and the pellets were washed using 500 µL of Doyle's wash buffer (76% Ethanol, 10 mM ammonium acetate) and centrifuged for 5 minutes at 15,000 x *g.* Finally, the wash buffer was removed, and the pellet was air-dried for at least 2 h before resuspending the DNA in 200 µL of TE buffer (10 mM Tris-HCl pH 8 and 1 mM EDTA). RNA was removed by incubation at 37°C in the presence of RNase A (10 µg/ml).

3.3 Primer Design

Primers were designed using Primer-BLAST, a free online primer designing tool from NCBI. They were designed to anneal at the beginning and the end of exon boundaries selected to be PCR amplified and/or sequenced. The selected primers were checked for potential problems using BLASTX from NCBI to ensure that they were not hitting other members of a particular

gene family (Zhang *et al*., 2000). Then, the primers were synthesized by Integrated DNA Technologies (IDT, Coralville, Iowa, USA). Initial PCR experiments were conducted to optimize and validate that the primers are amplifying the gene of interest and not amplifying other members of the same gene family. A compiled list of the primer pairs used for amplification, along with their names, sequences, and anticipated PCR product sizes (bp) are provided in Table S1 (See Appendix A).

3.4 Polymerase Chain Reaction (PCR)

This technique was used to study the presence/absence of genes of interest from the genome of different plants. All PCR reactions were conducted using an Eppendorf Mastercycler Thermal Cycler (Eppendorf Canada, Mississauga, Ontario). Each PCR reaction was performed using ~100 ng of DNA, 2 µL of forward and reverse primers (10 pmol/µL), 4 µL of ddH_2O, and 10 µL of Quick-Load® Taq 2X Master Mix [10mM Tris-HCL (pH 8.6), 50 mM KCl, 1.5 mM MgCl2, 50 units/mL *Taq* DNA Polymerase, 0.2 mM each dNTP, 5% glycerol, 0.08% IGEPAL CA-630, 0.05% Tween 20, 0.024% Orange G, 0.0025% Xylene Cyanol FF] (New England BioLabs, Ipswich, MA, USA) for a total of 20 µL reaction volume. PCR cycling parameters were: initial 5 min denaturation (94°C) followed by 35 cycles of denaturation (30 seconds at 94°C), annealing (30 seconds at 58°C), extension (30 seconds at 68°C) and followed by 5 min extension at 68°C. For amplicons greater than 1 kb, annealing temperature and extension times were increased as appropriate (60 sec for 1 kb). After completion, samples were held at 4°C, run on agarose gels (see below) or frozen at -20°C for later use.

3.5 Agarose Gel Electrophoresis

Agarose gels (0.5 - 1% agarose) were prepared with molecular biology grade agarose and 1X TAE buffer (0.04M Tris-Acetate, 0.001M EDTA). All gels were stained with ethidium bromide (0.5 µg/ml) and run at 85-100V. Either 100 bp or 1kb ladders were used (New England Biolabs, Massachusetts, USA). The gels were visualized with a UV light to confirm the migration of DNA and then photographed using an Alphaimager™ electronic ultra-violet photographic equipment and software present in the York University biology department.

3.6 Deletion Mapping

The deletion mapping helped in identifying the candidate genes for controlling distyly in *Turnera*. The generation and characterization of X-ray deletion mutants were reported by Labonne *et al*. (2010). A cross was made between a homozygous *SS* S-morph plant, F60SS, and a homozygous *ss* L-morph plant, S16L, both from diploid *T. subulata*. The pollen of F60SS, that was used to fertilize S16L, was irradiated with X-rays to generate mutations. All of the 3982 progeny grown were expected to be S-morph in the absence of mutations affecting the genes for determining distyly. Twelve progeny were mutants and showed the following phenotypes: 10 L-morph progeny (L1, L15, L16, L20-L26), one long-homostyle progeny (LH1), and one short-homostyle progeny (SH1) (Labonne *et al*., 2010). The incompatibility of the styles of SH1 is like that of an L-morph plant. The 12 mutants were used to localize the candidate genes determining the S-morph (see Figure 3).

The *S*-locus genes were PCR amplified in the X-ray mutants and parental plants (F60SS and S16L) using primers that targeted short portions of the genes (~250-400 bp). Polymorphisms

were detected using Single Stranded Conformation Polymorphisms (SSCP) (see below), polyacrylamide gel electrophoresis, and silver staining following the protocols of Labonne *et al*. (2009). For the three hemizygous S-morph-specific genes, agarose gel electrophoresis was used. The *TsWRKY* polymorphism figure was provided by JDJ Labonne where specific PCR primers were designed to amplify a polymorphic sequence ~560 bp upstream of the gene itself.

3.7 Single Strand Conformation Polymorphism (SSCP)

SSCP was used to identify which allele of the *S*-locus genes the X-ray mutant plants possessed. SSCP analysis was carried out using a Bio-Rad Mini Protein II Cell vertical gel electrophoresis unit (Bio-Rad Laboratories). PCR amplicons were denatured by adding 4 µL of PCR product to 20 µL of loading solution (95% formamide; 10 mM NaOH; 0.25% bromophenol blue and 0.25% xylene cyanol) and incubated at 95°C for 10 min in a Mastercycler PCR machine. Then, the samples were incubated on ice for 5 min. A small portion of each denatured PCR product (5 µL) was loaded on either a 7.5 or 10 % (37.5:1 acrylamide: bis-acrylamide) non-denaturing polyacrylamide gel (3 mL acrylamide 30%, 6.6 mL dH_2O, 2.4 mL 5X TBE buffer, 50 µL ammonium persulfate, 10 µL TEMED). The gel was placed in 1X Tris-borate EDTA buffer and subjected to electrophoresis for 1-2 hours at 200V at room temperature. The gel was silver-stained using 300 µL of silver nitrate solution (20% $AgNO_3$) in 30 mL of H_2O and incubated for 10 minutes (with gentle shaking every 2-3 min). The gel was rinsed 5 times with distilled water. A developer solution (8 grams NaOH, 500 mL ddH_2O, 2 mL Formaldehyde solution) was added until the gel was submersed (~ 30 mL) to allow the bands to develop. When sufficiently stained, 250 µL of 10% Acetic acid was added to stop the staining process.

3.8 RNA extraction and RT-PCR

Flower or flower bud tissue (100 mg) was frozen in liquid nitrogen, homogenized with a mortar and pestle, and total RNA was isolated using the RNeasy Plant Mini Kit (Qiagen, MA, USA) following the manufacturer's protocol coupled with "on column" DNA digestion. The RNA was quantified using a nanodrop 2000 spectrophotometer (Thermo Scientific, MA, USA).

The extracted RNA (100-200ng) was used to synthesize cDNA following the protocol of ProtoScript II First Strand cDNA Synthesis Kit (New England Biolabs, Massachusetts, USA). cDNA was used in RT-PCR to detect the expression of genes of interest using gene-specific primers (Table S1). PCR products were visualized on 0.8-1.5% agarose gels.

3.9 *TsBAHD* sequencing

Sanger sequencing was performed by the McGill University and Genome Quebec Innovation Centre (Montreal, QC) using Applied Biosystem's 3730xl DNA Analyzer technology. For each of the three samples, 40 µL (100-500 ng/µL) of the purified PCR product was sent for sequencing *TsBAHD* in the forward and reverse directions. 10 µL of each sequencing primers (5 pmol/ µL, Table S1) was also provided for each sample. Sequence chromatograms were visualized using Sequencher v. 5.4.6 (Gene Codes Corporation, Ann Arbor, MI, USA). The sequences were analyzed to trim poor-quality data from each sample. Sequences were inspected to verify the accuracy of base peak-calling, as well as to check for the presence of double peaks in the chromatograms.

The sequences were submitted to BLASTX searches to identify possible conserved domains using NCBI's Conserved Domain Database (Marchler-Bauer *et al*. 2015). The sequences were aligned using Clustal Omega (Sievers and Higgins, 2018) for comparison.

3.10 Purification of PCR products from agarose gel

Prior to purification, 5 µL of PCR amplicons were run on an agarose gel to ensure that the correct product had been amplified. For each sample, 60 – 80 µL of PCR amplicon (3-4 lanes with 20 µL each) were loaded on 0.8-1% agarose gels. After the bands separated, they were cut out of the gel using a clean razor blade and placed into a 1.5 mL microfuge tube. The DNA in the agarose was purified using the EZ-10 Spin Column PCR Products Purification Kit according to the manufacturer's instructions (BioBasic, Markham, Ontario).

3.11 Digestion with the Mfe1 Restriction Enzyme

The Mfe1 digestion experiment was used to determine whether a plant possesses the mutant or non-mutant form of *TsBAHD*. *TsBAHD* was PCR amplified from the genome of the samples being studied. For each sample, 60 – 80 µL of PCR product was purified using QIAquick PCR Purification Kit according to the manufacturer's instructions (QIAGEN, Hilden, Germany). The purified DNA was then digested using the Mfe1 restriction enzyme (Mfe1-HF, New England Bio Labs, Massachusetts, USA). Commonly, 6 µL (~600ng of DNA) of the purified DNA was added to 2.5 µL of Reaction Buffer, 15.8 µL of ddH_2O, and 0.7 µL of Mfe1 (7 units) restriction enzyme for a total volume of 25 µL. The volume of the restriction enzyme was adjusted based on the results with an additional amount of the enzyme added if digestion was incomplete.

Mfe1 cleaves at CAATTG restriction sites. Results, after digestion, were visualized on agarose gels (0.8-1 %). Images of the agarose gels were photographed as above.

3.12 Percent composition of the two *TsBAHD* alleles in long-homostyles

To quantify the extent of the two alleles (mutant and non-mutant) of the *TsBAHD* gene in individual plants, the intensity of each DNA band was measured from photographs of agarose gels using ImageJ software (Schneider *et al*., 2012). The intensities were measured by calculating the area of the intensity peaks produced by each band. Then the areas of the bands measured from each sample were added to calculate the percentage of each band with respect to other bands from the same sample. The percentages of each band were graphed.

3.13 Characterization of the self-compatible SSC mutant of *T. joelii*

A cross was made between a *Turnera joelii* self-incompatible (SI) S-morph plant and an SI L-morph plant. The S-morph parent had a *TsSPH1* gene inserted into its genome using pGreenII binary vector. The morph-ratio of the progeny produced deviated from the expected 1: 1 with 38 S-morph and 17 L-morph (chi-squared value= 6.33, P-value= 0.012). Note that the inserted *TsSPH1* gene was segregating in this family. Interestingly, one of the S-morph progeny that did not contain the *TsSPH1* gene insert, had a spontaneous mutation that caused its upper branches to become self-compatible (SC), while the lower branches remained SI. The mutant plant (referred to as SSC) was self-pollinated (selfing hereafter) and crossed on to normal L- and S-morph plants (see results). Using a pair of forceps, 3-5 anthers were brushed across the stigmas of the same flower (selfing) or another plant (cross), forceps were rinsed in ethanol and wiped dry prior to each pollination. Seeds were collected when fruits dehisced (~3 weeks after

pollination). Seeds were sown (~100 seeds per pot) in standard potting soil that had been autoclaved, in order to germinate them. Seedlings were transplanted into 15 cm pots ~ two weeks after they germinated. The morph of each progeny plant was recorded at flowering. From these first-generation crosses, four S-morph progeny were obtained (three from selfing and one from crossing SSC mutant on to a normal L-morph plant). These four self-compatible S-morph progeny were then selfed to produce a second-generation offspring.

To determine whether the second-generation progeny were SI or SC, each plant was selfed 10 times and seeds were counted when fruits ripened. The seeds were collected and germinated to produce the third generation of offspring. Plants of the third generation were tested for their compatibility by selfing each at least 5 times. Several L-morph progeny from each family/generation were selfed to test their incompatibility. Seeds set by the S-morph progeny of the third generation were stored in envelopes, and the mean number of seeds for each plant was recorded.

4. Results

4.1 *S*-locus assembly and identifying candidate genes for distyly

Recent studies led to the discovery of *S*-locus BACs comprising the *s*-haplotype and the *S*-haplotype (Labonne and Shore, 2011; Shore *et al*., 2019). A draft assembly of the *T. subulata* genome has also now been constructed, genome scaffolds at the *S*-locus have been identified, and the BAC clones and genome scaffolds have been assembled were overlaps in the sequence occurred (Shore *et al*., 2019).

The *s*-haplotype constituted of two BAC clones; BAC-J10 (155,226 bp) and BAC-L22 (192,366 bp) where the ends of the two clones overlapped by ~23 kb (Figure 3; Labonne and Shore, 2011; Shore *et al*., 2019). The total length of the assembled recessive haplotype is 324,544 bp and it contains 21 genes. For the *S*-haplotype, the ends of four BAC clones and a single scaffold overlapped to produce a fully assembled portion of the haplotype with a total length of 719,731 bp (Figure 3; Shore *et al*., 2019). This assembled portion of the *S*-haplotype contains nine genes compared to the 21 possessed by its recessive counterpart. Interestingly, the dominant haplotype possesses three additional genes, *TsSPH1*, *TsYUC6*, and *TsBAHD* (Figure 3 and Table 2). All three genes occur on BAC-A24 whereas BAC-G5i contains *TsSPH1*, and BAC-I7 possesses *TsBAHD* (Figure 3). Both BAC-G5i and BAC-I7 overlap with BAC-A24 (Figure 3). There are no recessive copies of the three additional genes indicating they are hemizygous. Finally, BAC-C7e possesses the *TsAP2* gene, however, the BAC clone does not overlap with any of the other BACs (Figure 3).

The remaining genes of the *S*-haplotype that correspond to the recessive counterpart occur on six scaffolds from the *T. subulata* genome assembly (Figure 3; Shore *et al*., 2019). These scaffolds were aligned using both deletion mapping and the gene order of the recessive haplotype (Figures 3, 4, and 5). The six genomic scaffolds possessed 14 genes that are complementary to their recessive counterparts from the *s*-haplotype. The six genome scaffolds have a total length of 786,998 bp which increases the total length of the *S*-haplotype to 1,484,000 bp which is ~4.6 times larger than the length of the *s*-haplotype.

In addition to the insertion of the three *S*-haplotype genes, it appears that there are at least two inversions in the dominant *S*-haplotype compared to the recessive. The first inversion encompasses three genes while the other includes 14 genes. The inversion of the 14 gene segment is supported by the occurrence of *TsSUMO, TsCYCD, TsADPF* and *TsEB1C* all on the same scaffold (Scf289) of the *S*-haplotype whereas the latter three are connected to *TsSKIP2* on the recessive *s*-haplotype (Figure 3). In addition, the genes *TsEB1C* to *TsDUF247* on the *S*-haplotype are all found in the genome of the X-ray deletion mutant L22 supporting the order of these genes on the dominant S-haplotype (Figure 5). However, since the scaffolds carrying those genes do not overlap, the position of the genes within this segment of the *S*-locus is yet to be determined with certainty. In contrast, the *s*-haplotype copy of *TsSUMO* is linked to *TsRBP* and not to *TsCYCD* as observed on the *S*-haplotype (Figure 3). For the second inversion, the order of *TsCBX, TsERF,* and *TsRBP* genes is inverted on the *S*-haplotype compared to the recessive (Figure 3).

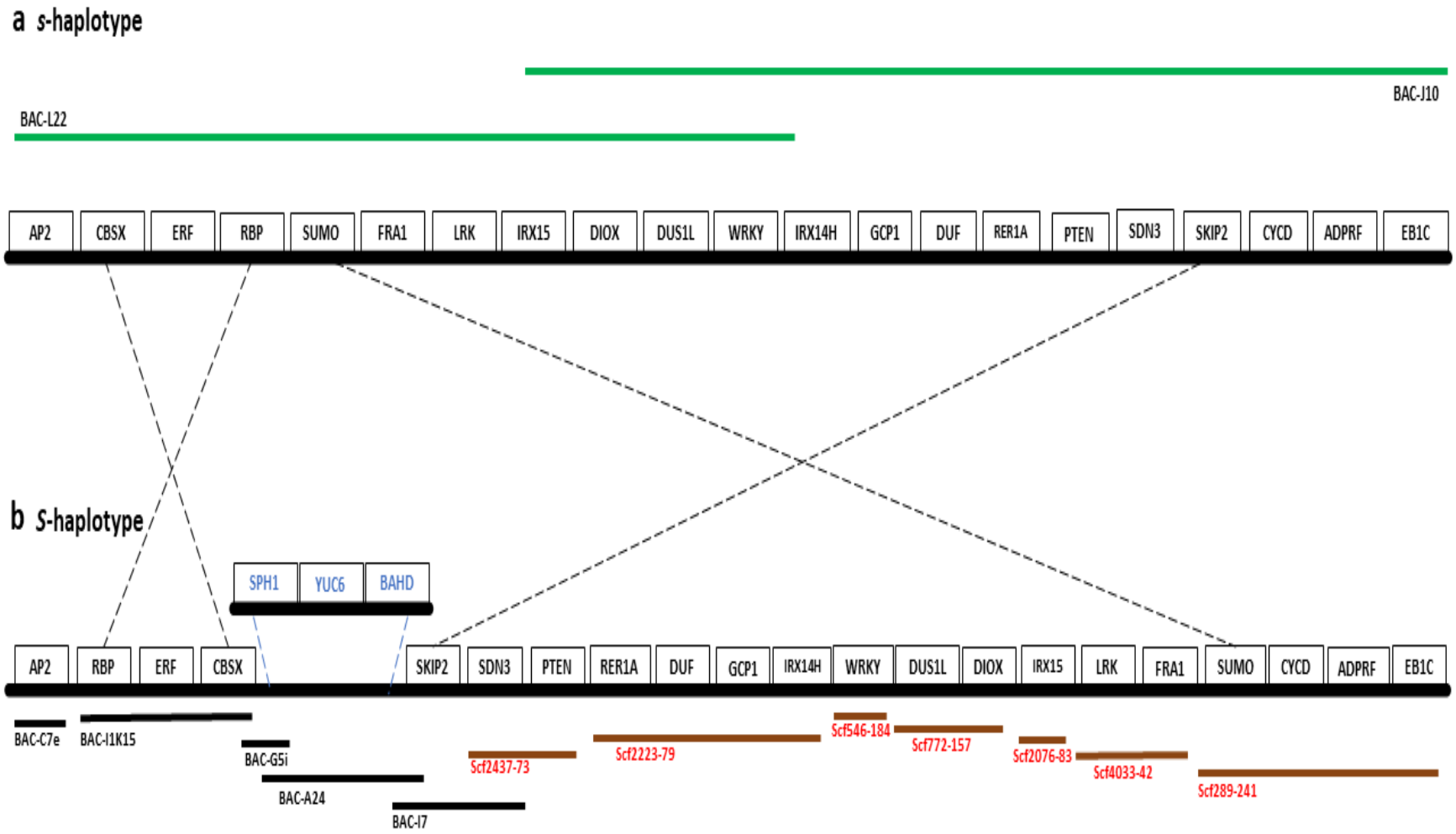

a s-haplotype
BAC-J10
BAC-L22
AP2
CBSX
ERF
RBP
SUMO
FRA1
LRK
IRX15
DIOX
DUS1L
WRKY
IRX14H
GCP1
DUF
RER1A
PTEN
SDN3
SKIP2
CYCD
ADPRF
EB1C
b S-haplotype
SPH1
YUC6
BAHD
AP2
RBP
ERF
CBSX
SKIP2
SDN3
PTEN
RER1A
DUF
GCP1
IRX14H
WRKY
DUS1L
DIOX
IRX15
LRK
FRA1
SUMO
CYCD
ADPRF
EB1C
BAC-C7e
BAC-I1K15
BAC-G5i
BAC-A24
BAC-I7
Scf2437-73
Scf2223-79
Scf546-184
Scf772-157
Scf2076-83
Scf4033-42
Scf289-241

Figure 3. Genetic Map of the S-locus haplotypes of *Turnera subulata*. Inverted gene orders are indicated by the dotted lines. **(a)** The *s*-haplotype is composed of two overlapping BAC clones (BAC-L22 and BAC-J10 in green). **(b)** The *S*-haplotype is composed of 5 BAC clones (black) and 7 genome scaffolds (brown). Three additional genes are inserted in the *S*-haplotype (blue). The BAC clones and scaffolds are represented by the solid lines above the *s*-haplotype and below the *S*-haplotype. The haplotypes, BAC clones, and scaffolds are not drawn to scale. All genes listed on both haplotypes are found in Table S2 (Appendix B). Note, the haplotypes were constructed and annotated by my co-authors, JDJ Labonne, JS Shore, PDJ Chafe, and A McCubbin (see Shore *et al.*, 2019).

4.2 Localizing *S-locus* genes using deletion mapping

Twelve X-ray mutants from *T. subulata* expressed different phenotypes than the expected S-morph as follows: ten L-morph (L1, L15, L16, L20-L26), one long-homostyle (LH1), and one short-homostyle (SH1) (Labonne *et al*., 2010). Fifteen genes from the *S*-haplotype were amplified using PCR on genomic DNA from the 12 X-ray mutants with gene-specific primers (Figures 4 and 5, Table S1). None of the three candidate genes amplified from the genomic DNA of the deletion mutants except from the long-homostyle LH1 where both *TsSPH1* and *TsYUC6* amplified but not *TsBAHD* (Figure 4).

SSCP was used on the PCR products of the remaining 11 genes to determine the presence/absence of the *S*-haplotype alleles from each deletion mutant and the breakpoints of X-ray mutants (Figure 5a and b). The approximate location of breakpoints for different mutants is illustrated in Figure 5b. The *S*-haplotype alleles of *TsSKIP2, TsSDN3,* and *TsPTEN* are absent from all mutants, but they do possess the recessive *s* copies of those genes. No markers were developed to determine if *TsRER1A* occurred in any of the mutants. The alignment of the X-ray mutants against the *S*-locus extends beyond the genes investigated in this study. Labonne *et al*. (2010) and Labonne and Shore (2011) used markers specific to regions up and downstream of the *S*-locus. Three markers used by Labonne and Shore (2011) are illustrated in this study; *O12SP1* (0.05 cM distal side of *S*-locus)*, H5SP1* (0.1 cM distal side of *S*-locus), and *N24SP1* (0.35 cM proximal side of *S*-locus) (Figure 5a and b).

Deletion mapping of the X-ray mutants provides insights into the function of the three hemizygous genes in distyly. All X-ray mutants are missing the *TsBAHD* gene including the X-ray deletion mutant, LH1, implicating *TsBAHD* as the gene responsible for short-style length and

possibly incompatibility of short styles (Figures 4 and 5). Only the long-homostyle mutant (LH1) possesses both the *TsSPH1* and *TsYUC6* genes implying that both genes determine the stamen characteristics of the S-morph (i.e. elongated filaments, large pollen, and possibly pollen incompatibility).

Table 2. List of candidate genes used in this study. The three candidate genes were identified in *Turnera subulata* using deletion mapping. All three genes occur on BAC-A24 clone. See Appendix C for the alignment of *TsSPH1*, *TsYUC6*, and *TsBAHD* against their homologs.

Name	Description	Size (bp)
TsSPH1	*SPH1* incompatibility protein homologue; one exon	469
TsYUC6	Flavin-dependent monooxygenase YUCCA family; four exons and three introns	11,131
TsBAHD	BAHD acyltransferase enzyme; one exon	1,317

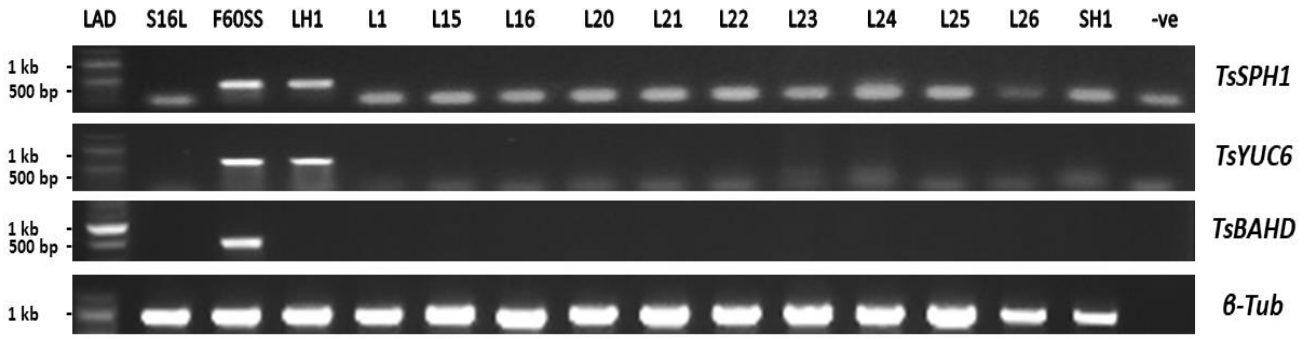


Figure 4. Screening for the three candidate genes, *TsSPH1, TsYUC6, and TsBAHD* in the genome of the X-ray deletion mutants and parental plants. The portions of the genes were PCR-amplified and visualized on a 1% agarose gel. The samples used were: *T. subulata (ss)* L-morph S16L parental plant; *T. subulata (SS)* homozygote S-morph F60SS parental plant; long-homostyle X-ray mutant LH1; L-morph X-ray mutants L1 - L26; short-homostyle X-ray mutant SH1. -ve is the negative control. *β-Tubulin* was used as a control. LAD: 100 bp ladder. The faint bands observed under the bright bands in the -ve control lane represent primer dimers as evidenced by the presence of a band in the –ve control lanes for both *TsPSH1* and *TsYUC6*.

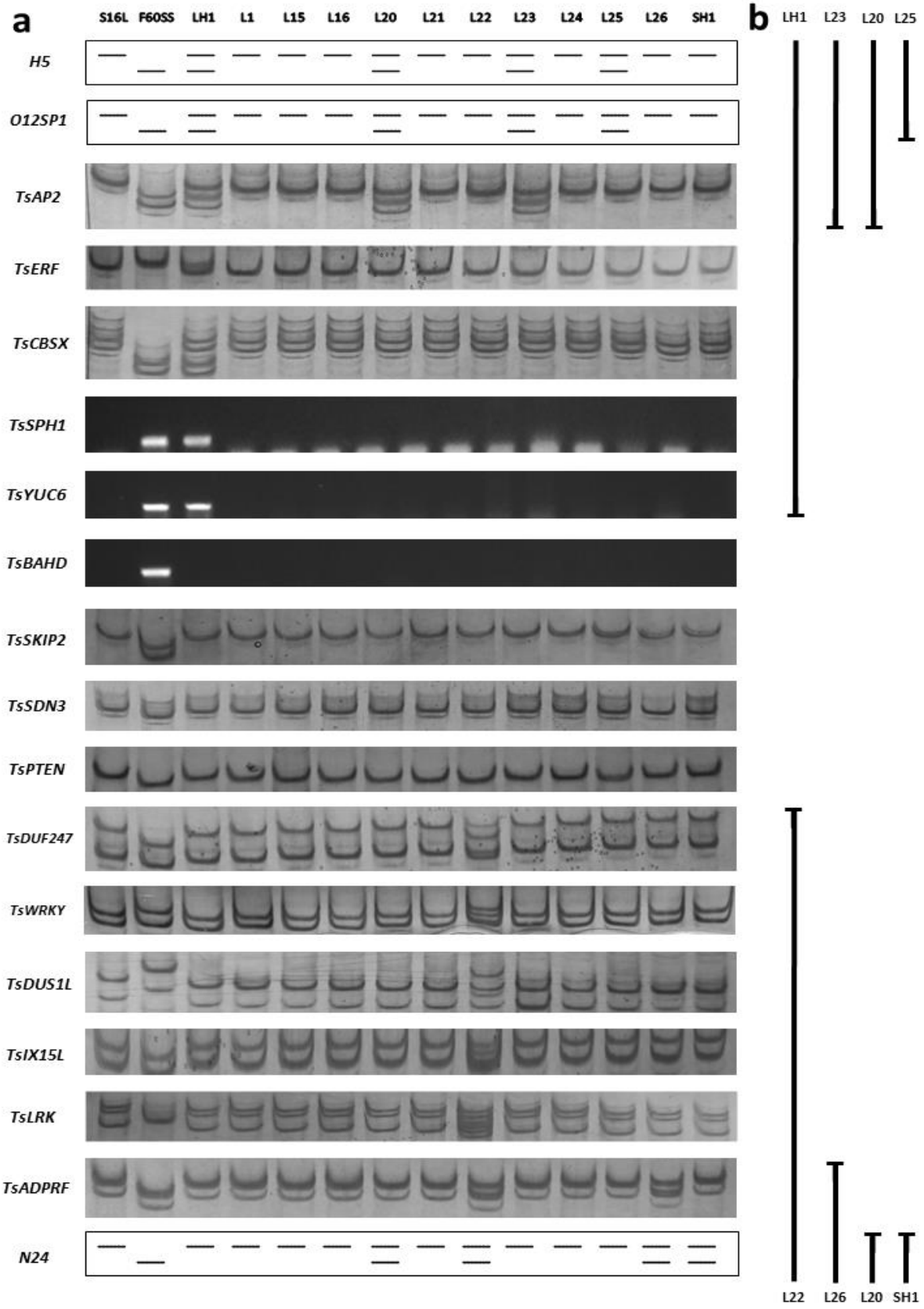

a
S16L
F60SS
LH1
L1
L15
L16
L20
L21
L22
L23
L24
L25
L26
SH1
H5
O12SP1
TsAP2
TsERF
TsCBSX
TsSPH1
TsYUC6
TsBAHD
TsSKIP2
TsSDN3
TsPTEN
TsDUF247
TsWRKY
TsDUS1L
TsIX15L
TsLRK
TsADPRF
N24
b
LH1
L23
L20
L25
L22
L26
L20
SH1

Figure 5. Deletion mapping using Single Strand Conformational Polymorphism (SSCP) and agarose gel electrophoresis of deletion mutants and parental plants. The three candidate genes, *TsSPH1, TsYUC6,* and *TsBAHD* were all visualized on 1% agarose gels. Samples used are *T. subulata* L-morph S16L; *T. subulata* homozygote S-morph F60SS; long-homostyle X-ray mutant LH1; L-morph X-ray mutants L1 - L26; short-homostyle X-ray mutant SH1. *H5*, *O12SP1,* and *N24* are schematic of *S*-locus markers used by Labonne *et al.* (2010). For SSCP, mutants possessing alleles of both parental plants, F60SS and S16L, should have additive banding patterns. Mutants with just the S16L banding patterns have a deletion of the *S*-haplotype allele of the gene. The faint bands observed under the bright bands in *TsSPH1* agarose gel represent primer-dimers (see Figure 4 above).

4.3 *TsSPH1*, *TsYUC6*, and *TsBAHD* genes in five *Turnera* species

PCR amplification was used to study the hemizygosity of the three candidate genes, *TsSPH1, TsYUC6,* and *TsBAHD* in the genome of S-morph and L-morph plants from five *Turnera species*. The three *S*-locus specific genes were amplified using gene-specific primers (Table S1) and their results were visualized on 1% agarose gels (Figure 6).

Five S- and five L-morph plants from *Turnera scabra, Turnera grandiflora, Turnera joelii, Turnera subulata* (an autotetraploid population), and *Turnera concinna* (3 L-morph and 7 S-morph plants) were used in this screening. The S-morph plants of all species studied showed amplification of the three genes while the genes did not amplify from all L-morph plants (Figure 6). These results provide further evidence for the hemizygosity of the three candidate genes as well as their specificity to S-morph plants.

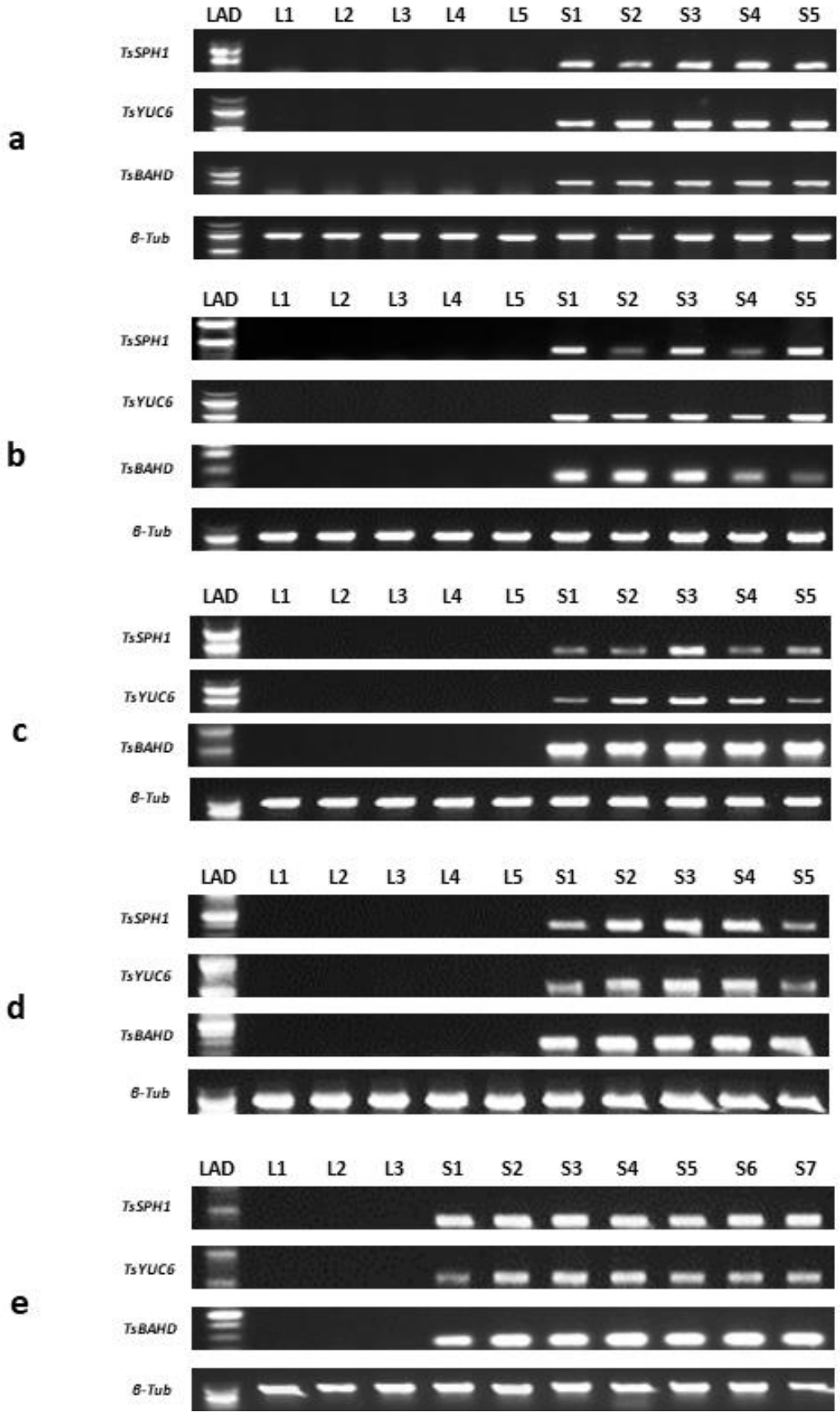

a
LAD L1 L2 L3 L4 L5 S1 S2 S3 S4 S5
TsSPH1
TsYUC6
TsBAHD
β-Tub
b
LAD L1 L2 L3 L4 L5 S1 S2 S3 S4 S5
TsSPH1
TsYUC6
TsBAHD
β-Tub
c
LAD L1 L2 L3 L4 L5 S1 S2 S3 S4 S5
TsSPH1
TsYUC6
TsBAHD
β-Tub
d
LAD L1 L2 L3 L4 L5 S1 S2 S3 S4 S5
TsSPH1
TsYUC6
TsBAHD
β-Tub
e
LAD L1 L2 L3 S1 S2 S3 S4 S5 S6 S7
TsSPH1
TsYUC6
TsBAHD
β-Tub

Figure 6. Amplification of the three candidate genes *TsSPH1*, *TsYUC6,* and *TsBAHD* from the genome of S- and L-morph plants from five *Turnera* species. The genes were visualized on 1% agarose gels. Lanes beginning with an (L) refer to L-morph plants while (S) refers to S-morph plants. (**a**) *Turnera scabra* from Managua Nicaragua; (**b**) *Turnera grandiflora*; (**c**) *Turnera joelii;* (**d**) *Turnera subulata* autotetraploid accession; (**e**) *Turnera concinna*. *β-Tubulin* (*β-Tub*) gene which was used as a control. LAD: 100 bp ladder (Upper band is 1kb and lower band is 500 bp).

4.4 Expression of the three candidate genes and protein sequence alignments

Semi-quantitative RT-PCR was used to determine the organ and time of expression of *TsSPH1*, *TsYUC6* and, *TsBAHD* in S- and L-morph plants of *T.subulata* (Figure 7). None of the three genes were expressed in any of the L-morph plants. *TsSPH1* and *TsYUC6* were both expressed in anthers, mainly those of smaller size/age, and the expression of *TsYUC6* was greater than that of *TsSPH1* (Figure 7a). *TsYUC6* showed two alternatively spliced transcripts of different sizes. Only *TsSPH1* was expressed in filaments with greater expression at the largest size/age buds (Figure 7b). Only *TsBAHD* was expressed in pistils (styles) with its expression peaking at intermediate size/age buds (Figure 7c).

Aligning the amino acid sequence of the stamen-specific gene *TsSPH1* against putative homologs from *A. thaliana* and *P. trichocarpa* showed that they all maintained the conserved regions in common with the *SPH1* gene (*S8*-allele) from *Papaver rhoeas* (Appendix C, Figure S2). In addition, they all possessed a putative signal peptide at the beginning of the sequences (Shore *et al*., 2019). Similarly, the amino acid sequence of *TsBAHD* gene was aligned against that of *A. thaliana* (AT3G26040) and *Populus trichocarpa* (XP_002314550) (Appendix C, Figure S3). The sequences maintained the functionally essential conserved regions of BAHD acyltransferases (Bayer *et al*., 2004; Tuominen *et al*., 2011; Appendix C, Figure S3). Finally, the amino acid sequence of *TsYUC6* was aligned against *YUC6* from *A. thaliana* (AT5G25620) and showed 63% identity (Appendix C, Figure S4).

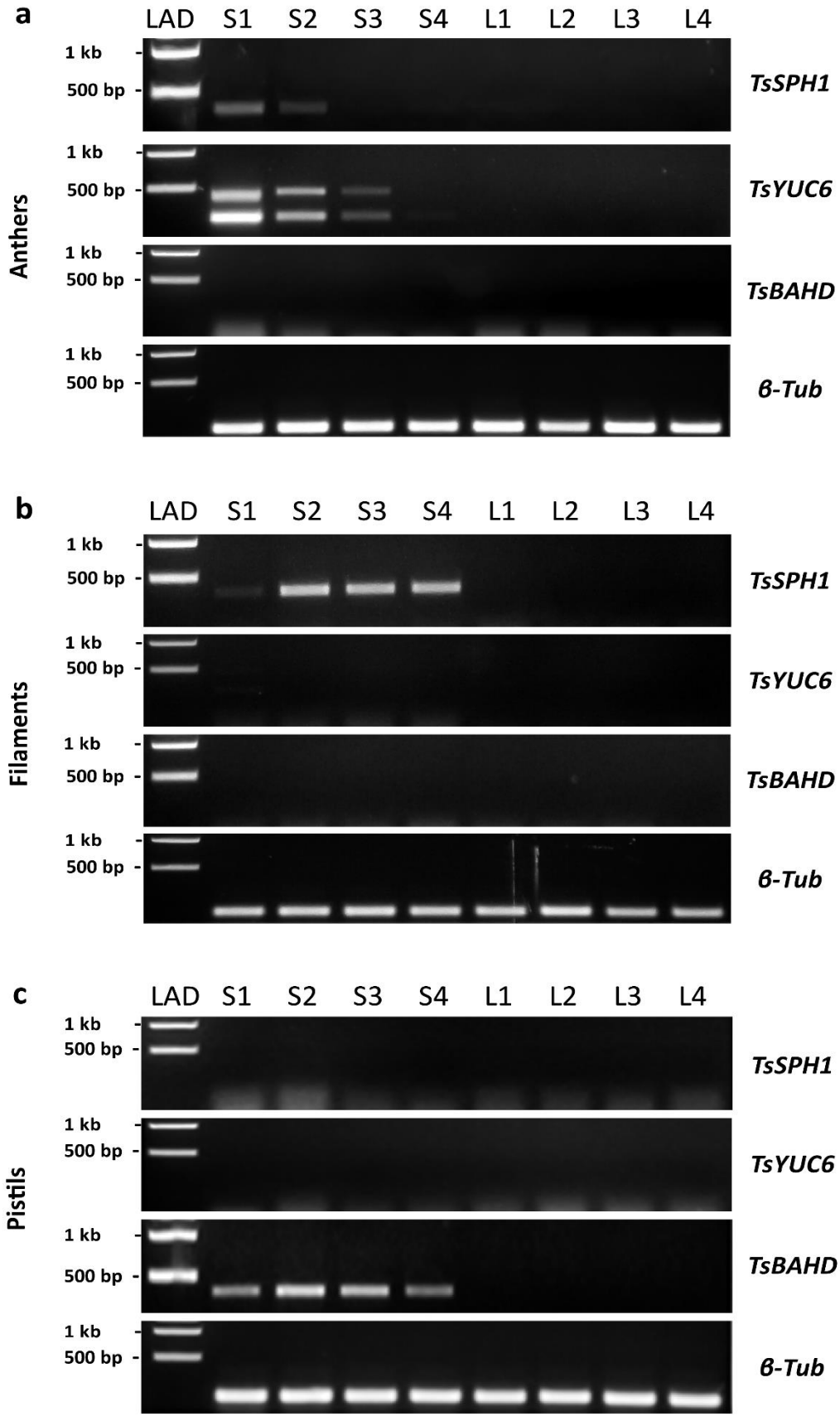

a
LAD S1 S2 S3 S4 L1 L2 L3 L4
1 kb
500 bp
TsSPH1
TsYUC6
TsBAHD
β-Tub
Anthers
b
Filaments
c
Pistils

Figure 7. Expression of the three candidate genes *TsSPH1*, *TsYUC6* and *TsBAHD* in different floral organs of S- and L-morph plants of *T. subulata* using semi-quantitative RT-PCR. Lanes S1-S4 refer to S-morph flower buds while L1-L4 refer to L-morph flower buds. RNA was extracted for (a) anthers, (b) filaments and (c) pistils from floral buds of increasing age (measured by size): (1): 6-8 mm; (2): 10-12 mm; (3): 14-16; and (4): >18 mm). *β-Tubulin* (*β-Tub*) is expressed in both morphs and all buds. LAD: 1 kb ladder. Note that the RT-PCR was conducted by my co-authors P. Henning and A. McCubbin (see Shore *et al*. ,2019). The faint bands observed in *TsSPH1* and *TsYUC6* (Pistil) and *TsBAHD* (Anthers) represent primer dimers.

4.5 Testing the function of candidate genes in distyly using mutants

The deletion mapping and expression studies implied that *TsBAHD* is the candidate gene controlling the style length and possibly the style incompatibility system of S-morph plants (Figures 5 and 7). Thus, the presence and expression of *TsBAHD* were studied in two long-homostyle mutants: Drh and Mhomo-H, where both mutants are self-compatible and possess long styles and long stamens.

4.5.1 Long-homostyle mutant Drh

The self-compatible long-homostyle mutant Drh of autotetraploid *Turnera scabra* from the Dominican Republic possesses both long styles and long stamens (Figure 2h; Tamari *et al.*, 2005). Thus, there is a mutation(s) affecting its style length and incompatibility. To determine the gene(s) responsible for this mutant phenotype, the three hemizygous candidate genes were PCR amplified in four S- and L-morph plants of *T. scabra*, along with the long-homostyle mutant Drh. As expected, the four S-morph plants possessed the three genes in their genomes while the L-morph plants did not possess any of them (Figure 8). While the long-homostyle mutant, Drh, possessed both stamen-specific genes, *TsSPH1* and *TsYUC6*, in its genome, the style-specific gene *TsBAHD*, was absent (Figure 8). Similar results were reported for the X-ray long-homostyle mutant LH1 where *TsBAHD* was absent from its genome (Figure 4). These results suggested that the absence of *TsBAHD* led to the mutant long style phenotype in Drh (and LH1) and to self-compatibility. Therefore, *TsBAHD* appears to be responsible for short-style length and possibly the style incompatibility system of S-morph plants.

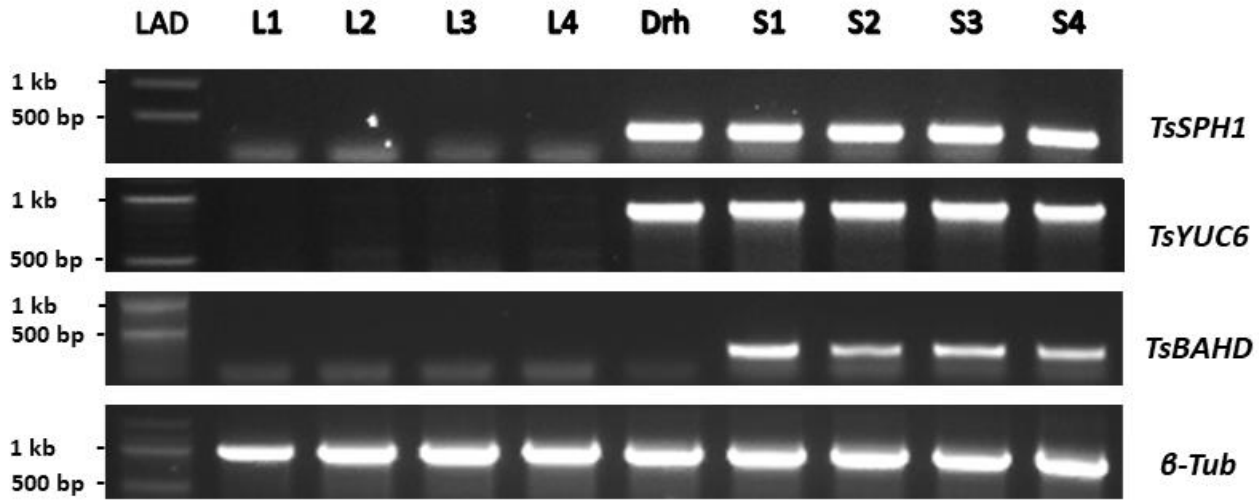


Figure 8. Screening for the three candidate genes, *TsSPH1*, *TsYUC6,* and *TsBAHD* from genomic DNA of *T. scabra* control plants and a long-homostyle (Drh) mutant. Each of the hemizygous genes was PCR amplified using genomic DNA from S- and L-morph control plants and the mutant Drh. Results were visualized on a 1% agarose gel. DNA samples used: L1–L4 L-morph plants; long-homostyle Drh; S1–S4 S-morph plants. *β-Tubulin* (*β-Tub*) was used as a positive control. LAD: 100 bp ladder. The faint bands observed in the L-morphs of *TsSPH1*, *TsBAHD*, and under the bright bands of the S-morphs represent primer dimers.

4.5.2 Long-homostyle Mhomo-H

4.5.2.1 The three candidate genes in Mhomo-H

The long-homostyle mutant, Mhomo-H, is a self-compatible plant that produces flowers with long styles and stamens (Tamari *et al*., 2005). On several occasions, branches of Mhomo-H produce revertant flowers that have the normal S-morph phenotype, referred to as Mhomo-S hereafter (see objectives). Mhomo-H, its long-homostyle progeny (Figure 2e), and Mhomo-S were screened for the three candidate genes using PCR as above. The stamen-specific genes *TsSPH1* and *TsYUC6* amplified in all samples. Interestingly, the style-specific gene *TsBAHD* also amplified from all samples including the long homostyle mutants (Figure 9).

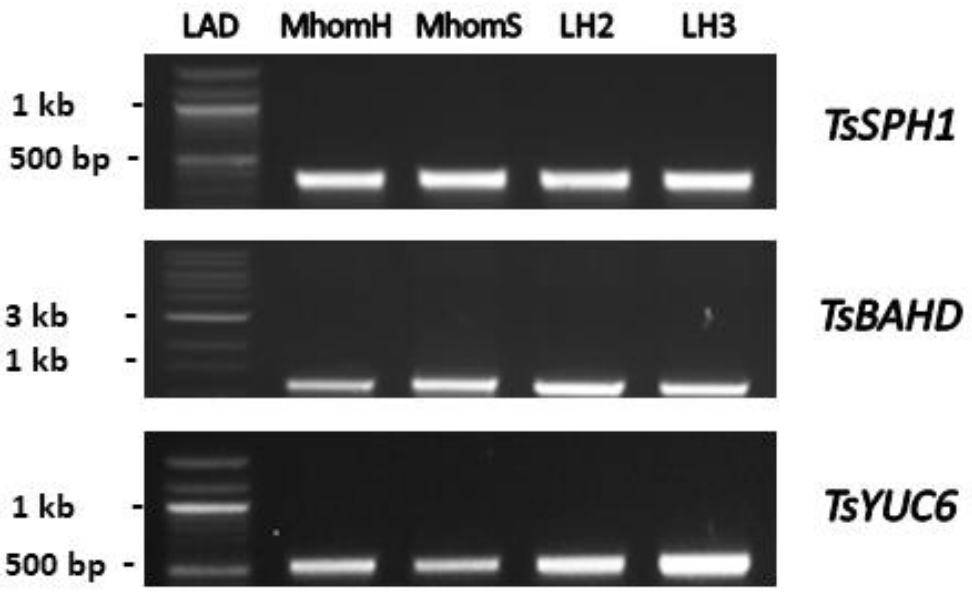


Figure 9. Screening for the three candidate genes in long-homostyle mutant Mhomo-H and its progeny. PCR was used to amplify three candidate genes from genomic DNA using specific primers (Table S1) and results were visualized on a 1% agarose gel. Samples used: (MhomH) long-homostyle mutant Mhomo-H; (MhomS) S-morph revertant Mhomo-S; (LH2) Mhomo-H long-homostyle progeny 734H; and (LH3) Mhomo-H homozygote long-homostyle progeny HH12. LAD: 100 bp ladder for *TsSPH1* and *TsYUC6*; 1kb ladder for *TsBAHD*.

4.5.2.2 Expression of *TsBAHD* in long-homostyle mutants

To test the hypothesis that *TsBAHD* is responsible for controlling the style length, the expression of *TsBAHD* was studied in Mhomo-H long-homostyle mutant and its progeny. RT-PCR was used to study the expression of *TsBAHD. TsBAHD* was expressed in two long-homostyle progeny and one S-morph progeny, but not in an L-morph plant, as expected (Figure 10). An additional three long-homostyle and four S-morph progeny of Mhomo-H also showed expression of *TsBAHD* (data not shown).

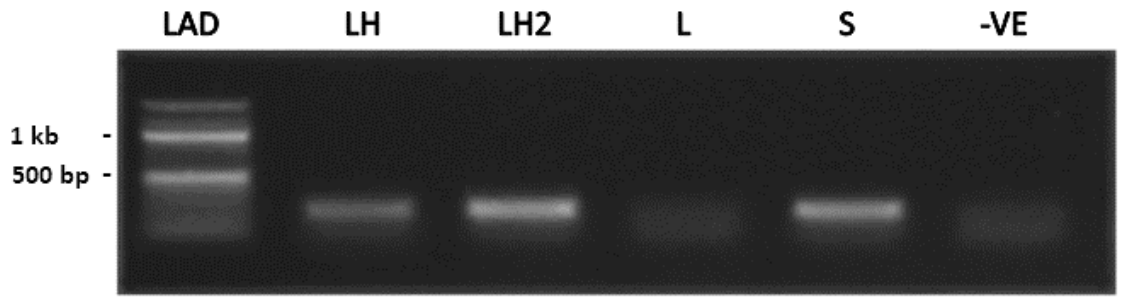


Figure 10. Expression of *TsBAHD* in style-tissue of long-homostyle and S-morph progeny of Mhomo-H. The expression of BAHD was studied using RT-PCR on RNA extracted from stylar tissue of flower buds. RT-PCR was used to amplify *TsBAHD* using gene-specific primers (Table S1) with a product size of 228 bp. Results were visualized on a 0.8% agarose gel. Samples used: Mhomo-H long-homostyle progeny: (LH) 275H, and (LH2) 734H; (L) *T. subulata* L-morph E1L; and (S) Mhomo-H S-morph progeny 918S. 100 bp ladder was used in Lane 1. The faint bands observed in the L-morph and -VE lanes represent primer dimers.

4.5.2.3 *TsBAHD* sequence and substitution mutation in long-homostyles

To determine whether a mutation occurred in *TsBAHD*, the gene was sequenced from the genome of two long-homostyle progeny and the S-morph revertant, Mhomo-S (see Appendix, Figure S1). The amino acid sequences of the gene were generated from the resulting DNA sequences (Figure 11). The resulting protein sequences were aligned and compared to BAHD enzymes from other members of the BAHD gene family, such as *AtBAHD* from *Arabidopsis* (Figure 11; Appendix C, Figure S3), to identify conserved regions (Tuominen *et al*., 2011). All four sequences possessed regions that are conserved among all members of the BAHD family which are essential for interacting with the substrates: the HxxxD, GN, and DFGWGKF motifs located at 160, 289, and 376 amino acids respectively (Figure 11; Unno *et al*., 2007; Tuominen *et al*., 2011). However, the GN motif in the two long-homostyle progeny possessed a substitution mutation at amino acid residue position 290 (Figure 11). The mutation changed an AAT codon to AGT resulting in an amino acid substitution from asparagine (N) to serine (S). The amino acid sequences were used to generate a 3D structure of the BAHD protein from Mhomo-S and a long-homostyle mutant progeny using Raptor X and Pymol Software (Appendix D; Peng and Xu, 2011; The Pymol Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC). Based on the 3D model of *TsBAHD*, there is no evidence for any differences in the tertiary protein structures of both Mhomo-S and the long-homostyle mutant except for the amino acid substitution of S with N in the GN motif (region 2 in the figure) (Figures S5 and S6, Appendix D).

The mutation also eliminated a Mfe1 restriction site in the gene. Thus, the mutant copy of *TsBAHD* has lost the restriction site. Therefore, it is possible to use restriction digests with Mfe1 following PCR amplification, to determine whether a plant possesses the mutant or non-mutant form of *TsBAHD*.

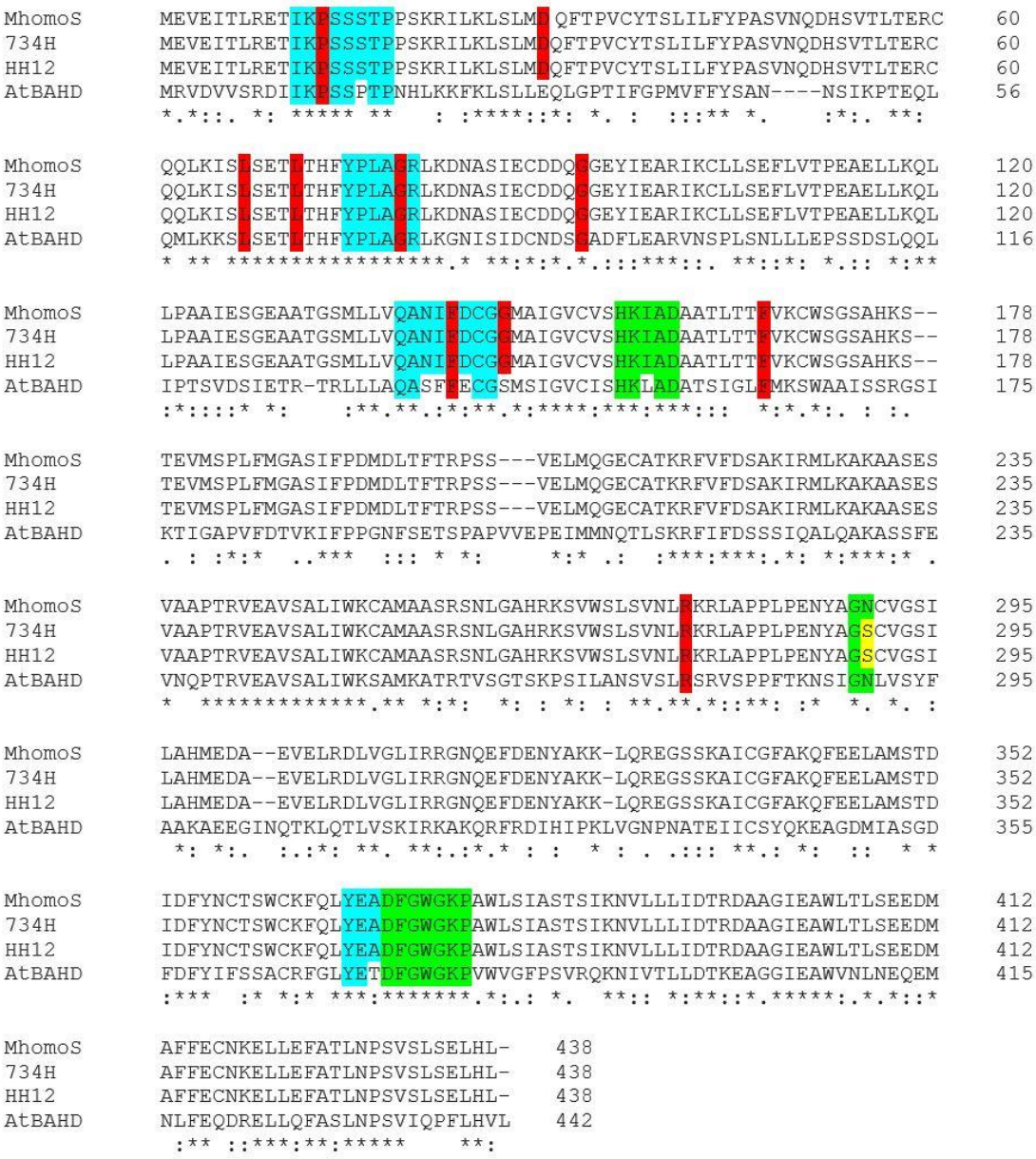


Figure 11. Alignment of *TsBAHD* amino acid sequences from Mhomo-H long-homostyle progeny, the S-morph revertant Mhomo-S, and *AtBAHD* from *Arabidopsis*. *AtBAHD* belongs to the BAHD family from *Arabidopsis* and the sequence was acquired from "The Arabidopsis Information Resource" (TAIR) (Lamesch *et al*., 2011). Regions highlighted in green indicate three conserved regions important for *TsBAHD* and *AtBAHD* function: HxxxD, GN, and DFGWG motifs. Regions highlighted in red indicate conserved amino acids among most *TsBAHD* genes from different BAHD families. Regions highlighted in blue represent conserved regions among the alcohol acyltransferase like enzymes (ACTLs). The amino acids highlighted in yellow indicate a mutation that substituted asparagine (N) with serine (S) in both 734H and HH12 long-homostyle mutant progeny, but not in Mhomo-S or *AtBAHD*.

Table 3. Genotypes of Mhomo-H, Mhomo-S, and progeny. S_{sub}*: S*-allele from *T.subulata*. S^{H}_{krap}: the homostyle allele from *T. krapovickasii*. $S^{revertant}_{krap}$: *S*-allele from *T. krapovickasii*. s_{sub}: recessive *s*-allele from *T.subulata*. Genotypes were inferred from crosses made by Tamari *et al*. (2005).

Name	Mutant Phenotype	Genotype	Predictions
Mhomo-H	Long-homostyle	$s_{sub}\ S^{H}_{krap}$	Mutant copy of *TsBAHD* from *T. krapovickasii*
Mhomo-S	S-morph	s_{sub} $S^{revertant}_{krap}$	Non-mutant copy of *TsBAHD* from *T. krapovickasii*
MhBRY-9S	S-morph	$S_{sub}\ S^{H}_{krap}$	Heterozygote where each nucleus contains both the non-mutant copy of *TsBAHD* from *T. subulata* and the mutant copy of *TsBAHD* from *T. krapovickasii*
275H, 734H, 859H	Long-homostyle	$s_{sub}\ S^{H}_{krap}$	Mutant copy of *TsBAHD* from *T. krapovickasii*
HH12	Long-homostyle	$S^{H}_{krap}\ S^{H}_{krap}$	Two mutant copies of *TsBAHD* from *T. krapovickasii*
723S, 886S	S-morph	$S_{sub}\ s_{sub}$	Non-mutant copy of *TsBAHD* from *T. subulata*

4.5.2.4 Detecting the two forms of *TsBAHD* (mutant and non-mutant) in long-homostyle progeny

Since the substitution mutation in *TsBAHD* eliminates the Mfe1 restriction site, it is possible to use restriction digests to determine which form of *TsBAHD* a particular plant possesses. Mfe1 digests of PCR amplicons of mutant *TsBAHD* will show only a single band of 1210 bp (due to the absence of the restriction site) while non-mutant *TsBAHD* will be digested into two fragments of 343 and 867 bp. *TsBAHD* was digested from four long-homostyle and two S-morph progeny of Mhomo-H (Figure 12). All long-homostyle progeny possessed only the undigested mutant form of *TsBAHD* indicated by a single band at 1210 bp. In contrast, all S-morph progeny possessed only the non-mutant copy of the gene indicated by the two bands (Figure 12). This is consistent with the mutant form of *TsBAHD* causing the long-homostyle phenotype.

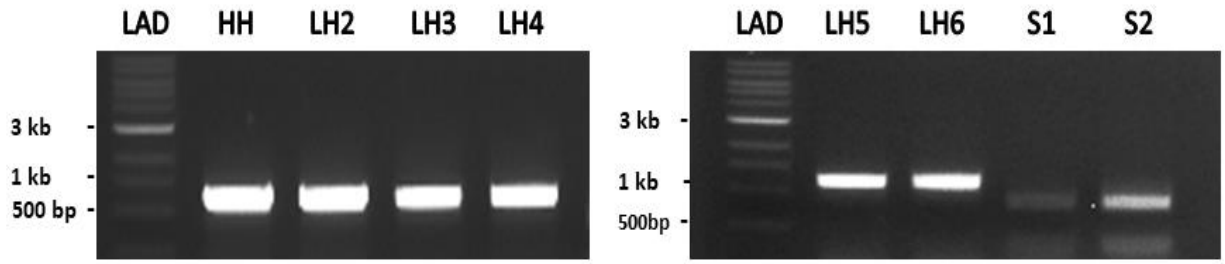


Figure 12. Digestion of PCR amplicons of the *TsBAHD* gene in long-homostyle and S-morph progeny of Mhomo-H with the Mfe1 restriction enzyme. PCR was used to amplify *TsBAHD* which was then digested with the Mfe1 restriction enzyme. The results were visualized on 0.8% agarose gels. Undigested copies (i.e. the mutant form) of *TsBAHD* are represented by a single band at 1210 bp while the non-mutant copies are represented by DNA bands at 867 and 343 bp. DNA samples used: (HH) homozygote long-homostyle S^HS^H progeny HH12; long-homostyle progeny: (LH2) 734H, (LH3 and LH6) 275H, and (LH4 and LH5) 859H; S-morph progeny (S1) 723S and (S2) 886S. LAD: 1 kb ladder.

4.5.2.5 Is Mhomo-H a periclinal chimera?

Mhomo-H frequently produces flowers possessing the non-mutant S-morph phenotype (known as Mhomo-S). Therefore, this suggests that Mhomo-H is a periclinal chimera. That is, different cells at the shoot apical meristem possess different nuclei some of which have the mutant form of *TsBAHD* and others the non-mutant form. In a periclinal chimera, different apical meristem layers contain either mutant or non-mutant copies of a gene. Depending on the patterns of cell divisions at these layers, different lateral shoots or organs may end up with varying numbers of cells with mutant vs non-mutant copies of *TsBAHD*. This can result in mutant long-homostyle flowers or revertant S-morph flowers. If this is true, we would expect to be able to find both mutant and non-mutant copies of *TsBAHD* in Mhomo-H.

TsBAHD was PCR amplified and digested with Mfe1 restriction enzyme. The results were compared to a known heterozygote S-morph plant, MhBRY-9S (Table 3), that carries both the normal allele S_{sub} and the homostyle mutant allele S^{H}_{krap} within each nucleus (Tamari *et al*., 2005; Figure 13, Lane 1). Thus, MhBRY-9S is expected to carry both the mutant and non-mutant forms of *TsBAHD* and using PCR, we might expect both copies to amplify equivalently. MhBRY-9S indeed possessed both copies of *TsBAHD* indicated by the single band at 1210 bp for the mutant allele, and the two additional bands at 867 and 343 bp for the normal allele (Figure 13). Similarly, the long-homostyle Mhomo-H mutant and its S-morph revertant Mhomo-S (Lanes 2, 5 and 6) possessed both copies of *TsBAHD* (Figure 13). As expected, the long-homostyle progeny of Mhomo-H possessed only the mutant copy of *TsBAHD* (Figure 13, Lanes 3 and 4). Therefore, the results support the claim that Mhomo-H and Mhomo-S are periclinal chimeras as they possess both copies of *TsBAHD*.

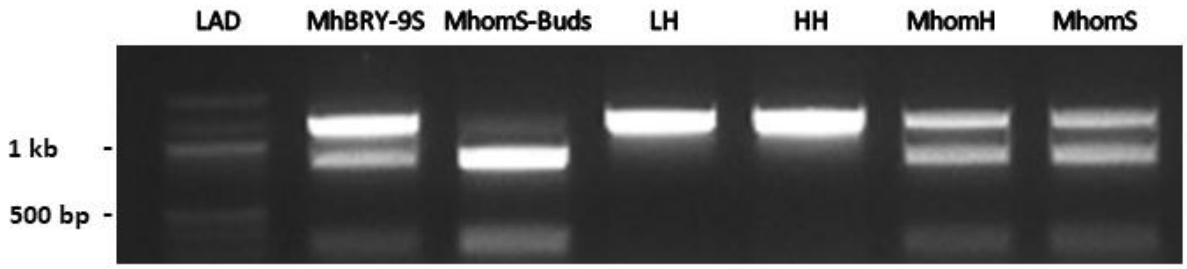


Figure 13. Digestion of PCR amplicons of *TsBAHD* gene in long-homostyle Mhomo H and its progeny with Mfe1 restriction enzyme. The samples were visualized on a 0.8% agarose gel. If the DNA remains undigested, a single band with a size of 1210 bp should occur. If the DNA is digested, it is cut into two fragments of 867 bp and 343 bp. DNA samples used: (MhBRY-9S) a heterozygote S-morph of a cross between Mhomo-H x BRY (Table 3); (MhomS-Buds) DNA from floral buds of an S-morph revertant Mhomo-S; (LH) Mhomo-H long-homostyle progeny 734H; (HH) Mhomo-H homozygote long-homostyle progeny HH12; (MhomH) long-homostyle Mhomo-H; and (MhomS) DNA from mature flowers of S-morph revertant Mhomo-S. LAD: 100 bp ladder.

The amount of each form of *TsBAHD* DNA (mutant and non-mutant) possessed by the periclinal chimeras is a function of how many cells carry one or the other copy of *TsBAHD*. This could vary from one tissue to another and/or among different shoots of the plants used. To quantify the amount of the two forms of *TsBAHD*, the relative intensity was used to calculate the percent composition of both copies using ImageJ (Figure 13; Schneider *et al*., 2012). As expected, the true heterozygote S-morph plant, MhBRY9S, possessed approximately equal proportions of both forms of *TsBAHD*: 52% for mutant allele and 48% for the non-mutant (Figure 14). In contrast, DNA from the putative periclinal chimera Mhomo-S (using both buds and mature flowers) possessed a greater percentage of the non-mutant copy of *TsBAHD* (77% and 68%, Figure 14). On the other hand, using flower buds of Mhomo-H, the non-mutant copy of *TsBAHD* was 56% and the mutant 44% (Figure 14). An important issue is to determine which tissues of a flower or bud possess mutant to non-mutant copies.

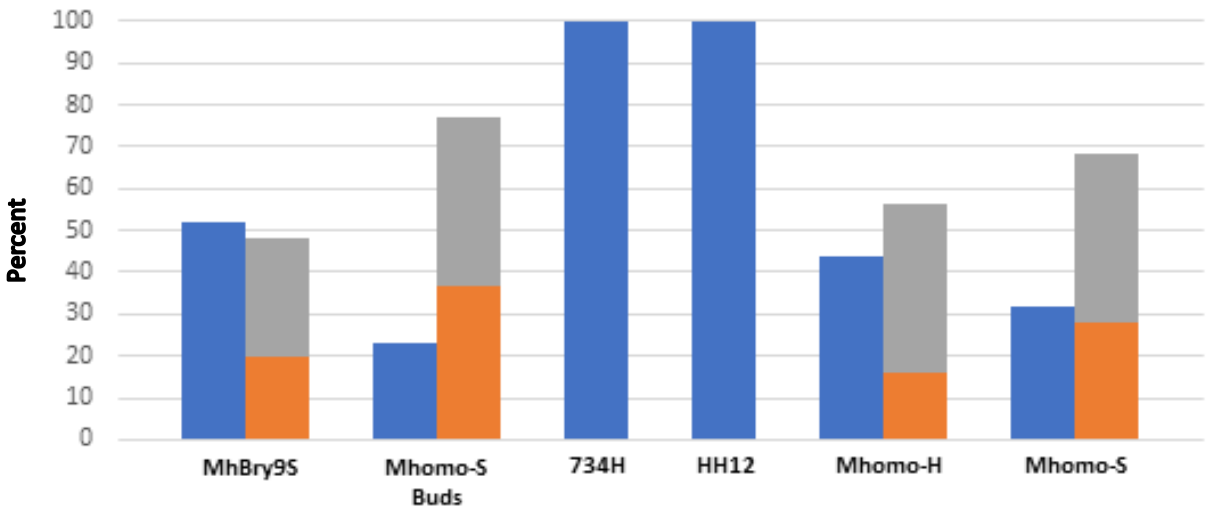


Figure 14. Percent composition of mutant vs non-mutant forms of *TsBAHD* in S-morph and long-homostyle plants from Mhomo-H family. Each bar represents the band intensity from a single digest experiment. Blue indicates the mutant copy (1210 bp) of *TsBAHD* while Grey and Orange indicate the non-mutant copy of *TsBAHD* (where grey represents the 867bp band and orange the 343 bp). The intensity of the mutant and non-mutant copies of *TsBAHD* (Figure 13) was measured using ImageJ (Schneider *et al*., 2012).

As a periclinal chimera grows, different lateral branches may contain different proportions of mutant to non-mutant genes as a function of patterns of cell division. Likewise, different floral organs and tissues may contain different proportions of mutant to non-mutant copies of the gene. To explore variation among branches, DNA was extracted from flower buds from separate branches from the long-homostyle mutant Mhomo-H and its S-morph revertant Mhomo-S to study their *TsBAHD* composition.

For Mhomo-S, *TsBAHD* was PCR amplified from the DNA of flower buds from five different branches and digested with Mfe1 restriction enzyme. All five samples possessed both copies of *TsBAHD* in their genome indicated by a band at 1210 bp for the mutant allele (it lacks the Mfe1 restriction site and so yields a single 1210 bp band) and two additional bands at 867/343 bp for the non-mutant allele (Figure 15a). The non-mutant allele of *TsBAHD* displayed a greater percentage (~67%) compared to the mutant copy (33%) in all five branches of Mhomo-S studied (Figure 15b).

The composition of *TsBAHD* in Mhomo-H was studied in greater detail. DNA was extracted from open flowers of four different branches. First, the styles were removed from all flowers, pooled together (due to their small size), and their DNA was extracted. Then, the DNA of the remaining organs was extracted from each flower separately. The use of open flowers ensures that they are long-homostyles (i.e. they are not short revertant). Furthermore, the composition of *TsBAHD* in the styles was also being studied.

All tissues studied possessed both copies of the *TsBAHD* gene indicated by the presence of the three bands: 1210 bp for the mutant and 867/343 bp for the non-mutant (Figure 16a). The

two S-morph progeny of Mhomo-H were included as positive controls to ensure that digestion went to completion (Figure 16a, lanes 5 and 6). The percent composition of *TsBAHD* was studied in the different floral organs of Mhomo-H (Figure 16b). The pooled styles possessed an equal percentage of the non-mutant copy of *TsBAHD* (51%) and the non-mutant copy (49%). The percentages for the remaining floral tissues of the four Mhomo-H flowers varied somewhat; (mutant/non-mutant) 35/65%, 45/55%, 46/54%, and 38/62% (Figure 16b).

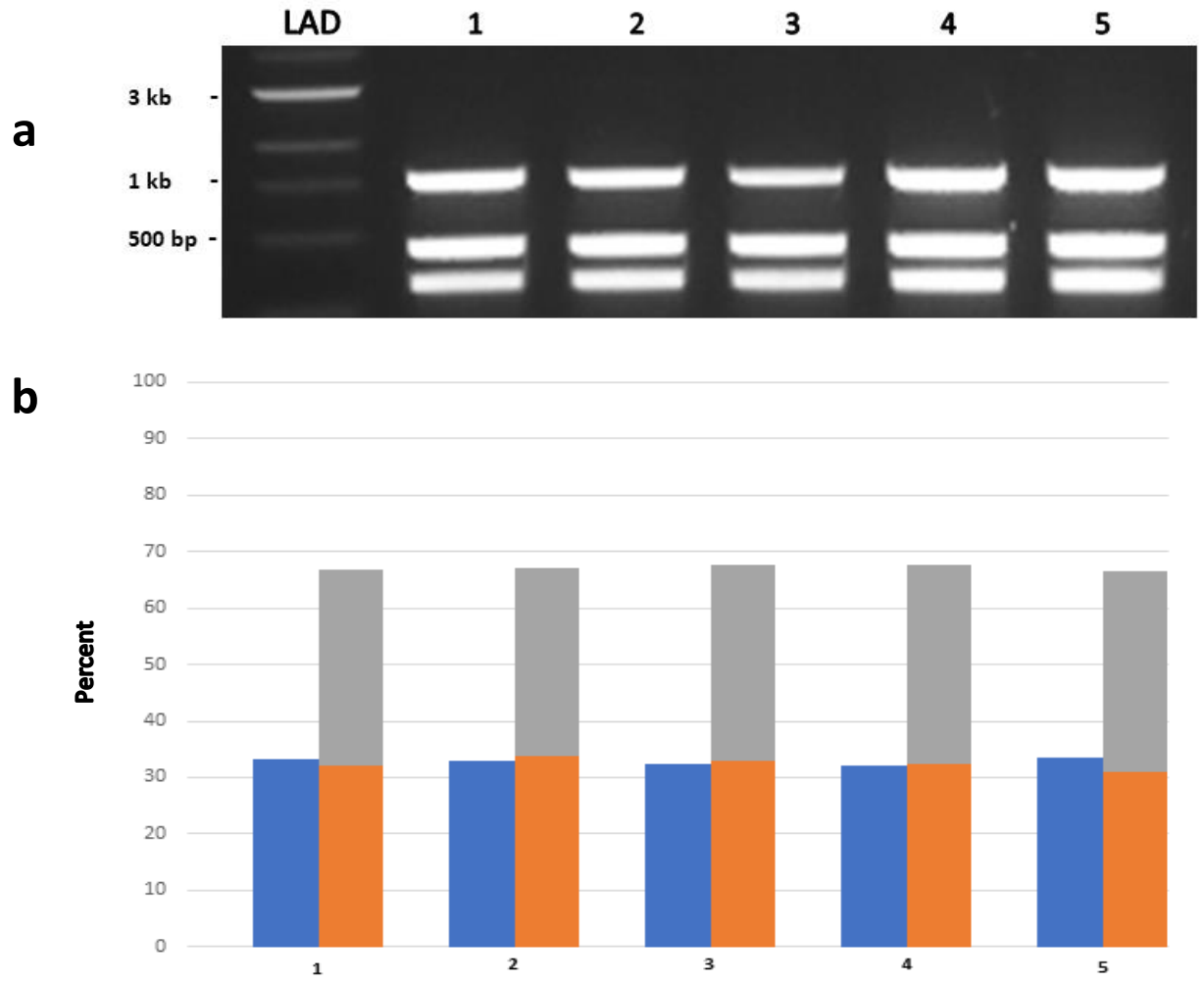


Figure 15. Digestion of *TsBAHD* gene in S-morph Mhomo-S revertant branches with the Mfe1 restriction enzyme. (a) DNA extracted from flower buds was PCR amplified and then digested with Mfe1 restriction enzyme. Results were visualized on a 0.8% agarose gel. Each bar represents the band intensity from a single digest experiment. The mutant form of *TsBAHD* is represented as a single band at 1210 bp while the non-mutant copy is represented by DNA bands at 867 bp and 343 bp. DNA samples used: 1-5 each represent DNA extracted from flower buds from separate branches of the S-morph revertant Mhomo-S. LAD: 1 kb ladder. (**b**) Blue: the mutant copy of *TsBAHD*. Grey and orange: the non-mutant copy of *TsBAHD*. The intensities of the bands in (**a**) were measured using ImageJ software (Schneider *et al*., 2012) and the percent composition was calculated using Excel.

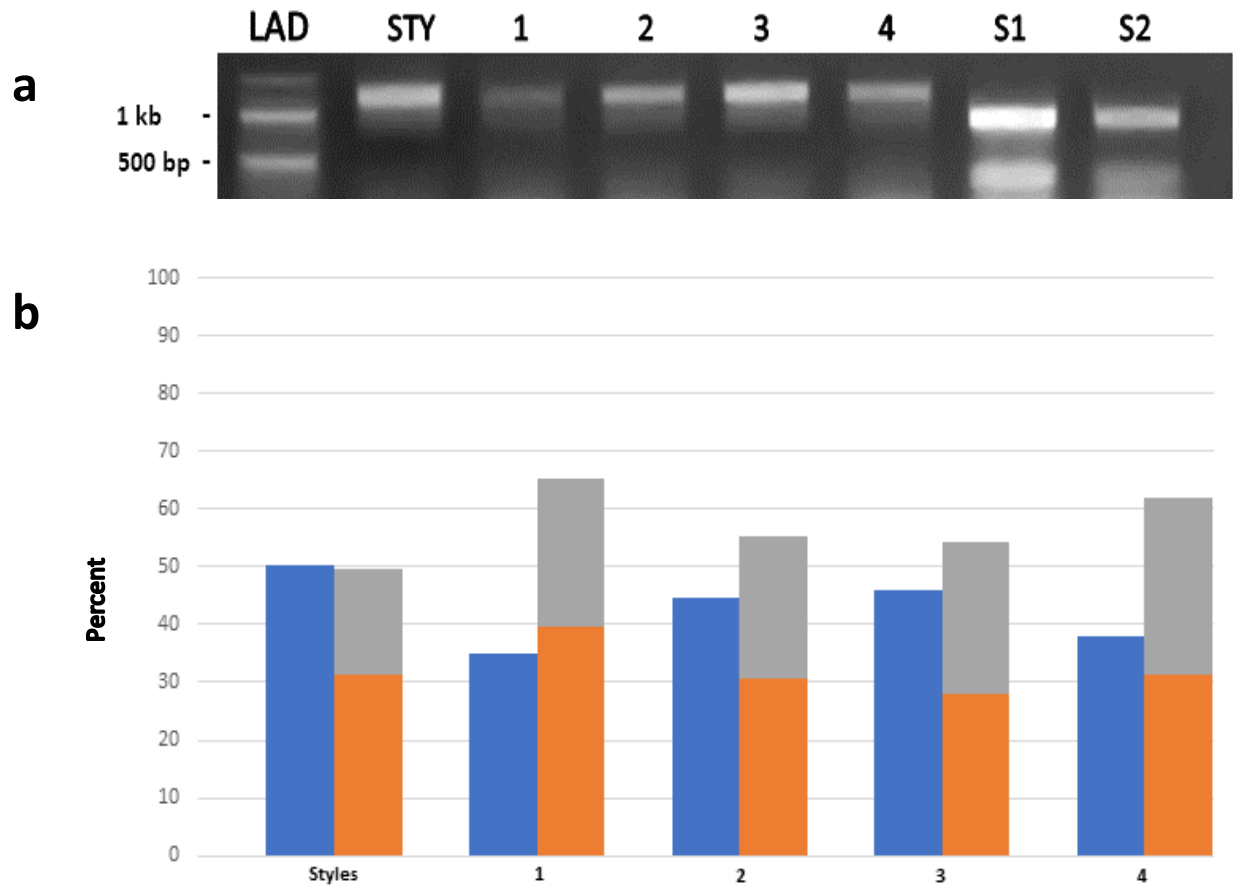


Figure 16. Digestion of *TsBAHD* gene in long-homostyle Mhomo-H flowers with the Mfe1 restriction enzyme. (**a**) DNA extracted from mature flower organs was PCR amplified then digested with Mfe1. Results were visualized on a 0.8% agarose gel. Each bar represents the band intensity from a single digest experiment. The mutant form of *TsBAHD* is represented by a single band at 1210 bp (blue) while the non-mutant form is represented by DNA bands at 867 (orange) and 343 bp (grey). Samples used: (STY) Pooled stylar tissue of four long-homostyle flowers from different branches of Mhomo-H, (1–4) remaining flower organs pooled excluding styles and much of the petals of four different flowers; S-morph progeny of Mhomo-H (S1) 886S, and (S2) 726S. LAD: 100 bp ladder. (**b**) Blue: the mutant form of *TsBAHD*. Grey and orange: the non-mutant form of *TsBAHD*. The intensities of the bands in (**a**) were measured using ImageJ software (Schneider *et al*., 2012) and the percent composition was calculated using Excel.

4.5.3 Genes controlling stamen length using the short-homostyle TrinSH

To determine the genes responsible for controlling the stamen length, the three *S-locus* hemizygous genes were PCR amplified from the DNA of a self-incompatible short-homostyle mutant of autotetraploid *T. subulata*, TrinSH, and three of its progeny (two short-homostyle and one S-morph). All genes amplified in the genome of TrinSH, its three progeny, and the S-morph control plant (Figure 17).

Since the mutant phenotype of TrinSH is the shortened stamen, the expression of the stamen-specific genes, *TsSPH1* and *TsYUC6*, was studied using RT-PCR (Figure 18). There appeared to be no expression of *TsSPH1* in TrinSH or in its short-homostyle progeny. The S-morph plants used in this experiment expressed *TsSPH1*, although the single S-morph progeny of TrinSH (Lane S1) appeared to have a reduced expression (Figure 18). As expected, no expression was observed for the L-morph (Figure 18). In contrast, *TsYUC6* was expressed in all short-homostyle mutants and S-morph plants, although somewhat weakly in one of the short-homostyle progeny (Lane SH3) (Figure 18). As expected, no expression of *TsYUC6* was observed for the L-morph control plant.

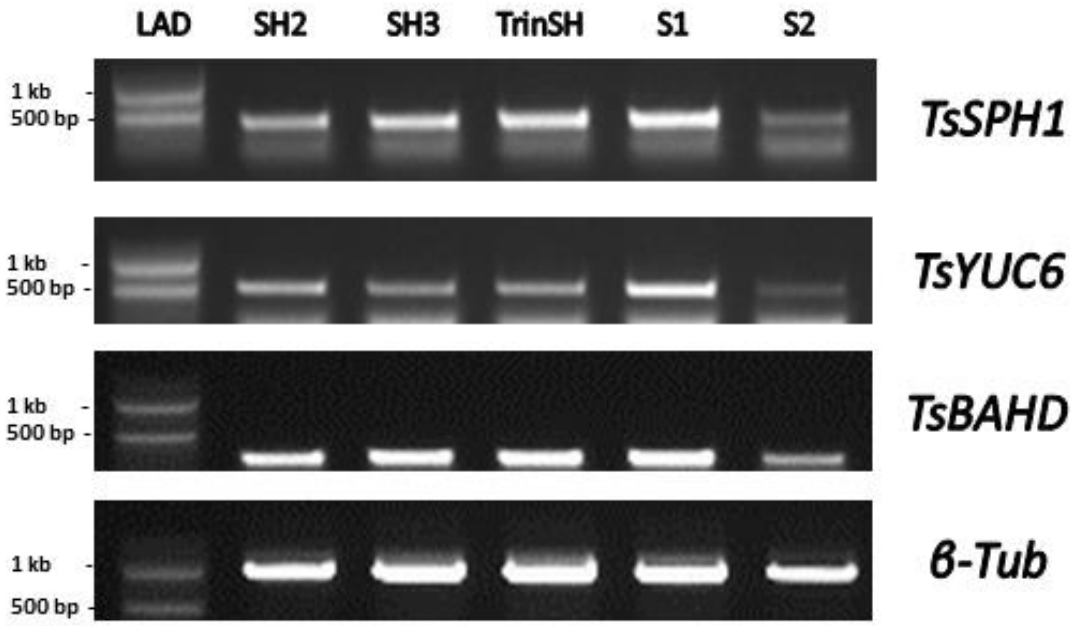


Figure 17. Screening for the three candidate genes, *TsSPH1*, *TsYUC6*, and *TsBAHD* in the genome of the short-homostyle mutant, TrinSH, and its progeny. DNA extracted from flowers buds was PCR amplified using gene-specific primers (Table S1). Results were visualized on a 0.8% agarose gel. Samples used: (SH2) TrinSH short-homostyle progeny L20SH; (SH3) TrinSH short-homostyle progeny L21SH; (TrinSH) TRINSH short-homostyle mutant; (S1) TrinSH S-morph progeny L38S; (S2) *T. subulata* S-morph control plant E71S. *β-Tubulin* (*β -Tub*) was used as a positive control. LAD: 100 bp ladder.

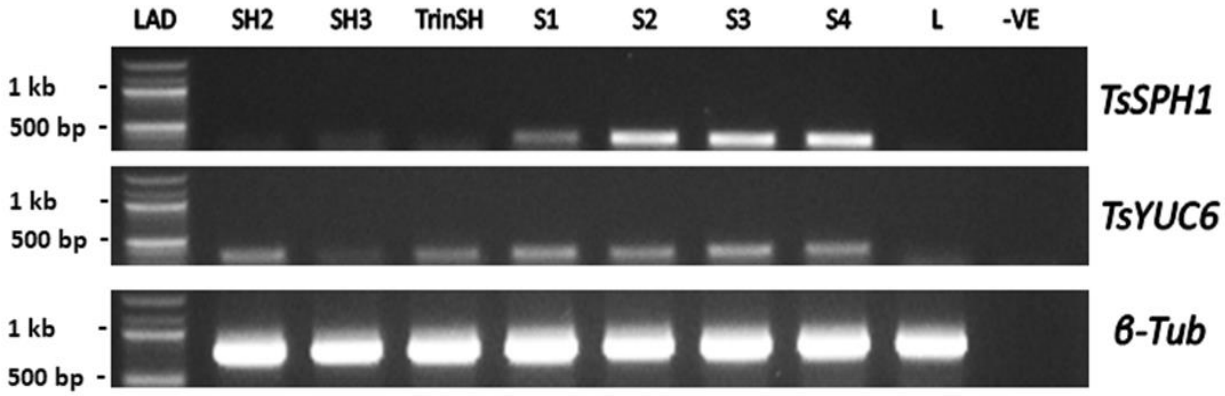


Figure 18. Expression of *TsSPH1* and *TsYUC6* in stamen-tissue of short-homostyle mutant TrinSH and its progeny. RT-PCR was used to study the expression of *TsSPH1* and *TsYUC6* from stamen tissue. Results were visualized on a 0.8% agarose gel. Samples used: (SH2) L20SH short-homostyle progeny of TrinSH, (SH3) L21SH short-homostyle progeny of TrinSH, (TrinSH) TrinSH short-homostyle mutant, (S1) L35S S-morph progeny of TrinSH; *T.subulata* S-morph plants (S2) E71S and (S3) E207S; (S4) S-morph plant *T. subulata* BOT4S; (L) *T.subulata* L-morph plant E214L. *β-Tubulin* (*β-Tub*) was used as a positive control. LAD: 100 bp ladder.

4.6 Self-incompatibility system in *Turnera*

4.6.1 Characterization of the self-compatibility in the SSC S-morph mutant

The plant giving rise to the S-morph SSC mutant was initially SI. However, it started producing some branches that were SC (Figure 19). To characterize and determine if the mutation is at or linked to the *S*-locus, cross and self-pollinations were performed using the normal (SI) and mutant (SC) branches of the original plant. In addition, cuttings rooted from various branches of the plant in order to maintain both the mutant and non-mutant portions of the plant in the longer term.

Does the mutation alter the pollen or style incompatibility response?

The self-compatibility mutation was characterized by reciprocally crossing normal (SI) and mutant branches (SC) onto the styles of a normal (SI) L-morph plant and onto those of a normal (SI) S-morph plant (Table 4). Pollen from the normal SI branches fertilized the L-morph plants setting an average of 19.55 ± 3.4 seeds per pollination but failed to fertilize the styles of the normal (SI) S-morph plants, as expected (Table 4). Pollen from the mutant SC branches fertilized the styles of the normal L-morph setting an average of 27.33 ± 1.7 seeds (Table 4). However, pollen from the mutant SC branches also successfully fertilized normal S-morph plants setting an average of 39.4 ± 2.9 seeds.

Then, a series of crosses were conducted to determine if the styles of the mutant have lost their incompatibility response. Styles of both normal SI and mutant SC branches rejected pollen from the normal (SI) S-morphs but set seeds when pollinated by normal L-morph plants with an average of 36.8 ± 3.6 and 25.8 ± 1.2 seeds, respectively (Table 5). These results suggest that a

mutation is altering the function of the pollen and not the styles of the mutant branches of SSC plant. Pollen of the SSC mutant will fertilize both L-morph and S-morph plants.

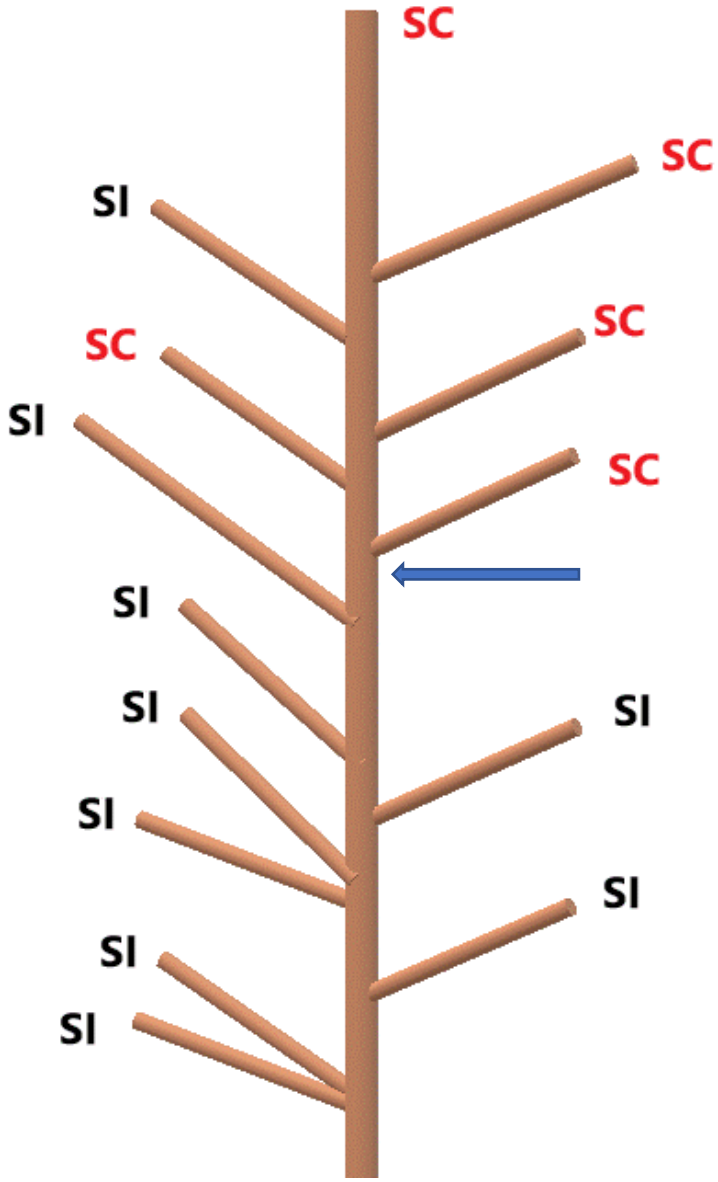


Figure 19. Schematic of the SSC S-morph mutant. SI refers to the normal self-incompatible branches. SC refers to the mutant self-compatible branches. The arrow indicates the potential site of which the mutation originated although note the occurrence of one SI branch above three SC branches.

Table 4. Results of crossing pollen of mutant SC and normal SI branches onto normal styles of S-morph plants and L-morph plants. The female parent (pollen recipient) is listed first. SC refers to the S-morph self-compatible mutant shoots whereas SI refers to self-incompatible normal shoots. The mean is based on n= 9 pollinations. SE is the standard error of the mean. "Fruit percentage" was calculated by dividing the number of pollinations that successfully resulted in fruits by the total number of pollinations. t-tests were used to compare the mean seed set when the L-morph plants were the pollen recipients and when the S-morph plants were the pollen recipients.

Cross	Mean (n=9) ± SE	Fruit Percentage (%)	t-test of mean seed set
SI L-morph style x SC mutant pollen	19.55 ± 3.4	100%	2.05 ns
SI L-morph style x SI S-morph pollen	27.33 ± 1.7	100%	
SI S-morph style x SC mutant pollen	39.4 ± 2.9	100%	13.92**
SI S-morph style x SI S-morph pollen	0	0%	

ns not significant, **P-value < 0.0001.

Table 5. Results of crossing normal SI pollen of S-morph and L-morph plants onto styles of SSC mutant branches and non-mutant (SI) branches. The female parent is listed first. SC refers to the S-morph self-compatible shoots whereas SI refers to the self-incompatible shoots. The mean is the average of seeds set by all replicates of each cross. SE is the standard error of the mean. "Fruit percentage %" was calculated by dividing the number of pollination that successfully resulted in fruits by the total number of pollinations.

Cross	Mean ± SE	Fruit Percentage (%)
SI S-morph style x SI S-morph pollen (n=7)	0	0%
SC mutant style x SI S-morph pollen (n=9)	0.22 ± 0.15	22.2%
SI S-morph style x SI L-morph pollen (n=8)	36.8 ± 3.6	100%
SC mutant style x SI L-morph pollen (n=10)	25.8 ± 1.2	100%

4.6.2 Inheritance of the mutation causing SC

The inheritance of the self-compatibility mutation was explored using self- and cross-pollinations to determine if the gene(s) conferring self-compatibility resides at the S-locus.

In the first generation, selfing the mutant (SSC) resulted in three SC S-morph progeny and 28 SI L-morph progeny. When SSC was crossed to an L-morph plant, this resulted in a single SC S-morph progeny and 28 SI L-morph progeny (Table 6). The segregation ratios showed an extreme excess of L-morph progeny. The cause(s) of these aberrant ratios is unknown but perhaps result from reduced viability of progeny carrying the mutation. The four S-morph progeny above were selfed to produce the second-generation of progeny. This gave 53 L- and 18 S-morph progeny which deviated from the expected 3 S-morph: 1 L-morph ratio (Table 7). Selfing each progeny 10 times, all 18 S-morph progeny of the second-generation set seeds indicating that they carry the SC mutation. On the other hand, the 53 L-morph progeny were SI as they produced no seeds. The seeds produced by the second-generation S-morphs were germinated to give rise to the third generation (Table 8).

The third-generation progeny were composed of 160 L-morph progeny and 64 S-morph progeny, deviating again from the expected 3 S-morph: 1 L-morph ratio (Table 8). One family (J30L x SSC 12S) yielded a greater proportion of S-morph progeny in the third generation compared with the other families; however, the ratio is still markedly distorted. The compatibility of the third-generation progeny was tested as above. At least five L-morph progeny from each family were tested for their compatibility and all are SI (Table 9). All 64 S-morph progeny were SC. However, two of these progeny yielded a low number of seeds upon

selfing, with the mean (n=10) of 2.25 ± 4.09 seeds set by both plants, and thus they were considered as SI. These results suggest that the gene causing the self-compatibility mutation of SSC is closely linked to, but not at the S-locus. The causes of the aberrant ratios (above) require further exploration.

Candidate genes in the SSC mutant

The results show that the mutation of SSC affects the function of the pollen, therefore the presence/absence of the stamen-specific genes, *TsSPH1* and *TsYUC6*, was studied in both normal SI and mutant SC branches using PCR amplification. Both branches possessed the two genes in their genome (Figures 20 and 21). If a mutation in *TsSPH1* and/or *TsYUC6* is responsible for SC in the SSC mutant, the sequence and expression of the two genes would need to be explored. The inheritance study above, however, indicates that a closely linked gene is likely responsible for the SC of the mutant, not the *TsSPH1* or *TsYUC6* genes.

Table 6. First-generation progeny of crosses between normal SI and mutant SC branches onto normal SI S- and L-morph plants. SC refers to the self-compatible mutant shoots whereas SI refers to self-incompatible normal shoots. The female parent is listed first. J30L is an L-morph plant of *T. joelii*; J211S is a control S-morph plant of *T. joelii*.

	Observed			
	L-morph progeny	**S-morph progeny**	**Expected ratio**	**Chi-squared**
J30L x SI	11	13	1 Long: 1 Short	0.167 ns
SC x J30L	28	1	1 Long: 1 Short	25.14**
***J211S x SC**	11	10	1 Long: 3 Short	8.39**
SC selfed	28	3	1 long: 3 short	70.55**

ns indicates that the differences are not significant. **P-value<0.05 indicates that the differences are significant
*all S-morph progeny from this cross were self- incompatible suggesting the "S-allele" they contain came from the normal J211S parental plant. The ratio is also consistent with this hypothesis.

Table 7. Second-generation progeny. The four self-compatible progeny (SSC-4S, SSC-13S, SSC-18S, and SSCxJ30L-12S) from the first-generation were selfed and the number of S-morph and L-morph progeny was determined. J30L x SSC 12S is the single S-morph progeny that arose from the cross SSCxJ30L.

Self	L-morph progeny	S-morph progeny	Chi-Squared
SSCxJ30L- 12S selfed	10	6	12*
SSC-18S selfed	4	2	5.55*
SSC- 4S selfed	8	4	11.11*
SSC- 13S selfed	31	6	68.18*
Total	53	18	96.84*

*P-value <0.0001 indicates a significant deviation from the 3 S-morph: 1 L-morph expected ratio

Table 8. Third-generation progeny of selfed S-morph progeny of SSC mutant. SSC refers to the S-morph self-compatible mutant shoot. SSC- 4S, 13S, 18S, and J30LxSSC- 12S refer to the 18 S-morph plants from Table 7 (second generation).

Self	L-morph progeny	S-morph progeny	Chi-squared
SSC- 4S selfed	64	9	152.92*
SSC- 13S selfed	24	10	37.69*
SSC- 18S selfed	19	4	46.19*
SSCxJ30L- 12S selfed	53	41	49.38*
Total	160	64	286.18*

*P-value <0.0001 indicates a significant deviation from the 3 S-morph: 1 L-morph expected ratio

Table 9. Seed counts from the selfing of the third-generation S-morph plants. SSC refers to the mutant S-morph self-compatible shoot. The mean seed count is the average number of seeds from 10 separate pollinations for each S-morph from each family.

Self	Mean seed count ± SE
SSC- 4S selfed	9.26 ± 4.14
SSC- 13S selfed	15.54 ± 4.15
SSC- 18S selfed	12.32 ± 4.18
J30LxSSC- 12S selfed	14.6 ± 9.75

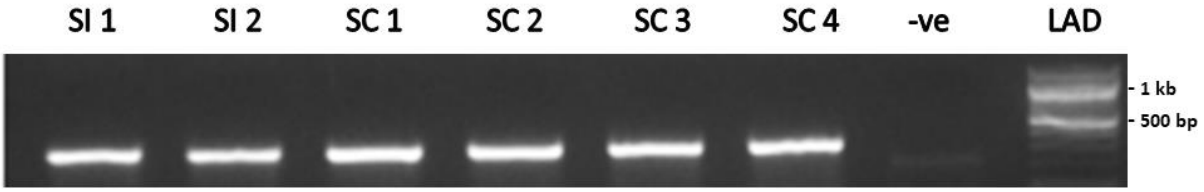


Figure 20. Screening for the stamen-specific gene, *TsSPH1*, in the self-compatible S-morph mutant plant, SSC. DNA was PCR amplified using specific primers (Table S1) and results were visualized on a 2% agarose gel. Samples used: (SI) refers to self-incompatible normal branches and (SC) refers to self-compatible mutant branches of the SSC plant; (-ve) negative control. LAD: 100 bp ladder.

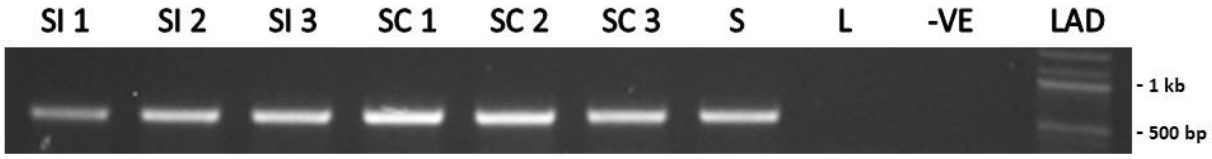


Figure 21. Screening for the stamen-specific gene, *TsYUC6*, in the self-compatible S-morph mutant plant, SSC. DNA was PCR amplified using primers specific for exons 3 and 4 of *TsYUC6* (Table S1). Results were visualized on a 1% agarose gel. Samples used: (SI) refers to self-incompatible normal branches and (SC) refers to self-compatible mutant branches of the SSC plant; (S) S-morph plant of *T. joelii*; (L) L-morph plant of *T. joelii* ; (-VE) negative control. LAD: 100 bp ladder.

5. Discussion

This study revealed major findings regarding the genes controlling distyly in *Turnera*, rivaling recent work on *Primula* (Li *et al*., 2008; Huu *et al*., 2016). A hemizygous region comprised of three genes, *TsSPH1*, *TsYUC6,* and *TsBAHD*, was found on the *S*-haplotype and appears to determine the S-morph of several species of *Turnera*. With three inserted genes and at least two inversions, the dominant *S*-haplotype possesses a novel gene arrangement while its recessive counterpart (the *s*-haplotype) possesses an ancestral gene order. Below I discuss the arrangement of *S*-locus genes and the potential role of the three hemizygous genes in controlling distyly in *Turnera*.

5.1 Organization of the hemizygous genes

The recessive *s*-haplotype of *T. subulata* showed considerable collinearity with several species in the order Malpighiales including: *Vitis vinifera* where a region of chromosome 7 contains an identical gene order with the exception of an additional gene; *Populus trichocarpa* (chromosome 7) which has a similar gene order with the exception of two missing genes, and two additional ones (Shore *et al*., 2019). Thus, the *s*-haplotype has an ancestral gene order while the *S*-haplotype possesses a novel arrangement with three inserted genes and at least two inversions. In addition, the *S*-haplotype is 4.6 times larger than the *s*-haplotype (Shore *et al*., 2019).

The three hemizygous genes occur on a single BAC clone (BAC-A24, 247288 bp) on the *S*-haplotype and are absent from the *s*-haplotype suggesting that they are the result of an insertion(s) relative to the latter. This provides an explanation for the suppressed

recombination in which there are no recessive alleles for these hemizygous genes to recombine with. In addition, it appears that at least two inversions occurred as well. Inversions will extend the degree of recombination suppression. As a result, we expect the accumulation of transposable elements on the *S*-haplotype (Bennetzen, 2000; Uyenoyama, 2005). In fact, the *S*-haplotype has a greater amount of transposable elements (approximately 77%) compared to the *s*-haplotype (Shore *et al*., 2019). Further rearrangements of the *S*-haplotype might be expected in divergent lineages because of the possibility of ectopic recombination (recombination among transposable elements). With the three genes occurring in a 247 kb region, translocations and other genetic processes could even move the genes to other genomic locations.

5.2 Other possible candidate genes

While three hemizygous genes on the *S*-haplotype appear to determine the S-morph, it is not clear whether any of the remaining genes from the haplotypes play any role in determining distyly. Based upon deletion mapping alone, *TsSKIP2*, *TsSDN3*, and *TsPTEN* remain as candidate genes because the *S*-haplotype alleles of these genes have been deleted from all X-ray mutants (the mutants possessed the recessive *s*-haplotype copy of these genes).

Interestingly, *TsSKIP2* showed homology to the *S*-locus F-box protein that determines the pollen SI function of the S-RNase-mediated self-incompatibility in *Antirrhinum, Petunia,* and *Prunus* species (Wang *et al*., 2004). In addition, *TsSKIP2* is functionally redundant with three other proteins from the *VIER F-BOX PROTIENE* family (German for four F-box proteins) where together, they regulate auxin and ethylene responses in *Arabidopsis* (Schwager *et al*., 2007) and

as both auxin and ethylene are phytohormones, they could play a role in morphological aspects of distyly. The *S*-haplotype copy of *TsDUF247* was absent from all X-ray mutants except for one L-morph plant, L22. *TsDUF247* showed homology to two genes that were recently discovered in *Lolium perenne, S-DUF247* and *Z-DUF247*, which are candidate genes for controlling the male self-incompatibility system (Manzanares *et al*., 2015; Thorogood *et al*., 2017). Further work is required to assess their roles, if any, in *Turnera*.

5.3 The roles of the hemizygous genes

Deletion mapping and genome walking identified three hemizygous genes that occur only in S-morph plants. *TsSPH1* (expressed filaments and anthers) and *TsYUC6* (expressed in anthers) show stamen-specific expression whereas *TsBAHD* exhibits style-specific expression. The occurrence of three genes determining distyly is remarkably similar to the classic model proposed for distyly in *Primula* (Ernst, 1955; Barrett and Shore, 2008; Cohen, 2010), the difference being that the three genes *Turnera* are hemizygous (i.e. they lack recessive allelic counterparts as in the classical *Primula* model).

5.3.1 *TsBAHD*: a candidate gene for style length

TsBAHD showed homology to *AtBAHD* (*AT3g26040*) from *Arabidopsis thaliana* that belongs to the BAHD acyltransferase family. The BAHD family has a wide array of biochemical functions and activities which includes utilizing CoA thioesters to catalyze the formation of various groups of plant metabolites (St Pierre and De Luca, 2000; D'Auria *et al*., 2006). Such products include small volatile esters, modified anthocyanin, constitutive defense compounds, and phytoalexins.

Both *AtBAHD* and *TsBAHD* belong to the alcohol acyltransferase like enzyme family (AATLs) of the BAHD gene family (Tuominen *et al*., 2011).

TsBAHD is the prime candidate gene for controlling the style length in *Turnera*. It was found to be specific to S-morph plants and expressed in pistils. *TsBAHD* was absent from the genome of the long-homostyle mutant Drh and from the X-ray deletion mutant LH1 (both of which have long styles). In addition, there is a mutation in the *TsBAHD* gene of the long-homostyle mutant Mhomo-H and its homostyle progeny providing further support that *TsBAHD* is responsible for short styles in S-morph plants.

In Mhomo-H, a single mutation was discovered in a functionally conserved motif (GN) that substituted serine (S) for asparagine (N) at amino acid residue 290. The GN motif is conserved among all members of the BAHD family and especially among the anthocyanin acyltransferases (AATs) and alcohol acyltransferase like enzyme (AATLs) families (D'Auria *et al*., 2006; Suzuki *et al*., 2007; Unno *et al*., 2007). The crystal structure of Vinorine synthase (a member of BAHD family) from the plant *Rauvolfia*, revealed that the GN motif is located in one of the β-loops that form the solvent channel located beside the active site of the enzyme (Ma *et al*., 2005; Suzuki *et al*., 2007). In addition, the crystal structure of an anthocyanin malonyltransferase homolog revealed that the catalytic histidine residue of the active site (HxxxD) forms hydrogen bonds with the asparagine residue of the GN motif (Ma *et al*., 2005). These hydrogen bonds are important for the overall catalytic activity of the enzyme (Suzuki *et al*., 2003; Unno *et al*., 2007). The importance of the GN motif in Vinorine synthase was highlighted by Bayer *et al*. (2004). They found that substituting asparagine (N) of the GN motif with alanine (A) significantly reduced the relative activity of the enzyme to 68%. Thus, the asparagine of the GN motif is

important for the greater catalytic function of Vinorine synthase. While *TsBAHD* is not the same gene as Vinorine synthase, these results suggest that the function of the mutant *TsBAHD* is likely impaired and responsible for elongated styles of the Mhomo-H mutant and its progeny.

BLAST searches of *TsBAHD* resulted in top hits for genes that inactivate brassinosteroids through acetylation (Wang *et al*., 2012; Choi *et al*., 2013). Recent studies in *Primula* revealed an *S*-locus gene, *CYP734A50*, that controls the length of styles (Neff *et al*., 1999; Huu *et al*., 2016). *CYP734A50* regulates the activity of brassinosteroids by inducing their degradation (Burrows and McCubbin, 2018). Brassinosteroids regulate cell expansion, and evidence suggests that reduced levels of brassinosteroids decrease the cellular expansions of style cells in the short styles of the S-morph plants. Thus, S-morph plants possess reduced levels of brassinosteroids compared to L-morph plants. Long-homostyles of mutants of *P. vulgaris* possessed a single base-pair insertion in *CYP734A50* gene that strongly reduced its expression. The reduced expression of *CYP734A50* resulted in eliminating suppression of the style length, and hence allowed style elongation to occur (Huu *et al*., 2016; Li *et al*., 2016). Virus-induced silencing of *CYP734A50* in *P. forbessi* S-morph plants resulted in a long-homostyle phenotype. The re-addition of brassinosteroids restored the phenotype back to S-morph flowers (Huu *et al*., 2016) providing strong support for *CYP734A50* as the determinant of short style length.

The findings presented here, identify *TsBAHD* as the candidate gene for short style length in *Turnera*, possibly through reducing the number of brassinosteroids to achieve the S-morph phenotype. The single substitution mutation found in the GN motif of *TsBAHD* might be reducing the catalytic function of the BAHD enzyme in the long-homostyle mutant (Mhomo-H)

leading to a long style length. Further research is necessary to investigate the function of *TsBAHD* and its role in controlling the activity of brassinosteroids in *Turnera*.

5.3.2 *TsSPH1*: a candidate gene for filament length

TsSPH1 was found to be expressed in S-morph filaments (Shore *et al*., 2019). *TsSPH1* showed homology to the *SPH1* gene from the S-protein homolog family (*SPH*) from *Papaver rhoeas* and *Arabidopsis thaliana* (Ride *et al*., 1998; Shore *et al*., 2019). The *SPH* family has 84 members of which their function is yet to be determined except for one, that is, *SPH1* of *Papaver rhoeas* (Ride *et al*., 1999; Wheeler *et al*., 2009 and 2010). *SPH1* was determined to be the female incompatibility gene in *Papaver rhoeas*, *PrsS* (Foote *et al*., 1994; Ride *et al*., 1999). The *PrsS* gene is expressed in the pistil and acts as the female determinant of the gametophytic SI system (Foote *et al*., 1994; Wheeler *et al*., 2009). It codes for a protein that acts as signaling peptide which interacts with its pollen-specific transmembrane receptor counterpart, *PrpS* (*P. rhoeas* pollen S) (Wheeler *et al*., 2009). Upon the binding of self PrsS peptide to a PrpS receptor, a Ca^{2+} signaling cascade is activated which results in the inhibition of the self-pollen tube and the activation of programmed cell death (Zhang *et al*., 2009; Wheeler *et al*., 2009; Wu *et al*., 2011; de Graaf *et al*., 2012). Unlike *SPH1* of *Papaver*, *TsSPH1* is associated with male function as it is expressed in S-morph filaments and anthers. Furthermore, the expression of *TsSPH1* was "suppressed" in the short-homostyle mutant (TrinSH) and its short-homostyle progeny. These findings indicate that *TsSPH1* might be playing a role in controlling the length of the filament. That is, it is responsible for elongating the filaments in S-morph plants. Further experiments are needed to investigate this role in *Turnera*.

5.3.3 *TsYUC6*: a candidate gene for anthers

The third and final candidate gene, *TsYUC6*, was expressed in anthers of the S-morph plants (Shore *et al*., 2019). BLAST searches of *TsYUC6* indicated that the gene belongs to the flavin-dependent monooxygenase YUCCA family (Cheng *et al*., 2006). Members of this family are responsible for catalyzing the second step of auxin synthesis from L-tryptophan (Cheng *et al*., 2006). The final product, auxin, is known to positively trigger filament elongation and negatively control anther dehiscence and pollen maturation (Checchetti *et al*., 2008; 2017). In addition, *TsYUC6* showed homology to *AtYUC6* in *A. thaliana* which is mainly expressed in anthers and pollen (Cheng *et al*., 2006; Shore *et al*., 2019). Studies have shown that *AtYUC6* functions with *AtYUC2* where both have been determined to be the main auxin biosynthetic flavin monooxygenases in anthers (Yao *et al*., 2018). Knocking-down *AtYUC6* and *AtYUC2* reduced the expression of auxin dramatically in the anthers and resulted in late maturation and extremely low pollen production (Cheng *et al*., 2006).

Recent genetic analysis revealed that both *AtYUC2* and *AtYUC6* act as sporophytic genes for pollen formation (Yao *et al*., 2018). That is, the production of auxin in sporophytic microsporocytes is required for the early stages of pollen development. Another study suggested that at the end of meiosis, auxin produced in anthers is transported to filaments where it triggers elongation and prevents pollen maturation (Cecchetti *et al*., 2017). This implies that lowering the expression of auxin would reduce the length of filaments whereas the opposite would induce an increase in length. Therefore, auxin, which is controlled by *AtYUC2* and *AtYUC6*, determines the filament length and the pollen maturation in *Arabidopsis* (Cecchetti *et al*., 2008; 2017).

The mechanisms and processes by which *AtYUC6* controls the length of the filament in *Arabidopsis* may be consistent with *TsYUC6* being the *S*-locus gene that controls the male characteristics of distyly in *Turnera,* possibly pollen size, pollen number, pollen SI and perhaps filament length, although *TsSPH1* is expressed in filaments, unlike *TsYUC6*. The fact that a short-homostyle mutant (TrinSH with short filaments) expresses *TsYUC6*, but not *TsSPH1*, suggests that *TsSPH1* is responsible for filament elongation (not *TsYUC6*), although the two gene products might interact. Further work is needed to explore the role(s) of *TsYUC6* in *Turnera*.

5.4 Self-incompatibility in *Turnera*

One outstanding question in distylous species is what causes the heteromorphic incompatibility system. While candidate genes for style and stamen lengths have been proposed in *Primula* (Ernst, 1955; Barrett and Shore, 2008; Cohen, 2010) and now here in *Turnera*, it is still unclear which genes are responsible for determining heteromorphic incompatibility. According to the Lloyd & Webb model for the evolution of distyly (1992), the self-incompatibility system may rise as a result of the co-adaptation between styles and pollen of the different morphs whose reproductive organs differ physiologically (Barrett and Shore, 2008). That is, there isn`t a self-recognition system, but physiological differences between styles and pollen lead to the inability of self (or intra-morph) pollen to grow through styles.

In line with Lloyd and Webb's model (1992), a mechanistic basis for incompatibility might be proposed as follows (Shore *et al*. 2019): *TsYUC6* is homologous to the auxin biosynthetic flavin monooxygenases that control the concentration of auxin (Cecchetti *et al*., 2008; Yao *et al*., 2018) while *TsBAHD* is homologous to *CYP734A50* which regulates brassinosteroids (Burrows

and McCubbin, 2018). A recent study found interactions between the auxin and brassinosteroids pathways that control growth and development in *Arabidopsis* (Tian *et al*., 2018). If the hypotheses introduced in this study regarding the floral morphology and organ development in *Turnera* are valid, we can propose the following mechanism for SI: pollen that developed under high concentrations of auxin produced by the *TsYUC6* gene of the S-morph, is only compatible with styles that developed under high concentration of brassinosteroids of the L-morph which doesn`t possess the *TsBAHD* gene. In contrast, pollen that developed under low concentrations of auxin in L-morph plants (which don`t possess *TsYUC6*) would only be compatible with styles that developed under low concentrations of brassinosteroids in S-morph plants, which possess TsBAHD that degrades brassinosteroids (Shore *et al*., 2019).

The analysis of transcriptomes of *Turnera* species can aid in identifying additional candidate genes that are differentially expressed between S- and L-morph plants, and that might represent "downstream" determinants of distyly. Such genes might have a role in controlling the SI system in *Turnera*. The mechanism proposed by Shore *et al*. (2019) requires more extensive studies to investigate the connection between the two hormones and their respective roles in the incompatibility system.

5.5.1 Long-homostyle Mhomo-H, a periclinal chimera?

Frequently, Mhomo-H produces flowers possessing the non-mutant S-morph phenotype (Mhomo-S). Mhomo-H and Mhomo-S appear to be periclinal chimeras (i.e. layer chimeras) as they possessed both copies of *TsBAHD* in their genomes (mutant and non-mutant; Figure 13). The shoot meristems of plants (SAM) are composed of a small population of pluripotent stem

cells, different layers of that give rise to various organs such as leaves, stems, and flowers (Fletcher, 2002). SAM is commonly composed of three layers: L1, L2, L3. For Mhomo-H, I hypothesize that these layers possess different copies of *TsBAHD* (mutant or non-mutant alleles) in their nuclei. Depending on the cell division patterns of the apical meristem, different lateral branches and organs (i.e. flowers) may end up with a varying number of cells that possess mutant vs non-mutant copies of *TsBAHD*. Since the fate of a cell in SAM is determined by its position (Taguchi-Shiubara *et al*., 2001; Fletcher, 2002), cells at the top of the L1 layer carrying the non-mutant copy of *TsBAHD* would give rise to a flower possessing an S-morph phenotype. In contrast, similar cells carrying the mutant copy of *TsBAHD* would give rise to a flower with a mutant long-homostyle phenotype. Additional experiments, such as *in-situ* PCR may aid in studying the location of *TsBAHD* alleles in the periclinal chimera long-homostyle Mhomo-H and its S-morph revertant, Mhomo-S.

5.5.2 Self-compatible S-morph SSC mutant

The S-morph self-compatible mutant SSC possessed a mutation that appears to have abolished its incompatibility phenotype where its pollen can fertilize both L- and S-morph plants. Inheritance studies indicated that the mutation is very closely linked to the *S*-locus. The mutant possessed both stamen-specific candidate genes, *TsSPH1 and TsYUC6*, in its genome. While inheritance studies suggest the mutation is not at the *S*-locus, it would be worth studying the expression and sequence of both genes to discover if they have been altered in this mutant. The cause of the aberrant ratios of S- and L-morph progenies of the SSC mutant throughout the three generations and its relation to the SC mutation have yet to be determined. It seems that progeny carrying the mutation must possess reduced viability. To test this hypothesis,

additional crosses, including the two SI progeny from the third generation, should be performed. If the produced progeny still shows aberrant segregation, then the lethality is likely at the *S*-locus. However, if no aberrant segregation is observed for the two SI progeny, this suggests the lethality is associated with a linked gene or gene region.

6. Conclusions and future work

This study provided for the first time, evidence that three genes determine the S-morph of Turnera species. The finding is parallel to the classical GPA model of distyly in Primula, except that the genes are hemizygous. The three hemizygous S-locus genes, TsSPH1, TsYUC6, and TsBAHD, were identified using chromosome walking and deletion mapping. Expression and sequencing experiments performed on mutants aided in characterizing the possible roles of the genes: that is, TsBAHD is the primary candidate gene for determining the length of the styles and possibly style SI. TsSPH1 is the candidate gene for determining the length of the filaments. TsYUC6 may determine pollen size, number, and/or SI of pollen.

Further experiments are needed to test the above hypotheses. Both knock-out and knock-in experiments would be valuable approaches for determining the roles of the hemizygous genes with certainty. Further experiments are required to explore the reduction of TsSPH1 expression in the short-homostyle mutant, possibly sequencing upstream of TsSPH1 to determine whether there have been mutations in the promoter region of the gene. The gene(s) responsible for the mutant SC phenotype in the SSC mutant plant requires further investigation. Transcriptome sequencing might be valuable in this context since it appears that a “non-S-locus” gene is responsible for the mutation, which might imply that a downstream gene is

responsible for the mutant phenotype. Finally, in-situ PCR might aid in localizing the TsBAHD alleles in the periclinal chimera long-homostyle mutant Mhomo-H and its S-morph revertant (Mhomo-S). This might clarify which particular cells of the style are responsible for generating a long versus short style and the role of TsBAHD, if any, in floral growth and development.

Literature Cited

Arbo, M. M., Gonzalez, A. M., & Sede, S. M. (2015). Phylogenetic relationships within Turneraceae based on morphological characters with emphasis on seed micromorphology. *Plant systematics and evolution*, *301*(7), 1907-1926.

Athanasiou, A., & Shore, J. S. (1997). Morph-specific proteins in pollen and styles of distylous Turnera (Turneraceae). *Genetics*, *146*(2), 669-679.

Athanasiou, A., Khosravi, D., Tamari, F., & Shore, J. S. (2003). Characterization and localization of short-specific polygalacturonase in distylous Turnera subulata (Turneraceae). *American Journal of Botany*, *90*(5), 675-682.

Bateson, W., & Gregory, R. P. (1905). On the inheritance of heterostylism in Primula. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character*, *76*(513), 581-586.

Barrett, S. C. H. (1992). Heterostylous genetic polymorphisms: model systems for evolutionary analysis. In *Evolution and function of heterostyly* (pp. 1-29). Springer, Berlin, Heidelberg.

Barrett, S. C. H. (2019). "A most complex marriage arrangement" Recent advances on heterostyly and unresolved questions. *New Phytologist*. Unpublished manuscript.

Barrett, S. C. H., & Shore, J. S. (2008). New insights on heterostyly: comparative biology, ecology and genetics. In *Self-incompatibility in flowering plants* (pp. 3-32). Springer, Berlin, Heidelberg.

Bayer, A., Ma, X., & Stöckigt, J. (2004). Acetyltransfer in natural product biosynthesis—functional cloning and molecular analysis of vinorine synthase. *Bioorganic & medicinal chemistry*, *12*(10), 2787-2795.

Bennetzen, J. L. (2000). Transposable element contributions to plant gene and genome evolution. *Plant molecular biology*, *42*(1), 251-269.

Burrows, B. A., & McCubbin, A. G. (2017). Sequencing the genomic regions flanking S-linked PvGLO sequences confirms the presence of two GLO loci, one of which lies adjacent to the style-length determinant gene CYP734A50. *Plant reproduction*, *30*(1), 53-67.

Burrows, B., & McCubbin, A. (2018). Examination of S-Locus Regulated Differential Expression in Primula vulgaris Floral Development. *Plants*, *7*(2), 38.

Charlesworth, D. (1982). On the nature of the self-incompatibility locus in homomorphic and heteromorphic systems. *The American Naturalist*, *119*(5), 732-735.

Cecchetti, V., Altamura, M. M., Falasca, G., Costantino, P., & Cardarelli, M. (2008). Auxin regulates Arabidopsis anther dehiscence, pollen maturation, and filament elongation. *The Plant Cell*, *20*(7), 1760-1774.

Cecchetti, V., Celebrin, D., Napoli, N., Ghelli, R., Brunetti, P., Costantino, P., & Cardarelli, M. (2017). An auxin maximum in the middle layer controls stamen development and pollen maturation in Arabidopsis. *New Phytologist*, *213*(3), 1194-1207.

Chen, G., Zhang, B., Liu, L., Li, Q., Zhang, Y. E., Xie, Q., & Xue, Y. (2012). Identification of a ubiquitin-binding structure in the S-locus F-box protein controlling S-RNase-based self-incompatibility. *Journal of Genetics and Genomics*, *39*(2), 93-102.

Cheng, Y., Dai, X., & Zhao, Y. (2006). Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in Arabidopsis. *Genes & development*, *20*(13), 1790-1799.

Choi, S., Cho, Y.H., Kim, K., Matsui, M., Son, S.H., Kim, S.K., Fujioka, S. & Hwang, I. (2013). BAT 1, a putative acyltransferase, modulates brassinosteroid levels in A rabidopsis. *The Plant Journal*, *73*(3), 380-391.

Cocker, J. M., Webster, M. A., Li, J., Wright, J., Kaithakottil, G., Swarbreck, D., & Gilmartin, P. M. (2015). Oakleaf: an S locus-linked mutation of Primula vulgaris that affects leaf and flower development. *New Phytologist*, *208*(1), 149-161.

Cohen, J. I. (2010). “A case to which no parallel exists”: The influence of Darwin's Different Forms of Flowers. *American Journal of Botany*, *97*(5), 701-716.

Darwin, C. (1877). *The different forms of flowers on plants of the same species*. John Murray.

D’Auria, J. C. (2006). Acyltransferases in plants: a good time to be BAHD. *Current opinion in plant biology*, *9*(3), 331-340.

De Graaf, B.H., Vatovec, S., Juárez-Díaz, J.A., Chai, L., Kooblall, K., Wilkins, K.A., Zou, H., Forbes, T., Franklin, F.C.H. & Franklin-Tong, V.E. (2012). The Papaver self-incompatibility pollen S-determinant, PrpS, functions in Arabidopsis thaliana. *Current Biology*, *22*(2), 154-159.

Dulberger, R. (1992). Floral polymorphisms and their functional significance in the heterostylous syndrome. In *Evolution and function of heterostyly*, 41-84. Springer, Berlin, Heidelberg.

Ernst, A. (1955). Self-fertility in monomorphic Primulas. *Genetica*, *27*(1), 391-448.

Franklin-Tong, V. E., Atwal, K. K., Howell, E. C., Lawrence, M. J., & Franklin, F. C. H. (1991). Self-incompatibility in Papaver rhoeas: there is no evidence for the involvement of stigmatic ribonuclease activity. *Plant, Cell & Environment*, *14*(4), 423-429.

Franklin-Tong, V. E., & Franklin, F. C. H. (2003). The different mechanisms of gametophytic self–incompatibility. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1434), 1025-1032.

Foote, H. C., Ride, J. P., Franklin-Tong, V. E., Walker, E. A., Lawrence, M. J., & Franklin, F. C. (1994). Cloning and expression of a distinctive class of self-incompatibility (S) gene from Papaver rhoeas L. *Proceedings of the National Academy of Sciences*, *91*(6), 2265-2269.

Ganders, F. R. (1979). The biology of heterostyly. *New Zealand Journal of Botany*, *17*(4), 607-635.

Gibbs, P. E. (1986). Do homomorphic and heteromorphic self-incompatibility systems have the same sporophytic mechanism?. *Plant Systematics and Evolution*, *154*(3-4), 285-323

Gilmartin, P. M. (2015). On the origins of observations of heterostyly in Primula. *New Phytologist*, *208*(1), 39-51.

Goldraij, A., Kondo, K., Lee, C.B., Hancock, C.N., Sivaguru, M., Vazquez-Santana, S., Kim, S., Phillips, T.E., Cruz-Garcia, F. & McClure, B., (2006). Compartmentalization of S-RNase and HT-B degradation in self-incompatible Nicotiana. *Nature*, *439*(7078), .805.

Heilmann, I. (2016). Phosphoinositide signaling in plant development. *Development*, *143*(12), 2044-2055.

Huu, C.N., Kappel, C., Keller, B., Sicard, A., Takebayashi, Y., Breuninger, H., Nowak, M.D., Baeurle, I., Himmelbach, A., Burkart, M. & Ebbing-Lohaus, T. (2016). Presence versus absence of CYP734A50 underlies the style-length dimorphism in primroses. *Elife*, *5*, 17956.

Indriolo, E., Tharmapalan, P., Wright, S. I., & Goring, D. R. (2012). The ARC1 E3 ligase gene is frequently deleted in self-compatible Brassicaceae species and has a conserved role in Arabidopsis lyrata self-pollen rejection. *The Plant Cell*, *24*(11), 4607-4620.

Kakita, M., Murase, K., Iwano, M., Matsumoto, T., Watanabe, M., Shiba, H., Isogai, A. & Takayama, S. (2007). Two distinct forms of M-locus protein kinase localize to the plasma membrane and interact directly with S-locus receptor kinase to transduce self-incompatibility signaling in Brassica rapa. *The Plant Cell*, *19*(12), 3961-3973.

Kamps-Hughes, N., Quimby, A., Zhu, Z., & Johnson, E. A. (2013). Massively parallel characterization of restriction endonucleases. *Nucleic acids research*, *41*(11), e119-e119.

Kao, T. H., & Tsukamoto, T. (2004). The molecular and genetic bases of S-RNase-based self-incompatibility. *The Plant Cell*, *16*(suppl 1), S72-S83.

Kappel, C., Huu, C. N., & Lenhard, M. (2017). A short story gets longer: recent insights into the molecular basis of heterostyly.

Khosravi, D., Joulaie, R., & Shore, J. S. (2003). Immunocytochemical distribution of polygalacturonase and pectins in styles of distylous and homostylous Turneraceae. *Sexual Plant Reproduction*, *16*(4), 179-190.

Khosravi, D., Yang, E. C., Siu, K. M., & Shore, J. S. (2004). High level of α-dioxygenase in short styles of distylous Turnera species. *International Journal of Plant Sciences*, *165*(6), 995-1006.

Kubo K. Entani T., Takara A., Wang N., Fields A.M., Hua Z., Toyoda M., Kawashima S., Ando T., Isogai A., Kao T., & Takayama S. (2010). Collaborative non-self recognition system in S-RNase-based self-incompatibility. *Science*. *330*(6005), 796-799.

Kubo K., Paape T., Hatakeyama M., Entani T., Takara A., Kajihara K., Tsukhara M., Shimizu-Inatsugi R., Shimizu K.K., & Takayama S. (2015). Gene duplication and genetic exchange drive the evolution of S-RNase-based self-incompatibility in Petunia. *The Plant Journal*. (1), 1-9.

Kurian, V., & Richards, A. J. (1997). A new recombinant in the heteromorphy 'S'supergene in Primula. *Heredity*, *78*(4), 383.

Labonne, J. J., Goultiaeva, A., & Shore, J. S. (2009). High-resolution mapping of the S-locus in Turnera leads to the discovery of three genes tightly associated with the S-alleles. *Molecular Genetics and Genomics*, *281*(6), 673.

Labonne, J. D. J., Tamari, F., & Shore, J. S. (2010). Characterization of X-ray-generated floral mutants carrying deletions at the S-locus of distylous Turnera subulata. *Heredity*, *105*(2), 235.

Labonne, J. D., & Shore, J. S. (2011). Positional cloning of the s haplotype determining the floral and incompatibility phenotype of the long-styled morph of distylous Turnera subulata. *Molecular Genetics and Genomics*, *285*(2), 101-111.

Lai, Z., Ma, W., Han, B., Liang, L., Zhang, Y., Hong, G., & Xue, Y. (2002). An F-box gene linked to the self-incompatibility (S) locus of Antirrhinum is expressed specifically in pollen and tapetum. *Plant molecular biology*, *50*(1), 29-41.

Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M. & Karthikeyan, A.S. (2011). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic acids research*, *40*(D1), D1202-D1210.

Lewis D. & Jones D.A. (1992). The genetics of heterostyly. In: Barrett S.C.H. (ed.) Evolution and Function of Heterostyly. Springer, Berlin, Heidelberg, 129-150.

Li, J., Webster, M., Dudas, B., Cook, H., Manfield, I., Davies, B., & Gilmartin, P. M. (2008). The S locus-linked Primula homeotic mutant sepaloid shows characteristics of a B-function mutant but does not result from mutation in a B-function gene. *The Plant Journal*, *56*(1), 1-12.

Li, J., Webster, M.A., Wright, J., Cocker, J.M., Smith, M.C., Badakshi, F., Heslop-Harrison, P. & Gilmartin, P.M. (2015). Integration of genetic and physical maps of the Primula vulgaris S locus and localization by chromosome in situ hybridization. *New Phytologist*, *208*(1), 137-148.

Li, J., Cocker, J.M., Wright, J., Webster, M.A., McMullan, M., Dyer, S., Swarbreck, D., Caccamo, M., van Oosterhout, C. & Gilmartin, P.M. (2016). Genetic architecture and evolution of the S locus supergene in Primula vulgaris. *Nature plants*, *2*(12), 16188.

Lloyd, D. G., & Webb, C. J. (1992). The evolution of heterostyly. In *Evolution and function of heterostyly* (pp. 151-178). Springer, Berlin, Heidelberg.

Ma, X., Koepke, J., Panjikar, S., Fritzsch, G., & Stöckigt, J. (2005). Crystal structure of vinorine synthase, the first representative of the BAHD superfamily. *Journal of Biological Chemistry*, *280*(14), 13576-13583.

Manzanares, C., Barth, S., Thorogood, D., Byrne, S.L., Yates, S., Czaban, A., Asp, T., Yang, B. & Studer, B. (2015). A gene encoding a DUF247 domain protein cosegregates with the S self-incompatibility locus in perennial ryegrass. *Molecular biology and evolution*, *33*(4), 870-884.

Matton, D. P., Nass, N., Clarke, A. E., & Newbigin, E. (1994). Self-incompatibility: how plants avoid illegitimate offspring. *Proceedings of the National Academy of Sciences*, *91*(6), 1992-1997.

Matsui, K., Nishio, T., & Tetsuka, T. (2004). Genes outside the S supergene suppress S functions in buckwheat (Fagopyrum esculentum). *Annals of Botany*, *94*(6), 805-809.

McClure, B. A., Haring, V., Ebert, P. R., Anderson, M. A., Simpson, R. J., Sakiyama, F., & Clarke, A. E. (1989). Style self-incompatibility gene products of Nicotlana alata are ribonucleases. *Nature*, *342*(6252), 955.

McClure B.A. (2010). Darwin's foundation for investigating self-incompatibility and the progress toward a physiological model for S-RNase-based SI. Journal of Experimental Botany. 60(4): 1069-1081.

McClure, B., Cruz-García, F., & Romero, C. (2011). Compatibility and incompatibility in S-RNase-based systems. *Annals of botany*, *108*(4), 647-658.

McDill, J., Repplinger, M., Simpson, B. B., & Kadereit, J. W. (2009). The phylogeny of Linum and Linaceae subfamily Linoideae, with implications for their systematics, biogeography, and evolution of heterostyly. *Systematic Botany*, *34*(2), 386-405.

Nasrallah, J. B. (2011). Self-incompatibility in the Brassicaceae. In *Genetics and Genomics of the Brassicaceae*(pp. 389-411). Springer, New York, NY.

Nasrallah & Nasrallah. (1989). The molecular genetics of self-incompatibility in Brassica. *Annual Review of Genetics*, 23, 121-139.

Nasrallah, J. B., & Nasrallah, M. E. (2014). S-locus receptor kinase signalling.*Biochemical Society Transactions*. 42(2), 313-319.

Neff, M.M., Nguyen, S.M., Malancharuvil, E.J., Fujioka, S., Noguchi, T., Seto, H., Tsubuki, M., Honda, T., Takatsuto, S., Yoshida, S. & Chory, J. (1999). BAS1: A gene regulating brassinosteroid levels and light responsiveness in Arabidopsis. *Proceedings of the National Academy of Sciences*, *96*(26), 15316-15323.

Newbigin, E., Anderson, M. A., & Clarke, A. E. (1993). Gametophytic self-incompatibility systems. *The Plant Cell*, *5*(10), 1315-1324.

Newbigin, E., Paape, T., & Kohn, J. R. (2008). RNase-based self-incompatibility: puzzled by pollen S. *The Plant Cell*, *20*(9), 2286-2292.

Nowak, M. D., Russo, G., Schlapbach, R., Huu, C. N., Lenhard, M., & Conti, E. (2015). The draft genome of Primula veris yields insights into the molecular basis of heterostyly. *Genome biology*, *16*(1), 12.

Peng, J., & Xu, J. (2011). RaptorX: exploiting structure information for protein alignment by statistical inference. *Proteins: Structure, Function, and Bioinformatics*, *79*(S10), 161-171.

Pribat, A., Sormani, R., Rousseau-Gueutin, M., Julkowska, M.M., Testerink, C., Joubès, J., Castroviejo, M., Laguerre, M., Meyer, C., Germain, V. & Rothan, C. (2012). A novel class of PTEN protein in Arabidopsis displays unusual phosphoinositide phosphatase activity and efficiently binds phosphatidic acid. *Biochemical Journal*, *441*(1), 161-171.

Ramachandran, V., & Chen, X. (2008). Degradation of microRNAs by a family of exoribonucleases in Arabidopsis. *Science*, *321*(5895), 1490-1492.

Richards, J. (2003). Primula. Portland, Oregon: Timber Press.

Ride, J. P., Davies, E. M., Franklin, F. C. H., & Marshall, D. F. (1999). Analysis of Arabidopsis genome sequence reveals a large new gene family in plants. *Plant molecular biology*, *39*(5), 927-932.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A., & Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics*, *16*(10), 944-945.

Samuel, M. A., Chong, Y. T., Haasen, K. E., Aldea-Brydges, M. G., Stone, S. L., & Goring, D. R. (2009). Cellular pathways regulating responses to compatible and self-incompatible pollen in Brassica and Arabidopsis stigmas intersect at Exo70A1, a putative component of the exocyst complex. *The Plant Cell*, *21*(9), 2655-2671.

Safavian, D., & Shore, J. S. (2010). Structure of styles and pollen tubes of distylous Turnera joelii and T. scabra (Turneraceae): are there different mechanisms of incompatibility between the morphs?. *Sexual plant reproduction*, *23*(3), 225-237.

Sievers, F., & Higgins, D. G. (2018). Clustal Omega for making accurate alignments of many protein sequences. *Protein Science*, *27*(1), 135-145.

Sijacic, P., Wang, X., Skirpan, A.L., Wang, Y., Dowd, P.E., McCubbin, A.G., Huang, S. & Kao, T.H. (2004). Identification of the pollen determinant of S-RNase-mediated self-incompatibility. *Nature*, *429*(6989), 302.

Schopfer, C. R., Nasrallah, M. E., & Nasrallah, J. B. (1999). The male determinant of self-incompatibility in Brassica. *Science*, *286*(5445), 1697-1700.

Schwander T, Libbrecht R, & Keller L. (2014). Supergenes and complex phenotypes. *Current Biology, 24*, 288-294.

Schwager, K. M., Calderon-Villalobos, L. I. A., Dohmann, E. M., Willige, B. C., Knierer, S., Nill, C., & Schwechheimer, C. (2007). Characterization of the VIER F-BOX PROTEINE genes from Arabidopsis reveals their importance for plant growth and development. *The Plant Cell*, *19*(4), 1163-1178.

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature methods*, *9*(7), 671.

Sharma, K. D., & Boyes, J. W. (1961). Modified incompatibility of buckwheat following irradiation. *Canadian Journal of Botany*, *39*(5), 1241-1246.

Shore, J. S., & Barrett, S. C. (1985). Morphological differentiation and crossability among populations of the Turnera ulminfolia L. complex (Turneraceae). *Systematic Botany*, 308-321.

Shore, J. S., Arbo, M. M., & Fernández, A. (2006). Breeding system variation, genetics and evolution in the Turneraceae. *New Phytologist*, *171*(3), 539-551.

Shore, J. S., Hamam, H. J., Labonne, J. D., Henning, P., & McCubbin, A. (2019). The long and short of the *S*-locus in *Turnera* (Passifloraceae). *New Phytologist*. DOI:10.1111/nph.15970

Stevens, V. A. M., & Murray, B. G. (1982). Studies on heteromorphic self-incompatibility systems: Physiological aspects of the incompatibility system of Primula obconica. *Theoretical and Applied Genetics*, *61*(3), 245-256.

St-Pierre, B., & De Luca, V. (2000). Origin and diversification of the BAHD superfamily of acyltransferases involved in secondary metabolism. *Recent Adv. Phytochem*, *34*, 285-315.

Suzuki, H., Nakayama, T., & Nishino, T. (2003). Proposed Mechanism and Functional Amino Acid Residues of Malonyl-CoA: Anthocyanin 5-O-Glucoside-6 ‘’‘-O-Malonyltransferase from Flowers of Salvia splendens, a Member of the Versatile Plant Acyltransferase Family. *Biochemistry*, *42*(6), 1764-1771.

Suzuki, H., Nishino, T., & Nakayama, T. (2007). cDNA cloning of a BAHD acyltransferase from soybean (Glycine max): isoflavone 7-O-glucoside-6 ”-O-malonyltransferase. *Phytochemistry*, *68*(15), 2035-2042.

Taguchi-Shiobara, F., Yuan, Z., Hake, S., & Jackson, D. (2001). The fasciated ear2 gene encodes a leucine-rich repeat receptor-like protein that regulates shoot meristem proliferation in maize. *Genes & Development*, *15*(20), 2755-2766.

Takasaki, T., Hatakeyama, K., Suzuki, G., Watanabe, M., Isogai, A., & Hinata, K. (2000). The S receptor kinase determines self-incompatibility in Brassica stigma. *Nature*, *403*(6772), 913.

Takayama, S., Shimosato, H., Shiba, H., Funato, M., Che, F.S., Watanabe, M., Iwano, M. & Isogai, A. (2001). Direct ligand–receptor complex interaction controls Brassica self-incompatibility. *Nature*, *413*(6855), 534.

Takayama S. & Isogai A. (2005).Self-incompatibility in plants. Annual Review of Plant Biology. 56, 467-489.

Takeshima, R., Nishio, T., Komatsu, S., Kurauchi, N., & Matsui, K. (2019). Identification of a gene encoding polygalacturonase expressed specifically in short styles in distylous common buckwheat (Fagopyrum esculentum). *Heredity*, 1.

Tamari, F., & Shore, J. S. (2004). Distribution of style and pollen polygalacturonases among distylous and homostylous Turnera and Piriqueta spp.(Turneraceae). *Heredity*, *92*(5), 380.

Tamari, F., Khosravi, D., Hilliker, A. J., & Shore, J. S. (2005). Inheritance of spontaneous mutant homostyles in Turnera subulata× krapovickasii and in autotetraploid T. scabra (Turneraceae). *Heredity*, *94*(2), 207.

Tamari, F., & Shore, J. S. (2006). Allelic variation for a short-specific polygalacturonase in Turnera subulata: is it associated with the degree of self-compatibility?. *International journal of plant sciences*, *167*(1), 125-133.

Thorogood, D., Yates, S., Manzanares, C., Skot, L., Hegarty, M., Blackmore, T., Barth, S. & Studer, B. (2017). A novel multivariate approach to phenotyping and association mapping of multi-locus gametophytic self-incompatibility reveals S, Z, and other loci in a perennial ryegrass (Poaceae) population. *Frontiers in plant science*, *8*, 1331.

Tian, H., Lv, B., Ding, T., Bai, M., & Ding, Z. (2018). Auxin-BR interaction regulates plant growth and development. *Frontiers in plant science*, *8*, 2256.

Tuominen, L. K., Johnson, V. E., & Tsai, C. J. (2011). Differential phylogenetic expansions in BAHD acyltransferases across five angiosperm taxa and evidence of divergent expression among Populus paralogues. *BMC genomics*, *12*(1), 236.

Ueno, M., Yasui, Y., Aii, J., Matsui, K., Sato, S., & Ota, T. (2016). Genetic Analyses of the Heteromorphic Self-Incompatibility (*S*) Locus in Buckwheat. Molecular Breeding and Nutritional Aspects of Buckwheat, 411.

Unno, H., Ichimaida, F., Suzuki, H., Takahashi, S., Tanaka, Y., Saito, A., Nishino, T., Kusunoki, M. and Nakayama, T. (2007). Structural and mutational studies of anthocyanin malonyltransferases establish the features of BAHD enzyme catalysis. *Journal of Biological Chemistry*, *282*(21), 15812-15822.

Ushijima, K., Nakano, R., Bando, M., Shigezane, Y., Ikeda, K., Namba, Y., Kume, S., Kitabata, T., Mori, H. & Kubo, Y. (2012). Isolation of the floral morph-related genes in heterostylous flax (Linum grandiflorum): the genetic polymorphism and the transcriptional and post-transcriptional regulations of the S locus. *The Plant Journal*, *69*(2), 317-331.

Uyenoyama, M. K. (2005). Evolution under tight linkage to mating type. *New phytologist*, *165*(1), 63-70.

Wang, L., Dong, L., Zhang, Y. E., Zhang, Y., Wu, W., Deng, X., & Xue, Y. (2004). Genome-wide analysis of S-Locus F-box-like genes in Arabidopsis thaliana. *Plant molecular biology*, *56*(6), 929-945.

Wedderburn, F., & Richards, A. J. (1990). Variation in within-morph incompatibility inhibition sites in heteromorphic Primula L. *New Phytologist*, *116*(1), 149-162.

Wheeler MJ, de Graaf BH, Hadjiosif N, Perry RM, Poulter NS, *et al*. (2009) Identification of the pollen self-incompatibility determinant in Papaver rhoeas. Nature 459, 992–995.

Wheeler, M. J., Vatovec, S., & Franklin-Tong, V. E. (2010). The pollen S-determinant in Papaver: comparisons with known plant receptors and protein ligand partners. *Journal of experimental botany*, *61*(7), 2015-2025.

Wu, J., Wang, S., Gu, Y., Zhang, S., Publicover, S. J., & Franklin-Tong, V. E. (2011). Self-incompatibility in Papaver rhoeas activates nonspecific cation conductance permeable to Ca2+ and K+. *Plant physiology*, *155*(2), 963-973.

Yao, X., Tian, L., Yang, J., Zhao, Y.N., Zhu, Y.X., Dai, X., Zhao, Y. & Yang, Z.N. (2018). Auxin production in diploid microsporocytes is necessary and sufficient for early stages of pollen development. *PLoS genetics*, *14*(5), 1007397.

Yasui, Y., Mori, M., Aii, J., Abe, T., Matsumoto, D., Sato, S., Hayashi, Y., Ohnishi, O. & Ota, T., (2012). S-LOCUS EARLY FLOWERING 3 is exclusively present in the genomes of short-styled buckwheat plants that exhibit heteromorphic self-incompatibility. *PloS one*, *7*(2), 1264.

Zhang Z., Schwartz S., Wagner L., & Miller W. (2000), "A greedy algorithm for aligning DNA sequences" J Comput Biol 2000; 7(1-2):203-14

Zhang, Y., Zhao, Z., & Xue, Y. (2009). Roles of proteolysis in plant self-incompatibility. *Annual review of plant biology*, *60*, 21-42.

Appendices

Appendix A: PCR and sequencing primers

Table S1. List of primers used for amplifying genes of interest. (+) indicates the forward orientation and (–) indicates the reverse orientation. The product size for each primer pair used was calculated based on *Turnera subulata* species. The Application column refers to the experiment the primers were used for: Seq. = Sequencing, Map = Deletion Mapping, RT-PCR = Reverse Transcription Poly Chain Reaction, PCR = Poly Chain Reaction, Sp.Sur.= Species survey for the three candidate genes.

Gene name	Primer name	Sequence	Orientation	Size	Application
TsSPH1	SSTAUNIV1F	TCAGCCAATAGCTCCTTGAAGTT	+	369 bp	Map,Sp.Sur.
	SSTAUNIV1R	CCAGTCATACAACTTGAAATAGG	-		
	Asph1F	CTACCGGACGGGAGCTT	+	374 bp	RT-PCR
	Asph1R	AGCAATGCTATTGGCGG	-		
TsBAHD	HxxxD_ExonF	ATGGAAGTTGAGATCACGCTG	+	1210 bp	Map
	HxxxD_ExonR	ATGCTTCTATTCCGGCAGCA	-		
	16HxxxD2R	TTCCAGATGAGAGCGGACAC	-	Binds at 752 bp	RT-PCR
	HxxxD1500_1F	GTGTCCGCTCTCATCTGGAA	+	Binds at 731 bp	Seq,Sp.Sur.
	HxxxD500R	CGATCCGCTCCAGCATTTGA	-	Binds at 502 bp	Seq
	HxxxD1000F	GCATGGCGATTGGTGTTTGT	+	Binds at 439 bp	Seq
	HxxxD1000R	TTCTACCTCGGCATCCTCCA	-	Binds at 859 bp	Seq
	HxxxD1500_2F	AGATACAAGGGATGCTGCCG	+	Binds at 1178 bp	Seq
	HxxxDupdownF	AGCCTCCAGCAACAAAGTAA	+	1562 bp	Seq
	HxxxDupdownR	ACGGGGCTATCGGAGAAAGT	-		
	HxxxD 1F	CTGGGAGCGCATAGAAAGTC	+	242 bp	Map,Sp.Sur.
	HxxxD 1R	GCAAACCCACAAATTGCTTT	-		
	AbahdF	ACCCTCTAGCAGGCAGACTC	+	398 bp	RT-PCR

	AbahdR	CAACTCAACTGAGGAGGGTCT	-		
TsYUC6	YUC6_Ex1F	CTTCGCGCTGGGTATATGCT	+	615 bp	PCR
	YUC6_Ex1R	CTTTTACCACAAGCGAAGGCT	-		
	YUC6_Ex2F	GCATGTTTTACCACAAGAGATGCT	+	240 bp	PCR
	YUC6_Ex2R	TCCCCACTTTTGATCTTGGCT	-		
	YUC6_Ex34iF	ATTTGTGGAAGCATCAAAAGGCT	+	653 bp	Map, Sp.Sur.
	YUC6_Ex34iR	CATTCCATTGCCTGATTGCAT	-		
	AyucF	TGTACCGTCATGGCTAAAGGT	+	581 bp	RT-PCR
	AyucR	CAACATTGTGTTAGCTACAGGTCA	-		
TsAP2	ApetellaF	GTGGGACTTGAACGACTCTCC	+	260 bp	Map
	ApetellaR	CACTGTCCTCCTCATAGGGC	-		
TsERF	AP2L22F1	AACTCACCACCACCTGCTTC	+	283 bp	Map
	AP2L22R4	CTCGACCTTTTGTTGGGTTC	-		
TsCBSX	LEJ1L22F14	GACCACTGTTTCTCAGTCCCATA	+	309 bp	Map
	LEJ1L22R9	TGACCTGTTCCCTGAAGTTGA	-		
TsSKIP2	Skip_4F	CCAGTCTACATCCACCGCC	+	286 bp	Map
	Skip_2R	TTTGGTGACCGACTCGAAG	-		
TsSDN3	Exo_1iF	AAGACTTCAGATGGAAAGACACC	+	253 bp	Map
	Exo_1iR	CAGGCTAGACGTGGGAGGAA	-		
TsPTEN	Phosphad 3F	CTTGGACTGCGGTTCCCTTC	+	175 bp	Map
	Phosphad 3R	GCTTATTCTGCGAAACGGCG	-		
TsDUF247	Duf247_2F	CTGCGGAACATAGGCCTTGT	+	240 bp	Map

	Duf247_2R	GGCTCTTCTTGGGGCACTTC	-		
TsDIOX2	FSPF1	GGTTCACCCGATGAGTTTGA	+	236 bp	Map
	FSPR1	CAGTATGAGCTCCGCTGTGA	-		
TsDUS1L	tRNADSF1	GGGACGAATACTGCAGAGGA	+	512 bp	Map
	tRNADSR4	AATTATCCACCATGGGAGCA	-		
TsLRK	RCKL22F2	CGCTGCTCTCTTTCAAGTCC	+	255 bp	Map
	RCKL22R3	AGGTTGGAGAGAGGGGAGAG	-		
TsADPRF	ARF_F	ATCTCTTGTTCACCACTTGCCT	+	247 bp	Map
	ARF_R	AGCCTCAGATAAATGATGACTGAAA	-		
β-*Tubulin*	TUB1F	CAGCTGGAAAGGATCAATGTTTA	+	900 bp	Map,Sp.Sur.
	TUB1R	GTTCTTGGCATCCCACATTT	-		

Appendix B: Genes on the *S*-locus haplotypes of *T. subulata*

Table S2. Genes on the *S*-locus haplotypes of *T. subulata*. For each gene, its *Arabidopsis thaliana* homolog, as well as its function, is listed.

Code	*A.thaliana* Homolog	Function
TsAP2	AT5G67180	Floral homeotic protein APETALA 2
TsCBSX	AT4G36910	CBS domain-containing protein
TsERF	AT5G67190	Ethylene-responsive transcription factor
TsRPB	AT2G42240	RNA-binding protein
TsSUMO	AT3G57870	Sumo conjugating enzyme 1A
TsFRA1	AT5G47820	Kinesin-like protein
TsLRK	At5g67200	Leucine-rich repeat protein kinase family protein
TsIRX15L	AT5G67210	Protein essential for normal xylan synthesis
TsDIOX2	AT3G49630	2-oxoglutarate-dependent dioxygenase
TsDUS1L	AT5G67220	tRNA dihydrouridine synthase activity
TsWRKY7	AT4G24240	WRKY transcription factor 7
TsIRX14H	AT5G67230	Beta-1,4-xylosyltransferase
TsGCP1	AT4G36880	Germination-specific cysteine protease 1
TsDUF247	AT3G50120	Transmembrane protein-containing DUF247
TsRER1A	AT4G39220	Retrieval of early ER protein
TsPTEN	AT3G19420	Protein-tyrosine-phosphatase
TsSDN3	AT5G67240	Small RNA degrading nuclease 3
TsSKIP2	AT5G67250	F-box protein
TsCYCD	AT5G67260	Cyclin-dependent protein kinase, cell proliferation
TsADPRF	AT1G10630	ADP-ribosylation factor 1
TsEB1C	AT3G47690	Microtubule-associated protein RP/EB 1C
***S*-locus specific genes**		

TsSPH1	AT4G16295	Self-incompatibility (S) protein homolog of *Papaver somniferum*
TsYUC6	AT5G25620	Flavin monooxygenase with an important role in auxin biosynthesis
TsBAHD	AT3G26040	BAHD acyltransferase

Appendix C: Alignment of genomic and amino acid sequences of the candidate genes

Figure S1. *TsBAHD* genomic sequence from two long-homostyle mutant and the S-morph revertant Mhomo-S. PCR purified DNA (check methods) were used to sequence *TsBAHD* using specific primers (Table S1). *TsBAHD* was sequenced using Sanger sequencing (McGill University and Genome Quebec Innovation Centre). MhomoS refers to the S-morph revertant Mhomo-S. 734H is a heterozygote long-homostyle mutant progeny of Mhomo-H; HH12 is a homozygote long-homostyle progeny of Mhomo-H. Sequences were aligned using Clustal Omega (Sievers and Higgins, 2018).

```
MhomoS      -ATATGGAAGTTGAGATCACGCTGAGAGAAACTATTAAGCCATCTTCTTCTACACCCCCT    59
734H        -ATATGGAAGTTGAGATCACGCTGAGAGAAACTATTAAGCCATCTTCTTCTACACCCCCT    59
HH12        GATATGGAAGTTGAGATCACGCTGAGAGAAACTATTAAGCCATCTTCTTCTACACCCCCT    60
             ************************************************************

MhomoS      AGCAAAAGAATTCTCAAGCTTTCTCTCATGGACCAGTTCACTCCTGTGTGTTACACGTCT    119
734H        AGCAAAAGAATTCTCAAGCTTTCTCTCATGGACCAGTTCACTCCTGTGTGTTACACGTCT    119
HH12        AGCAAAAGAATTCTCAAGCTTTCTCTCATGGACCAGTTCACTCCTGTGTGTTACACGTCT    120
            ************************************************************

MhomoS      CTCATCCTTTTTTATCCTGCAAGTGTAAACCAAGATCACTCCGTAACACTGACAGAAAGA    179
734H        CTCATCCTTTTTTATCCTGCAAGTGTAAACCAAGATCACTCCGTAACACTGACAGAAAGA    179
HH12        CTCATCCTTTTTTATCCTGCAAGTGTAAACCAAGATCACTCCGTAACACTGACAGAAAGA    180
            ************************************************************

MhomoS      TGCCAGCAACTGAAAATTTCTCTGTCAGAAACATTAACTCACTTTTACCCTCTAGCAGGC    239
734H        TGCCAGCAACTGAAAATTTCTCTGTCAGAAACATTAACTCACTTTTACCCTCTAGCAGGC    239
HH12        TGCCAGCAACTGAAAATTTCTCTGTCAGAAACATTAACTCACTTTTACCCTCTAGCAGGC    240
            ************************************************************

MhomoS      AGACTCAAAGACAATGCCTCCATAGAATGTGACGACCAAGGAGGTGAGTACATTGAAGCT    299
734H        AGACTCAAAGACAATGCCTCCATAGAATGTGACGACCAAGGAGGTGAGTACATTGAAGCT    299
HH12        AGACTCAAAGACAATGCCTCCATAGAATGTGACGACCAAGGAGGTGAGTACATTGAAGCT    300
            ************************************************************

MhomoS      CGGATCAAATGTCTTCTCTCAGAGTTTCTTGTAACGCCCGAGGCAGAGTTGCTGAAACAG    359
734H        CGGATCAAATGTCTTCTCTCAGAGTTTCTTGTAACGCCCGAGGCAGAGTTGCTGAAACAG    359
HH12        CGGATCAAATGTCTTCTCTCAGAGTTTCTTGTAACGCCCGAGGCAGAGTTGCTGAAACAG    360
            ************************************************************

MhomoS      CTCCTGCCTGCAGCTATAGAATCAGGCGAAGCGGCCACCGGAAGCATGTTGCTTGTCCAA    419
734H        CTCCTGCCTGCAGCTATAGAATCAGGCGAAGCGGCCACCGGAAGCATGTTGCTTGTCCAA    419
HH12        CTCCTGCCTGCAGCTATAGAATCAGGCGAAGCGGCCACCGGAAGCATGTTGCTTGTCCAA    420
            ************************************************************
MhomoS      GCCAATATCTTTGATTGTGGTGGCATGGCGATTGGTGTTTGTGTTTCACATAAGATCGCT    479
734H        GCCAATATCTTTGATTGTGGTGGCATGGCGATTGGTGTTTGTGTTTCACATAAGATCGCT    479
HH12        GCCAATATCTTTGATTGTGGTGGCATGGCGATTGGTGTTTGTGTTTCACATAAGATCGCT    480
            ************************************************************

MhomoS      GATGCGGCAACGTTAACTACATTCGTCAAATGCTGGAGCGGATCGGCTCATAAATCCACA    539
734H        GATGCGGCAACGTTRACTACATTCRTCAAATGCTGGAGCGGATCGGCTCATAAATCCACA    539
HH12        GATGCGGCAACGTTAACTACATTCGTCAAATGCTGGAGCGGATCGGCTCATAAATCCACA    540
```

```
              ************** ********* **********************************

MhomoS        GAAGTAATGAGCCCTCTATTCATGGGTGCTTCCATTTTTCCGGACATGGATTTGACATTC     599
734H          GAAGTAATGAGCCCTCTATTCATGGGTGCTTCCATTTTTCCGGACATGGATTTGACATTC     599
HH12          GAAGTAATGAGCCCTCTATTCATGGGTGCTTCCATTTTTCCGGACATGGATTTGACATTC     600
              ************************************************************


MhomoS        ACCAGACCCTCCTCAGTTGAGTTGATGCAAGGAGAATGTGCAACGAAAAGGTTTGTTTTC     659
734H          ACCAGACCCTCCTCAGTTGAGTTGATGCAAGGAGAATGTGCAACGAAAAGGTTTGTTTTC     659
HH12          ACCAGACCCTCCTCAGTTGAGTTGATGCAAGGAGAATGTGCAACGAAAAGGTTTGTTTTC     660
              ************************************************************


MhomoS        GATTCCGC-AAGATACGCATGCTCAAAGCGAAAGCAGCAAGCGAAAGTGTGGCAGCGCCT     718
734H          GATTCCGCAAAGATACGCATGCTCAAAGCGAAAGCAGCAAGCGAAAGTGTGGCAGCGCCT     719
HH12          GATTCCGCAAAGATACGCATGCTCAAAGCGAAAGCAGCAAGCGAAAGTGTGGCAGCGCCT     720
              ******** ********************************************************


MhomoS        ACCCGGGTGGAAGCCGTGTCCGCTCTCATCTGGAAATGTGCAATGGCGGCATCAAGATCC     778
734H          ACCCGGGTGGAAGCCGTGTCCGCTCTCATCTGGAAATGTGCAATGGCGGCATCAAGATCC     779
HH12          ACCCGGGTGGAAGCCGTGTCCGCTCTCATCTGGAAATGTGCAATGGCGGCATCAAGATCC     780
              ************************************************************


MhomoS        AACCTGGGAGCGCATAGAAAGTCCGTGTGGTCTCTTAGCGTCAATTTGCGCAAAAGGCTT     838
734H          AACCTGGGAGCGCATAGAAAGTCCGTGTGGTCTCTTAGCGTCAATTTGCGCAAAAGGCTT     839
HH12          AACCTGGGAGCGCATAGAAAGTCCGTGTGGTCTCTTAGCGTCAATTTGCGCAAAAGGCTT     840
              ************************************************************


MhomoS        GCGCCGCCATTGCCAGAGAATTATGCTGGCAATTGTGTTGGCTCCATCCTAGCACATATG     898
734H          GCGCCGCCATTGCCAGAGAATTATGCTGGCAGTTGTGTTGGCTCCATCCTAGCACATATG     899
HH12          GCGCCGCCATTGCCAGAGAATTATGCTGGCAGTTGTGTTGGCTCCATCCTAGCACATATG     900
              ******************************** ***************************


MhomoS        GAGGATGCCGAGGTAGAATTGCGAGACTTAGTGGGTCTGATAAGAAGAGGGAACCAGGAA     958
734H          GAGGATGCCGAGGTAGAATTGCGAGACTTAGTGGGTCTGATAAGAAGAGGGAACCAGGAA     959
HH12          GAGGATGCCGAGGTAGAATTGCGAGACTTAGTGGGTCTGATAAGAAGAGGGAACCAGGAA     960
              ************************************************************


MhomoS        TTTGATGAAAATTACGCGAAGAAACTTCAACGAGAGGGGTCTTCGAAAGCAATTTGTGGA    1018
734H          TTTGATGAAAATTACGCGAAGAAACTTCAACGAGAGGGGTCTTCGAAAGCAATTTGTGGA    1019
HH12          TTTGATGAAAATTACGCGAAGAAACTTCAACGAGAGGGGTCTTCGAAAGCAATTTGTGGA    1020
              ************************************************************


MhomoS        TTTGCCAAACAATTTGAAGAGTTGGCTATGAGTACTGACATTGATTTTTATAACTGCACC    1078
734H          TTTGCCAAACAATTTGAAGAGTTGGCTATGAGTACTGACATTGATTTTTATAACTGCACC    1079
HH12          TTTGCCAAACAATTTGAAGAGTTGGCTATGAGTACTGACATTGATTTTTATAACTGCACC    1080
              ************************************************************


MhomoS        AGCTGGTGTAAATTTCAACTGTACGAGGCTGATTTTGGATGGGGGAAGCCAGCATGGTTG    1138
734H          AGCTGGTGTAAATTTCAACTGTACGAGGCTGATTTTGGATGGGGGAAGCCAGCATGGTTG    1139
HH12          AGCTGGTGTAAATTTCAACTGTACGAGGCTGATTTTGGATGGGGGAAGCCAGCATGGTTG    1140
              ************************************************************


MhomoS        AGCATAGCGAGTACATCAATAAAGAATGTGCTCTTGCTGATAGATACAAGGGATGCTGCC    1198
734H          AGCATAGCGAGTACATCAATAAAGAATGTGCTCTTGCTGATAGATACAAGGGATGCTGCC    1199
HH12          AGCATAGCGAGTACATCAATAAAGAATGTGCTCTTGCTGATAGATACAAGGGATGCTGCC    1200
              ************************************************************


MhomoS        GGAATAGAAGCATGGTTGACGTTGAGTGAAGAGGATATGGCCTTTTTTGAATGTAATAAA    1258
734H          GGAATAGAAGCATGGTTGACGTTGAGTGAAGAGGATATGGCCTTTTTTGAATGTAATAAA    1259
HH12          GGAATAGAAGCATGGTTGACGTTGAGTGAAGAGGATATGGCCTTTTTTGAATGTAATAAA    1260
              ************************************************************
```

```
MhomoS      GAGCTGCTAGAATTTGCAACGCTAAACCCTAGTGTAT-TCTGAGCGAACTCCATCTTTGA    1317
734H        GAGCTGCTAGAATTTGCAACGCTAAACCCTAGTGTATCTCTGAGCGAACTCCATCTTTGA    1319
HH12        GAGCTGCTAGAATTTGCAACGCTAAACCCTAGTGTATCTCTGAGCGAACTCCATCTTTGA    1320
            *********************************** ********************


MhomoS      CAGAATCATGCTTTTGCAATCCGGTCAATAGTCACCCAATCGCCGGGTAAACGGAAATTC    1377
734H        CAGAATCATGCTTTTGCAATCCGGTCAATAGTCACCCAATCGCCGGGTAAACGGAAATTC    1379
HH12        CAGAATCATGCTTTTGCAATCCGGTCAATAGTCACCCAATCGCCGGGTAAACGGAAATTC    1380
            ************************************************************


MhomoS      CAATTCCTGCCTTGGAATTTCCCGATTTCCTAGAGGTAAAGTATTCCGAAATAGAATTTA    1437
734H        CAATTCCTGCCTTGGAATTTCCCGATTTCCTAGAGGTAAAGTATTCCGAAATAGAATTTA    1439
HH12        CAATTCCTGCCTTGGAATTTCCCGATTTCCTAGAGGTAAAGTATTCCGAAATAGAATTTA    1440
            ************************************************************

MhomoS      ATCAATGCTTCAAACTCATA----------------------------------------    1457
734H        ATCAATGCTTCAAACTCATATTATCTATGGCCTACAGTAGCCACTGCCACCTCGGAGTCT    1499
HH12        ATCAATGCTTCAAACTCATATTATCTATGGCCTACAGTAGCCACTGCCACCTCGGAGTCT    1500
            ********************

MhomoS      --------------------  1457
734H        ATGCAGTTCGGCATGAACTT  1519
HH12        ATGCAGTTCGGCATGAACTT  1520
```

Figure S2. Inferred amino acid alignments of the *TsSPH1* gene of *T. subulata* against homologues from *A. thaliana*, *Populus trichocarpa,* and *Papaver rhoeas*. This figure was taken from Shore *et al*. (2019). Conserved residues are in bold font. The putative signal peptides are underlined. The signal peptide regions are arbitrarily aligned at the start codon.

```
T. subulata      MSSPTGRELTIYLYLFSIIAV-FPFLTSASSSLKFD----LFGNMYRVHVINGFSSNDLP 55
P. trichocarpa   MASSS---LSASFPLLFILAISYPVLAWGSKPIPFNDNKSFFCFKFRVHIINGFSSNKNP 57
A. thaliana      MNCIKQFLLAICFSLALTCQDHVLVGGTTTRDIIVP-----KISEWQVTVVNGLTTGET- 54
P. rhoeas        MKLLYAILFLSFLTLASSR----------------------FLPVIEVRIMN-KRGNGHS 37

T. subulata      FLLHCWSSNDDLGHHSLYIGGDFNFHFGLRIIPPSTRFWCDMNRGPKYIPQVSVFEEDEV 115
P. trichocarpa   LSLHCWSQDNDLGNHTLYIGGDFNFKFGLASFG-KTIFHCDFKWAEKHR-FANVFTDGME 115
A. thaliana      LFIHCKSKEDDLGEINLKFRNRFSWNFGENMLH-STFFWCYMN-KDNGHMNVNVFWDDVI 112
P. rhoeas        IGIHCRSKDDDLGYHRISDGQQVHFSFRENFFH-TTTFNCDIEWDSRRHFNFDSYRAQRD  96

T. subulata      LHLCSHTQQCYWRGQDNGLYFCNDNS-SYFKLYDWYIRKKER 156
P. trichocarpa   SSTCCDTNSCYWKTEDDGIYFSNDNK-NYIKRLDWLK----- 151
A. thaliana      LFHRCGWKNCIWTAKTDGLYLWNSASGEDVLSRKWEVGW--- 151
P. rhoeas        DHGRCTT-ECLWKTTDEGLYGYDQEH-EYWQLY—YLAKK--- 132
```

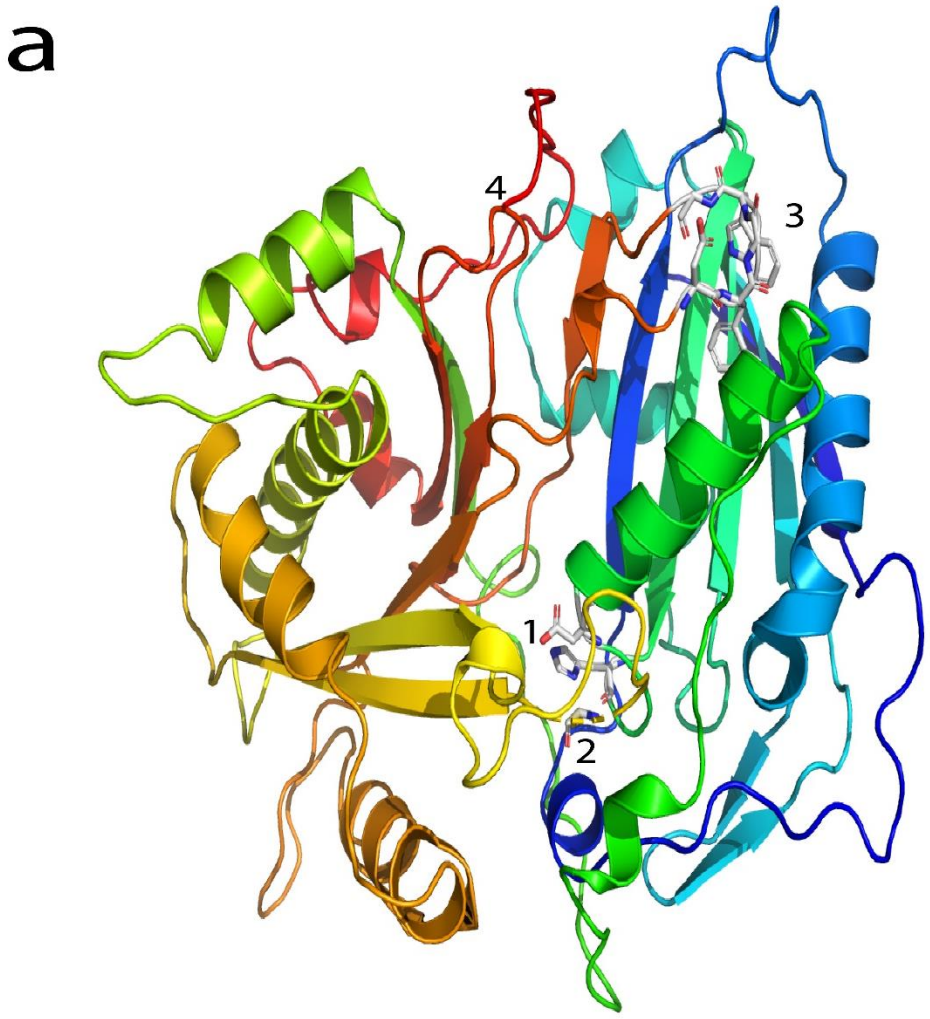
Figure S3. Inferred amino acid alignment of the *TsBAHD* gene of *T. subulata* against homologues from *A.thaliana*, *Populus trichocarpa*. This figure was taken from Shore *et al*. (2019). Functionally important regions are indicated in boxes: (1) HxxxD, (2) GN, and (3) DFGWGKP. Highly conserved amino acids highlighted in grey (Tuominen *et al*., 2011).

```
AtBAHD      ----MRVDVVSRDIIKPSSPTPNHLKKFKLSLLEQLGPTIFGPMVFFYSAN-----NSIK   51
TsBAHD      ----MEVEITLRETIKPSSSTPPSKRILKLSLLDQFTPVCYTSLILFYPASVNQ-DHSVT   55
PtBAHD      MSVPMRIEIMQRETIKPSSPTPLHLRSLKLSLLDQFMPVGHIPLQLFYPRNGNDTDHLAK   60
                *.:::  *: ***** **   : :*****:*: *. .  : :**  .     :  .

AtBAHD      PTEQLQMLKKSLSETLTHFYPLAGRLKGNISIDCNDSGADFLEARVNSPLSNLLLEPSSD   111
TsBAHD      LTERCQQLKISLSETLTHFYPLAGRLKDNASIECDDQGGEYIEARIKCLLSKFLVTPEAE   115
PtBAHD      ATERSLLLKTSLSEALTHFYPFAGRLKDNSSIECDDHGAEYIEARIHCILSDILKKPDTE   120
             **:   ** ****:******:*****.* **:*:* *.:::***::. **.:*  *.::
                                                      (1)
AtBAHD      SLQQLIPTSVDSIET-RTRLLLAQASFFECGSMSIGVCISHKLADATSIGLFMKSWAAIS   170
TsBAHD      LLKQLLPAAIESGEAATGSMLLVQANIFDCGGMAIGVCVSHKIADAATLTTFVKCWSGSA   175
PtBAHD      VLKQLMPAAVSEPATARDSQLIVQASFFDCGGLAIGVNLSHKVADAATLTSFIKCWAATA   180
             *:**:*:::..  :     *:.**.:*:**.::*** :***:***:::  *:*.*:. :

AtBAHD      SRGSIKTIGAPVFDTVKIFPPGNFSETSPAPVVEPEIMMNQTLSKRFIFDSSSIQALQAK   230
TsBAHD      HKS--TEVISPLFMGASIFPDMDLTFTRPSSV---ELMQGECATKRFVFDSAKIRMLKAK   230
PtBAHD      RRSSTEVVISPVFMGASIFPQMDLPISM-LPV---DLIQGESVMKRFVFEAPKITALKAK   236
             :.    : :*:*  ..***  ::  :    *   ::: .:   ***:*:: .*  *:**
                                                                      (2)
AtBAHD      ASSFEVNQPTRVEAVSALIWKSAMKATRTVSGTSKPSILANSVSLRSRVSPPFTKNSIGN   290
TsBAHD      AAGESVAAPTRVEAVSALIWKCAMAASRSNLGAHRKSVWSLSVNLRKRLAPPLPENYAGN   290
PtBAHD      AISASVPDPTRVESVTALIWKCAMSASRSNLGVPRKAVLSLGVNIRKRLVPTLPDNYGGN   296
            * . .*  *****:*:*****.** *:*:  *. : :: : .*.:*.*: * : .*  **

AtBAHD      LVSYFAAKAEEGINQTKLQTLVSKIRKAKQRFRDIHIPKLVGNPNATEIICSYQKEAGDM   350
TsBAHD      CVGSILAHMED-A-EVELRDLVGLIRRGNQEFDENYAKKL-QRDGSSKAICGFAKQFEEL   347
PtBAHD      YVGSISARIED-HDDLELQGIVSRIRKDLIEFGENYAKIT-QGDDTSLAICKAVEEFGKM   354
             *. : *: *:   : :*: :*. **:   .* : :        .::  **   ::  .:
                                  (3)
AtBAHD      IASGDFDFYIFSSACRFGLYETDFGWGKPVWVGFP-SVRQKNIVTLLDTKEAGGIEAWVN   409
TsBAHD      AMSTDIDFYNCSSWCKFQLYEADFGWGKPAWLSIA-STSIKNVLLLIDTRDAAGIEAWLT   406
PtBAHD      AMSKDIDSYNGTSWCRFELYDADFGWGKPTWLSNVFTIELKNIMCLMDTRDGDGIEACIS   414
              * *:* *  :* *:* **::*******.*:.   :   **:: *:**::. **** :.

AtBAHD      LNEQEMNLFEQDRELLQFASLNPSVIQPFLHVL     442
TsBAHD      LSEEDMALFECNKELLEFATLNPSVSLSELHL-     438
PtBAHD      LSREDMALFESNKELLEFAAANPSVSV------     441
            *..::* *** ::*** :**: ****
```

Figure S4. Alignment of inferred amino acid alignment of *TsYUC6* gene of *T. subulata* against *Yuc6* of *A. thaliana* (*AtYuc6*, AT5G25620.2). This figure was taken from Shore *et al*. (2019).

```
TsYuc6      MD-CL-GEIEDK--ENKANMSKPSRWVYAPGPVIVGAGPSGLAVAACLKEKGIPTVVIER    56
AtYuc6      MDFCWKREMEGKLAHDHRGMTSPRRICVVTGPVIVGAGPSGLATAACLKERGITSVLLER    60
            ** *   *:* *: .:: .*:.* *   . ************.******:** :*::**

TsYuc6      YSCVASLWKLKTYNRLRLHLPKQFCELPLMGFPSEYPTYPTKDQFIAYVEEYAKKFDIRP    116
AtYuc6      SNCIASLWQLKTYDRLHLHLPKQFCELPIIPFPGDFPTYPTKQQFIEYLEDYARRFDIKP    120
             .*:****:****:**:**********:: **.::******:*** *:*:**::***:*

TsYuc6      HFNETVSRAEYDQNLGFWHVKTIGLMAEETEYVCRWVVVATGENAEAVVPEIEGMREFV-    175
AtYuc6      EFNQTVESAAFDENLGMWRVTSVGEE-GTTEYVCRWLVAATGENAEPVVPRFEGMDKFAA    179
            .**:**. * :*:***:*:*.::*     *******:*.******* ***.:*** :*.

TsYuc6      -GDIRHTSLYKSGEDYKGKKVLVVGCGNSGMEVCLDLCNYNAKPSLVVKDKVHVLPQEML    234
AtYuc6      AGVVKHTCHYKTGGDFAGKRVLVVGCGNSGMEVCLDLCNFGAQPSLVVRDAVHVLPREML    239
             * ::**. **:* *: **:********************:.*:*****:* *****:***

TsYuc6      GTSTFGLSMFLLKWLPTRMVDRFLLIMSRLALGDTTRFGLYRPLLGPLEHKNLLGKTPVL    294
AtYuc6      GTSTFGLSMFLLKWLPIRLVDRFLLVVSRFILGDTTLLGLNRPRLGPLELKNISGKTPVL    299
            *************** *:******::**: ***** :** ** ***** **: ******

TsYuc6      DFGTLAKIKSGDVKICGSIKRLKHYTAEFVDGSTEKFDAIILATGYKSNVPSWLKERDFF    354
AtYuc6      DVGTLAKIKTGDIKVCSGIRRLKRHEVEFDNGKTERFDAIILATGYKSNVPSWLKENKMF    359
            *.*******:**:*:*..*:***:: .** :*.**:*********************..:*

TsYuc6      SEEDGFPRGTFQNGWIGERGLYAVGFAKRGILGASMEAKSVAEDIGRCWEAQAEHDPISF    414
AtYuc6      SKKDGFPIQEFPEGWRGECGLYAVGFTKRGISGASMDAKRIAEDIHKCWKQDEQLQCK-L    418
            *::****   * :** ** *******:**** ****:** :**** :**: : : :   :

TsYuc6      ARPLSDAIRQWNEN-----   428
AtYuc6      GKRM---KRKFSESDCGGN   434
            .: :    *::.*.
```

Appendix D: *TsBAHD* protein structures



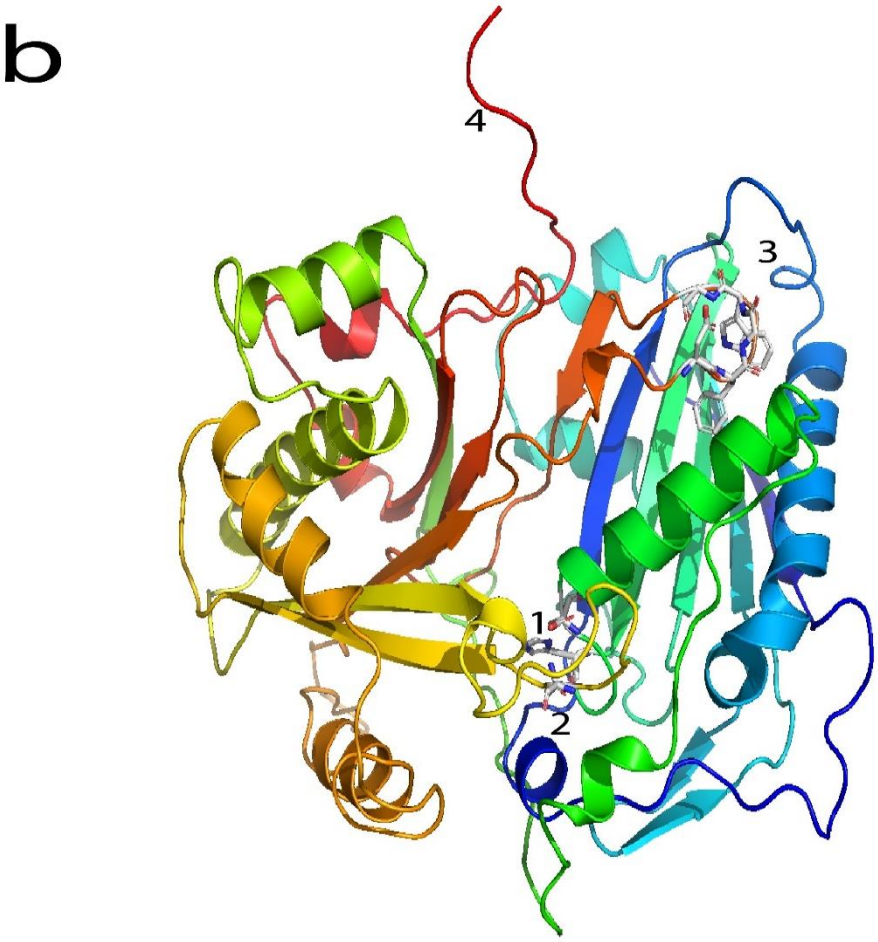


Figure S5. Protein structures of *TsBAHD* from long-homostyle mutant progeny and S-morph revertant Mhomo-S. The protein structures were generated by Raptor X (Peng and Xu, 2011) and edited using Pymol Software (The Pymol Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.). **(a)** The protein structure of *TsBAHD* from a long-homostyle mutant progeny of Mhomo-H. **(b)** The protein structure of *TsBAHD* from the S-morph revertant Mhomo-S. For both figures **a** and **b**: (1) represents the catalytic domain HxxxD; (2) represents the site of the serine amino acid substitution S for N in the GN motif in **a**, while in **b** the "normal" asparagine amino acid N occurs in the GN motif; (3) represents DFGWG domain; (4) represents the C-terminus of *TsBAHD*. The distance from the histidine of the HxxxD domain to the serine of the GN motif from the mutant *TsBAHD* and to the asparagine of the GN from the non-mutant *TsBAHD* is the same, 7.6 Å. No structural differences were observed between the mutant form of *TsBAHD* **(a)** and the non-mutant form of *TsBAHD* **(b)**.

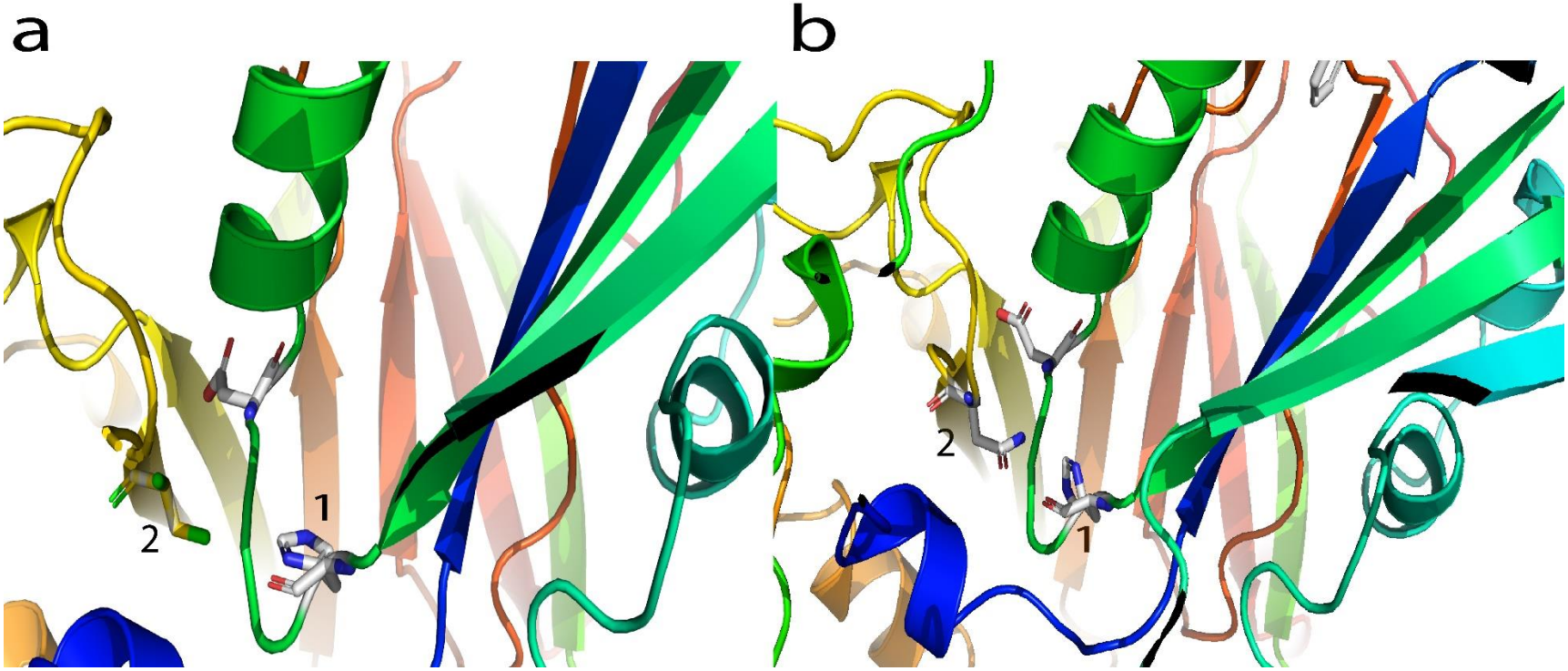

a
1
2
b
2
1

Figure S6. A closeup look at the HxxxD and GN domains of *TsBAHD* from the S-morph revertant Mhomo-S and a long-homostyle mutant progeny. The protein structures were generated by Raptor X (Peng and Xu, 2011) and edited using Pymol Software (The Pymol Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.). **(a)** The protein structure of *TsBAHD* from a long-homostyle mutant progeny of Mhomo-H. **(b)** The protein structure of *TsBAHD* from the S-morph revertant Mhomo-S. From both figures **a** and **b**: (1) represents the catalytic domain HxxxD; (2) represents the asparagine N of the GN motif in **a**, and the mutant serine amino acid S of the GN motif in **b**.