New Phytologist

The long and short of the *S*-locus in *Turnera* (Passifloraceae)

Joel S. Shore[1], Hasan J. Hamam[1], Paul D. J. Chafe[1], Jonathan D. J. Labonne[1], Paige M. Henning[2] and Andrew G. McCubbin[2]

[1]Department of Biology, York University, 4700 Keele Street, Toronto, ON M3J 1P3,Canada; [2]School of Biological Sciences, Washington State University, PO Box 644236, Pullman, WA 99164-4236, USA

Author for correspondence:
Joel S. Shore
Tel: +1 416 736 2100
Email: shore@yorku.ca

Received: *13 March 2019*
Accepted: *23 May 2019*

New Phytologist (2019) **224:** 1316–1329
doi: 10.1111/nph.15970

Key words: Distyly, Hemizygous, *S*-haplotypes, *S*-locus, *TsBAHD*, *TsSPH1*, *TsYUC6*, *Turnera*.

Summary

- Distyly is an intriguing floral adaptation that increases pollen transfer precision and restricts inbreeding. It has been a model system in evolutionary biology since Darwin. Although the *S*-locus determines the long- and short-styled morphs, the genes were unknown in *Turnera*. We have now identified these genes.
- We used deletion mapping to identify, and then sequence, BAC clones and genome scaffolds to construct *S*/*s* haplotypes. We investigated candidate gene expression, hemizygosity, and used mutants, to explore gene function.
- The *s*-haplotype possessed 21 genes collinear with a region of chromosome 7 of grape. The *S*-haplotype possessed three additional genes and two inversions. *TsSPH1* was expressed in filaments and anthers, *TsYUC6* in anthers and *TsBAHD* in pistils. Long-homostyle mutants did not possess *TsBAHD* and a short-homostyle mutant did not express *TsSPH1*.
- Three hemizygous genes appear to determine S-morph characteristics in *T. subulata*. Hemizygosity is common to all distylous species investigated, yet the genes differ. The pistil candidate gene, *TsBAHD*, differs from that of *Primula*, but both may inactivate brassinosteroids causing short styles. *TsYUC6* is involved in auxin synthesis and likely determines pollen characteristics. *TsSPH1* is likely involved in filament elongation. We propose an incompatibility mechanism involving *TsYUC6* and *TsBAHD*.

Introduction

Heterostyly is a plant mating system where two (distyly) or three (tristyly) mutually compatible floral morphs, differing in the arrangement of reproductive organs within flowers, occur in populations. It has been a subject of great interest in genetics and evolutionary biology at least since Darwin's (1877) book on the subject (Barrett, 1992; Gilmartin, 2015). Heterostyly has served as a model system in the rediscovery of Mendelian inheritance (Bateson & Gregory, 1905), an enduring example of a supergene (Dowrick, 1956; Charlesworth & Charlesworth, 1979; Thompson & Jiggins, 2014; Charlesworth, 2016), and a test of the operation and effects of frequency-dependent selection (Fisher, 1941). It also has provided a study system for the evolution of self-fertilization, especially in *Primula*, where the frequencies and relative fitness of derived self-compatible homostylous plants in otherwise distylous populations have been measured (Crosby, 1949; Bodmer, 1960; Piper *et al.*, 1986). Finally, there have been at least 23 independent origins of heterostyly (Lloyd & Webb, 1992a), implying that it is a solution to a common set of selection pressures for increased precision of cross-pollen transfer and the avoidance of inbreeding depression (Lloyd & Webb, 1992b; Barrett, 2002). Despite this sustained interest, the genes determining heterostyly are only now becoming known (reviewed in Kappel *et al.*, 2017).

Distyly is characterized by the presence of two genetically determined floral morphs: the long- and short-styled morphs (hereafter L- and S-morphs), with reciprocally positioned male and female reproductive organs. A self- and intra-morph incompatibility system usually restricts seed set to crosses between the morphs. Traits distinguishing the morphs exhibit single locus Mendelian inheritance with S-morph plants commonly heterozygous, *Ss*, and L-morph plants homozygous, *ss* (Lewis & Jones, 1992; Barrett & Shore, 2008).

Candidate genes for distyly have been proposed for a few species in unrelated families (Primulaceae, Linaceae and Polygonaceae; Kappel *et al.*, 2017). Diverse methods have been used to identify candidate genes, as well as to map and sequence the *S*-locus in distylous genera including *Fagopyrum*, *Linum*, *Primula* and *Turnera* (see references below). Given the numerous independent origins, it is perhaps not surprising that different genes might determine distyly in various lineages.

Genes hypothesized to control style length have been reported in three genera. In *Linum*, the *THRUM STYLE SPECIFIC GENE 1* (*TSS1*) is expressed in S-morph styles. *TSS1* is a novel gene of unknown function (Ushijima *et al.*, 2012, 2015). In *Fagopyrum esculentum*, *c.* 5.4 Mb of sequence is specific to the S-morph. Almost 75% of this region is comprised of transposable element (TE) derived sequences and 32 predicted genes lie within it. One of these, *S-Locus Early Flowering 3* (*S-ELF3*), has been

proposed to control style length (Yasui *et al.*, 2012). *S-ELF3* is a homologue of *ELF3* of *Arabidopsis,* which encodes a nuclear protein that regulates flowering. Functional *S-ELF3* is specific to the S-morph of *F. esculentum*, as well as other heterostylous *Fagopyrum* spp. Furthermore, self-compatible long-homostyle cultivars and homostyle *Fagopyrum* species possess inactivating mutations in *S-ELF3* (Yasui *et al.*, 2012). Other genes that might control features of distyly are the focus of ongoing research (Yasui *et al.*, 2016). To date, the only candidate for an *S*-locus gene that governs stamen characteristics comes from studies on *Primula*, the most intensively studied genus, in which the entire *S*-locus has been mapped and sequenced (Li *et al.*, 2016).

The *S*-locus in *Primula* involves a *c.* 280-kb region containing five genes, restricted to the dominant haplotype. Contrary to historical expectation, none of these genes have recessive allelic counterparts thus identifying the *S*-locus as hemizygous. The inserted region is flanked by two copies of a duplicated cyclin-like F-Box gene (*CFB*), present as a single copy in the recessive haplotype (Li *et al.*, 2016). The five genes are *CYP734A50*, *GLO2*, a Pumilio-like RNA-binding protein (*PUM*), a kelch-repeat F-box protein (*KFB*) and a gene encoding a protein with a conserved C-terminal domain (*CCM*) (Li *et al.*, 2016; Burrows & McCubbin, 2017). Potential functions for *KFB*, *PUM* and *CCM* in distyly have not been determined, but evidence suggests that *CYP734A50* controls pistil characteristics and *GLO2* stamen characteristics.

Primula CYP734A50 was first reported as an S-morph specific style expressed gene identified using tissue-resolved transcriptomics (Huu *et al.*, 2016). Several lines of evidence support *CYP734A50* as controlling style length. A functional copy of this gene is present only in S-morph individuals of distylous species and is absent from long-homostyle *Primula* species. In addition, independent long-homostyle mutants of *P. vulgaris* carry deletions within *CYP734A50* (Huu *et al.*, 2016). Most persuasively, virus-induced gene silencing of *CYP734A50* in the S-morph of *P. forbesii*, converted short styles to a long-homostyle phenotype (Huu *et al.*, 2016). *CYP734A50* is homologous to *BAS1* of *Arabidopsis*, a functionally characterized gene with brassinosteroid (BR) inactivating activity (Neff *et al.*, 1999). BRs regulate cell expansion and evidence suggests that reduced BR concentrations (reducing cell expansion) are responsible for short styles in the S-morph: short styles have reduced BR content relative to long styles, and application of exogenous BR is sufficient to convert short styles to a long phenotype (Huu *et al.*, 2016). *GLO2* of *Primula* is a MADS-box transcription factor, first identified as an *S*-linked gene (Li *et al.*, 2008), and is the most highly differentially expressed transcript between floral morphs with its expression specific to S-morph flowers (Nowak *et al.*, 2015). *GLO2* is a strong candidate for the gene controlling anther height, as a short-homostyle mutant carries a transposon insertion in *GLO2* whilst the other *S*-locus genes appear unchanged (Li *et al.*, 2016). As a transcription factor, *GLO2* has the potential to affect downstream gene expression, but exactly how it might regulate anther height is unclear.

Here we set out to identify the genes responsible for distyly in *Turnera* species. We ask: Are the *S*-locus genes different from those discovered recently in *Primula*, *Fagopyrum* and *Linum*? Is the S-morph in *Turnera* determined by hemizygous genes? Finally, can the organization of genes at the *S*-locus lead to suppressed recombination? We study gene expression and exploit mutants to explore the roles of candidate genes in distyly.

Materials and Methods

The plants used were grown in the glasshouse at York University or Washington State University. Most of the plants have been in cultivation for 15 or more years and were originally derived from open-pollinated seed collected in South and Central America, or the Caribbean. Distylous species included *Turnera subulata* Smith, *T. joelii* Arbo, *T. grandiflora* (Urb.) Arbo, *T. concinna* Arbo and *T. scabra* Millspaugh (Nicaragua and Dominican Republic). Long-homostyle species included *T. aurelii* Arbo, *T. campaniflora* Arbo, Barrett&Shore, *T. occidentalis* Arbo&Shore and *T. ulmifolia* var. *acuta* (Spreng.) Urb. The *Turnera* species studied are perennials, and cuttings are easily rooted. Older population samples have been re-grown from bulk seed for as many as six generations.

Polymerase chain reaction

For PCR we used an annealing temperature of 58°C unless indicated otherwise. Gene-specific primer sequences are provided (Supporting Information Table S1). PCR cycling parameters were: 5 min initial denaturation (94°C), followed by 35 cycles of: 30 s at 94°C, 30 s at 58°C, 30 s at 68°C, followed by 5 min at 68°C. For amplicons > 1 kb, extension times were increased as appropriate. Each PCR was performed using *c.* 100 ng of genomic DNA, 2 µl of forward and reverse primers (10 pmol μl^{-1}), and 10 µl of Quick-Load® Taq 2× Master Mix (NEB, Ipswich, MA, USA), in a 20-µl volume. Negative controls were included. PCR products were run on 0.8–1.5% agarose gels, stained with ethidium bromide.

Bacterial artificial chromosome sequencing and assembly

We identified bacterial artificial chromosome (BAC) clones of the *S*-haplotype by using PCR to screen for the presence of genes in the *S*-locus region in BAC DNA isolated from row and column pools of 96- or 384-well plates of two BAC libraries. One library (Ss library) had been prepared previously (Labonne & Shore, 2011). This was constructed from leaves of S-morph progeny (*Ss*) from a cross between two plants, S16L (*ss*) × F60SS (*SS*) (Fig. 1a) and contains clones derived from both haplotypes. A second BAC library (SS library) was constructed from a single homozygous (*SS*) S-morph plant (Fig. 1a). This library was constructed following Labonne & Shore (2011) but using the *Bam*HI cloning site of the vector. Coverage for the SS library was *c.* 6×, with an average insert size of 125 kb. We identified the allelic source of clones/genes (i.e. from the *S*- or *s*-haplotypes) using SSCP (single-stranded conformational polymorphisms) as described in Labonne & Shore (2011).

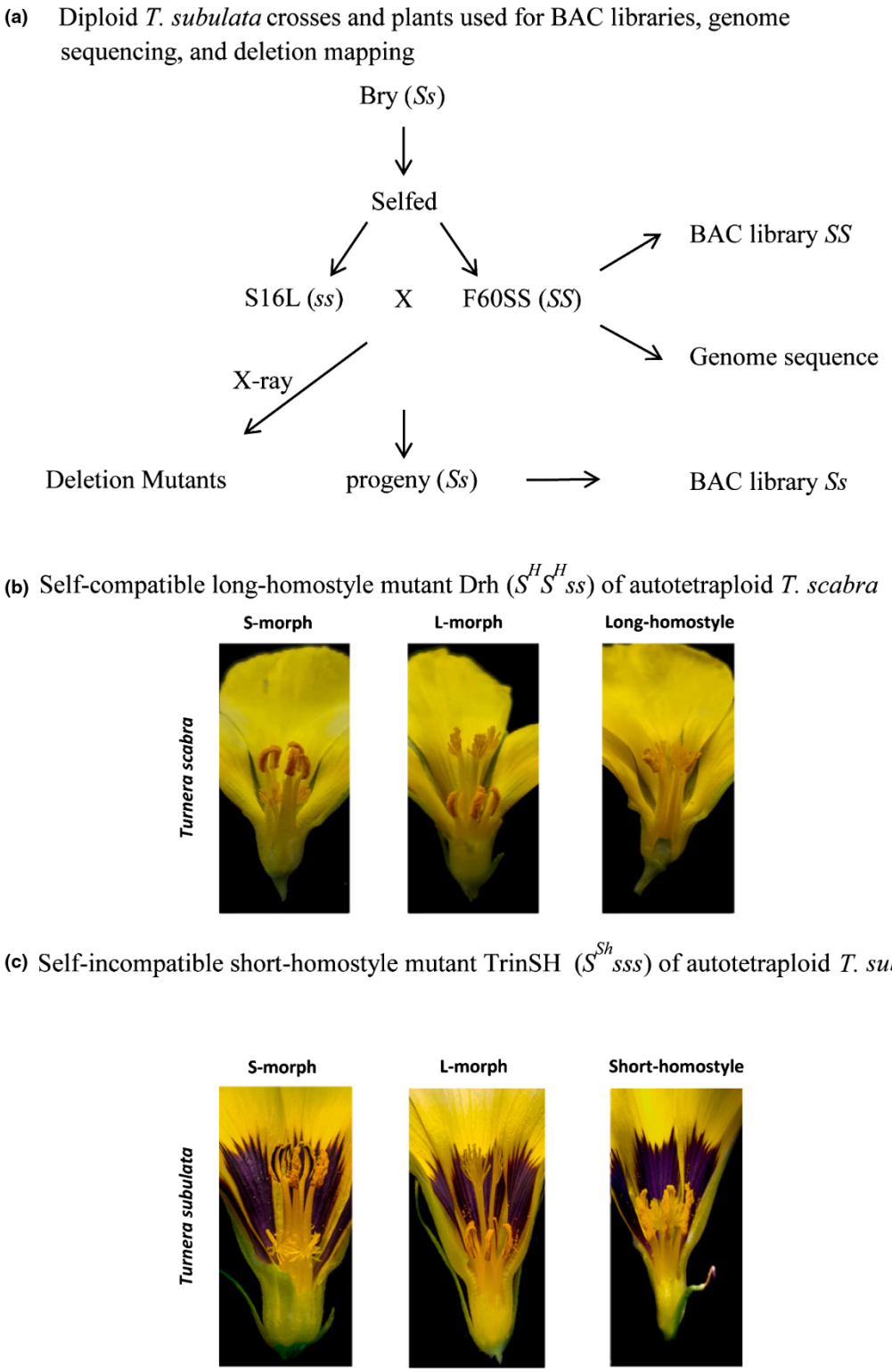

Fig. 1 Plants of *Turnera* used in this study showing species, genotypes, crosses, mutants and resources generated from them. (a) Plants used for bacterial artificial chromosome (BAC) library construction, genome and BAC sequencing, and generation of deletion mutants. (b) Genotype of self-compatible long-homostyle mutant, Drh, and photographs of L-morph, S-morph and Drh mutant flowers of *T. scabra*. (c) Genotype of self-incompatible short-homostyle mutant TrinSH and photographs of L-morph, S-morph and TrinSH mutant flowers of *T. subulata*.

Positive BAC clones were grown in large-scale cultures and purified using the QIAGEN Large-Construct Kit (Qiagen). Library construction, sequencing and assembly of reads were carried out at Genome Quebec and Washington State University. DNA libraries and BAC sequencing at Genome Quebec used the following protocols for each BAC clone: A DNA library was

prepared following the Pacific Biosciences 20 kb Template Preparation using the BluePippin Size-Selection System protocol. Approximately 7.5 μg of BAC DNA (100 μl final volume) was sheared using Covaris g-TUBES (Covaris Inc., Woburn, MA, USA) at *c.* 1500 ***g*** for 60 s on each side, using an Eppendorf centrifuge 5424 (Eppendorf, Hamburg, Germany). Sheared DNA was size-selected on a BluePippin system (Sage Science Inc., Beverly, MA, USA) using a cut-off range of 15–50 kb. DNA damage repair, end repair and SMRT bell ligation steps were performed following the protocol in the SMRTbell Template Prep Kit 1.0 (Pacific Biosciences, Menlo Park, CA, USA). Sequencing primer was annealed at a final concentration of 0.83 nM and the P6v2 polymerase was bound at 0.5 nM. The library was sequenced on a PacBio RSII instrument at a loading concentration of 100 pM using the MagBead OneCellPerWell loading protocol, DNA sequencing kit 4.0, SMRT cells v.3 and 6 h movies.

The BAC clone reads were assembled using the HGAP workflow (Chin *et al.*, 2013). Raw subreads were generated from PacBio data files. A subread length cut-off value was extracted and used in preassembly (BLASR) which aligns short subreads to long subreads allowing correction of errors on long reads. The corrected reads were then used in a Celera assembly, generating contigs. The contigs were polished by aligning raw reads (BLASR processed through a variant calling algorithm (Quiver) resulting in high-quality consensus sequences using local realignments and PacBio quality scores (Chaisson & Tesler, 2012; Chin *et al.*, 2013). BAC clone sequencing and assembly for two overlapping clones (BacI1 and BacK15) was performed by the WSU Genomics Core Laboratory, using minor modifications of the methods above, but pooling BAC DNAs at the outset of the process.

We sequenced one BAC clone (BAC-G5i) using Roche 454 FLX Titanium sequencing, and assembly was conducted using MIRA-3 (Chevreux *et al.*, 1999), with assembly parameters set to: genome, de novo, accurate, highly-repetitive. Once sequenced, overlapping BAC clones were assembled into scaffolds for each haplotype.

Genome sequence and assembly

We extracted DNA from young leaves of an S-morph homozygous (*SS*) plant of *T. subulata* (F60SS, Fig. 1a) using nuclei isolation, followed by CTAB extraction (modified from Williamson *et al.*, 2014). Briefly, leaves were ground in liquid nitrogen, resuspended in a nuclei isolation buffer, filtered through Miracloth (Millipore) and organelles lysed by addition of Triton X-100. Nuclei were pelleted and DNA was extracted using CTAB. The DNA was further purified using a DNeasy Plant Mini Kit (Qiagen).

We constructed a PCR-free Illumina shotgun library with an average insert size of *c.* 450 bp and two Nextera mate pair libraries with *c.* 5 kb inserts. The first two libraries were sequenced with Illumina HiSeq Rapid PE250 whereas the third was sequenced with Illumina HiSeq 2500 PE125. Reads had average quality values of 34. There were 72 685 390 paired-end reads from the first library, 15 734 190 from the second, and 279 876 652 from the third. We used DISCOVAR (2013) DE NOVO (Weisenfeld *et al.*, 2014) to assemble reads from the first library. We used NEXTCLIP (Leggett *et al.*, 2014) to filter Nextera reads for quality, and scaffolded the genome using SSPACE (Boetzer *et al.*, 2011) with default parameters but setting the maximum link ratio to 0.85 and the minimum number of links to three. Scaffolds were gap-filled (GAPCLOSER v.1.12; Luo *et al.*, 2012) using the default parameters and a maximum read length of 155 bp. This was followed with a final round of scaffolding.

Annotation

We annotated genome scaffolds using WQ-MAKER 2.31.9 running on XSEDE/JETSTREAM, and BAC clones and *S*-locus scaffolds using MAKER 2.31.8 (Holt & Yandell, 2011). We focus on genome scaffolds that had BLAST (Altschul *et al.*, 1990) hits to the *S*-locus region. A full account of genome assembly will be published elsewhere. During annotation we used databases of Uniprot proteins (The UniProt Consortium, 2019), transposable elements, repeats generated for *T. subulata* using REPEATMODELER v.1.0.10 (Smit & Hubley, 2015), and two *in silico* gene predictors (SNAP and Augustus) run as part of the annotation pipeline. Genome annotation was run three times to train SNAP parameters. Augustus parameters were trained by running BUSCO v.3.0.3 (see the Results section). During re-annotation, we included protein data bases from *A. thaliana*, *Populus trichocarpa* and *Vitis vinifera*.

Deletion mapping

We used 12 previously reported X-ray deletion mutants of *T. subulata* (Labonne *et al.*, 2010), to localize candidate genes determining the S-morph. These mutants have varying lengths of deletions of the dominant *S*-haplotype. Ten mutants were of the L-morph, one was a long-homostyle, and the other resembles a short-homostyle, but its style has the incompatibility behaviour of an L-morph plant (Labonne *et al.*, 2010).

For the deletion mutants and parental plants (F60SS and S16L; Fig. 1a), we PCR-amplified portions of the genes (*c.* 200–350 bp) that occur in the *S*-locus region and detected polymorphisms (SSCPs) using polyacrylamide gel electrophoresis (PAGE), and silver-staining, following the protocols of Labonne *et al.* (2008). For morph-specific genes, we used agarose gel electrophoresis. For *TsWRKY*, we designed PCR primers to a polymorphic sequence *c.* 560 bp upstream of the gene (Table S1).

Organ and time of gene expression

We carried out semi-quantitative reverse transcription polymerase chain reactions (RT-PCR) to determine floral organ specificity and timing of expression of three candidate genes. Anthers, filaments and pistils were collected separately from flower buds of each morph of autotetraploid *T. subulata*. Floral organs were collected at four developmental time points designated by bud length: 6–8, 10–12, 14–16 and 18+ mm. RNA was extracted using Concert Plant RNA Reagent (Invitrogen) following the manufacturer's 'Small-Scale Isolation' protocol. RNA

samples were treated with DNase I (Thermoscientific, Waltham, MA, USA) and removal of genomic DNA confirmed by gel electrophoresis. cDNA was synthesized using an oligo (dT) primer and Episcript reverse transcriptase (Epicentre, Madison, WI, USA), according to the manufacturer's protocol. Samples were normalized using a housekeeping gene, β-*tubulin*. Once normalized, expression of each *S*-locus gene was assessed using gene-specific primers (Table S1). For semi-quantitative RT-PCR, we used MyTaq 2× Red mix (Bioline, Memphis, TN, USA), 5 μM of each primer in a final volume of 15 μl. Thermocycler parameters were optimized for each primer pair. PCR products were separated on 1.5% agarose gels for 20 min at 100 V and stained with ethidium bromide.

Analyses of *Turnera* species and mutants

We determined whether the candidate genes were restricted to the S-morph in five distylous species including diploids *T. scabra*, *T. grandiflora*, *T. concinna*, *T. joelii* and autotetraploid *T. subulata*. We extracted genomic DNA using the CTAB protocol (Labonne & Shore, 2011). We PCR-amplified the genes commonly analysing five plants of each morph with gene-specific primers, and used β-*tubulin* as a positive control (Table S1). PCR products were run on agarose gels (as above).

One of the candidate genes, *TsSPH1*, had been sequenced from four homostylous species (allo-octoploid: *T. aurelii*, and allohexaploids: *T. campaniflora*, *T. occidentalis* and *T. ulmifolia* var. *acuta*) and from a short-homostyle mutant, TrinSH, and an S-morph plant of autotetraploid *T. subulata* (Table S2). Sanger sequencing of gel-purified PCR products was carried out in both directions with the forward and reverse primers (at the Génome Québec Innovation Centre using an Applied Biosystems 3730xl DNA Analyzer; Foster City, CA, USA).

We analyzed two mutants to investigate the roles that candidate genes might play in distyly. We studied a self-compatible long-homostyle mutant, Drh (Fig. 1b), of distylous autotetraploid *T. scabra* from the Dominican Republic (DR), characterized by Tamari *et al.* (2005). We used PCR to amplify each of three candidate genes from this mutant, as well as from L- and S-morph plants of the *T. scabra* DR population. PCR amplicons were run on agarose gels, as above.

A self-incompatible short-homostyle mutant of distylous autotetraploid *T. subulata* was collected from a garden in Trinidad and has been cultivated in the glasshouse at York University for *c.* 25 yr (Fig. 1c). The mutant has the compatibility behaviour of a normal S-morph plant. We crossed the mutant to an L-morph plant and investigated the presence/absence and expression of candidate genes in this mutant, three offspring and control plants (as above).

Results

Genome sequencing

We will provide a full account of the genome assembly of *T. subulata* elsewhere. Here we focus on scaffolds in the *S*-locus region but provide a summary of the genome assembly statistics. The sequence reads provided an average coverage of *c.* 110×. The *de novo* draft assembly yielded 54 996 contigs (> 500 bp) which were merged into 22 870 scaffolds. The assembly covered 579.6 Mb or *c.* 85.9% of the *T. subulata* estimated genome size of 674.8 Mb (0.69 pg; López *et al.*, 2011). There was an N50 contig size of 51.9 kb, an N50 scaffold size of 103.1 kb, and an estimated 22 989 genes. We searched for the presence of 1440 core eukaryotic genes (from *A. thaliana*), using BUSCO v.3.0.3 (Simão *et al.*, 2015), which revealed the presence of 94.1% of the genes in the genome assembly. We used BLAST searches of *S*-locus genes from BAC sequences to identify genomic scaffolds in the *S*-locus region.

BAC sequencing and assembly across the *s*- and *S*-haplotypes

We obtained sequences of eight BAC clones from the *S*-locus region. All but one BAC clone were sequenced with PacBio long-read sequencing and yielded single contigs with very high coverage (Table S3). Average read lengths exceeded 5000 bp for PacBio sequencing. For BAC-G5i, Roche-454-Titanium sequences had an average read length of 469 bp after trimming.

We consider two BAC clones to constitute the *s*-haplotype (BAC-J10, 155 226 bp and BAC-L22, 192 366 bp; Table S3; Labonne & Shore, 2011). Ends of the two clones overlap by 23 048 bp (with five mismatched bp adjacent to homopolymeric sequences). The length of the assembled *s*-haplotype is 324 544 bp. It contains 21 putative genes based upon BLAST searches (Table S4; Fig. 2a).

Labonne & Shore (2011) identified two BAC clones corresponding to portions of the *S*-haplotype (BAC-I1 and BAC-K15) and we identified four additional BAC clones and have sequenced them all (Table S3). BAC-I1 and BAC-K15 were sequenced and assembled in a single sequencing run (now called BAC-I1K15) and it overlaps with BAC-G5i for 47 407 bp (showing eight mismatches). The opposite end of BAC-G5i overlaps identically with BAC-A24 for 54 542 bp. BAC-A24 and BAC-I7 overlap identically for 79 624 bp. Lastly, a scaffold from the *T. subulata* genome assembly (Scf2437, 73 192 bp) overlaps identically (11 786 bp) with the end of BAC-I7. The total length of this assembled portion of the *S*-haplotype is 724 069 bp (Fig. 2).

Despite its considerable length, the assembled *S*-haplotype (comprising BAC clones and one genome scaffold) contains nine genes as opposed to 21 possessed by the recessive *s*-haplotype, but significantly, the *S*-haplotype contains three additional genes (*TsSPH1*, *TsYUC6* and *TsBAHD*; Fig. 2; Table S4). All three genes occur on BAC-A24, whereas the adjacent clones, BAC-G5i and BAC-I7, each share one of these genes in regions of overlap with BAC-A24. BAC-C7e, which possesses the *TsAp2* gene, does not overlap with the remaining BAC clones.

The remaining genes corresponding to those of the *s*-haplotype occur on six scaffolds from the *T. subulata* genome assembly (Fig. 2b; Table S4). We have aligned these scaffolds both using deletion mapping and the gene order of the *s*-haplotype (Fig. 2). These genome scaffolds have 14 genes in common with

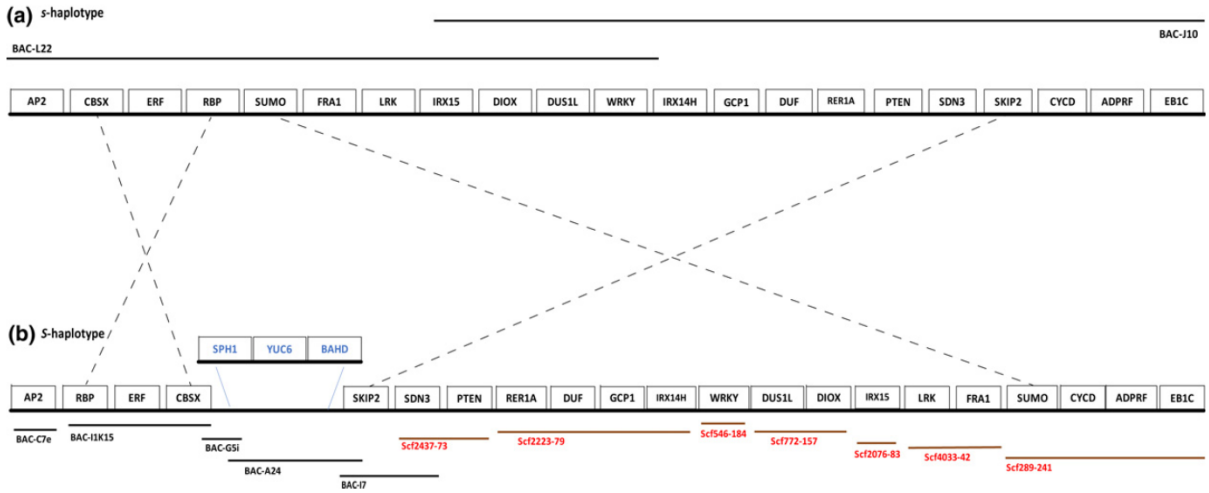

Fig. 2 Schematic showing *s*- and *S*-haplotypes of *Turnera subulata* including the genes they possess, the BAC clones (black lines) and/or genome scaffolds (brown lines) of which they are comprised and overlap between them. Inverted gene orders are indicated by the dashed lines. (a) The *s*-haplotype comprises BAC-L22 and BAC-J10. (b) The *S*-haplotype is comprised of five BAC clones and seven genome scaffolds. An insertion of three genes (blue font) on the *S*-haplotype is shown. The genome scaffold identifiers are the particular scaffold number followed by its size in kb (e.g. Scf289-241 is Scf289 which is 241 kb). Haplotypes, clones and scaffolds, are not drawn to scale.

the *s*-haplotype and a total length of 786 998 bp. Thus, the total length of the *S*-haplotype (BACs plus genome scaffolds which includes genes *TsAp2* to *TsEB1C*; Fig. 2b) is *c.* 1488 900 bp, *c.* 4.6 times the length of the *s*-haplotype.

In addition to the insertion of three genes (*TsSPH1*, *TsYUC6* and *TsBAHD*), there appear to be at least two inversions (one of three genes and another of 14 genes) in the dominant *S*-haplotype relative to the recessive (Fig. 2). The inversion of a 14-gene segment is supported at the distal end by the occurrence of four genes; *TsSUMO*, *TsCYCD*, *TsADPRF* and *TsEB1C*, all on scaffold Scf289, whereas the latter three genes are linked to *TsSKIP2* on the *s*-haplotype (Fig. 2; Table S4). Furthermore, genes of the *S*-haplotype (from *TsEB1C* to *TsDUF247*) are all possessed by deletion mutant L22 (Fig. 3) supporting this gene order. In addition, *TsSUMO* is linked to *TsRBP* on the *s*-haplotype, not to *TsCYCD* as on the *S*-haplotype Scf289 scaffold (Fig. 2). At the proximal ends of the haplotypes, the gene order *TsCBSX*, *TsERF*, *TsRBP*, is inverted. Finally, five genome scaffolds do not overlap (Fig. 2) and therefore their order and orientation are not known with certainty raising the possibility of additional rearrangements.

Much of the additional sequence of the *S*-haplotype appears to be due to TE-derived sequences. Approximately 76.6% of the *S*-haplotype sequence appears to be due to TEs, as opposed to 45.1% for the *s*-haplotype (Table S5). A sample of control scaffolds (i.e. non-*S*-locus scaffolds) matched by size to each *S*-locus BAC and scaffold, had a mean of 59.9% of sequence comprised of TEs (range 28.6–91.7%). Interestingly, the *S*-haplotype BACs had greater percentages of TE sequence (range 84.4–92.5%) whereas scaffolds at the distal end of the *S*-haplotype (Scf4033 – 35.4%; Scf289 – 57.2%) had the least.

Deletion mapping

We PCR-amplified 15 genes across the *S*-haplotype from genomic DNA of X-ray deletion mutants. The *TsSPH1*, *TsYUC6* and *TsBAHD* genes do not amplify from genomic DNA of the deletion mutants (Fig. 3a), with the exception of the long-homostyle mutant (LH1) for which both *TsSPH1* and *TsYUC6* amplified, but not *TsBAHD*. Approximate locations of breakpoints for various mutants are illustrated (Fig. 3b). The *S*-haplotype alleles of *TsSKIP2*, *TsSDN3* and *TsPTEN* also are absent from all deletion mutants, whereas the mutants possess the *s*-haplotype alleles of those genes. We have not developed a marker to determine whether *TsRER1A* occurred in any deletion mutants. The alignment of mutants against the *S*-locus region extends beyond the genes investigated here, as Labonne *et al.* (2010) and Labonne & Shore (2011) used markers further up- and downstream of the *S*-locus. Here we include the *H5SP1* and *N24SP1* markers of Labonne & Shore (2011), which lie 0.1 and 0.3 cM, respectively, on opposite sides of the *S*-locus (Fig. 3a). The deletion mapping implicates *TsBAHD* as the determinant of the short style phenotype, and both *TsSPH1* and *TsYUC6* as determinants of S-morph stamen characteristics.

Hemizygosity of *TsSPH1*, *TsYUC6* and *TsBAHD* in five *Turnera* species

PCR-amplification of *TsSPH1*, *TsYUC6* and *TsBAHD* genes from plants of the S-morph for the five *Turnera* species investigated provided further evidence of their hemizygosity (Fig. 4). We also had previously amplified and sequenced *TsSPH1* from four long-homostyle species of *Turnera*. All possess what appears to be a functional copy of the gene (Table S2).

Gene expression

We used semi-quantitative RT-PCR to determine organ and time of expression of *TsSPH1*, *TsYUC6* and *TsBAHD* (Fig. 5). None of the three genes are expressed in L-morph plants. *TsSPH1* and *TsYUC6* both showed expression in anthers, with

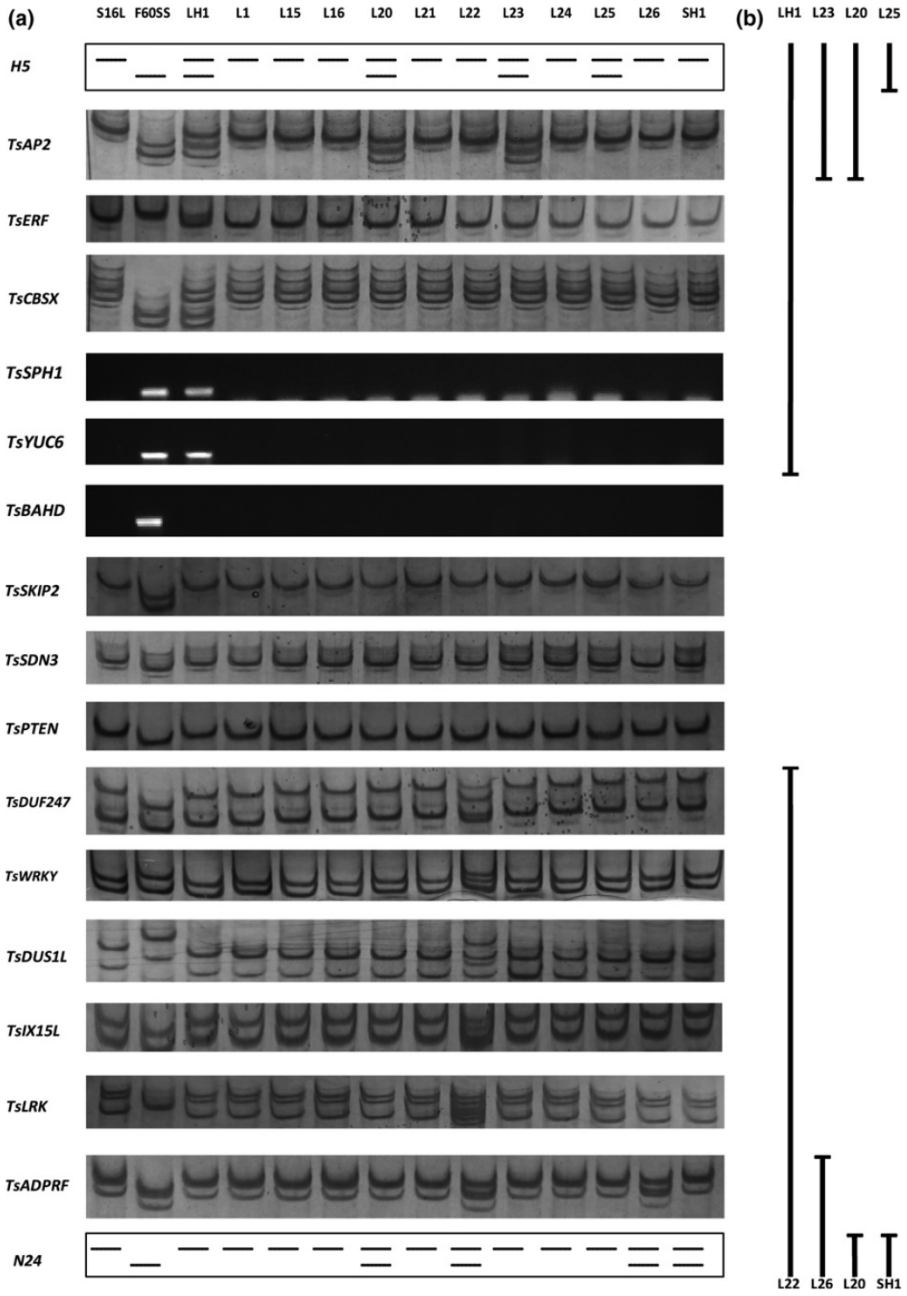

Fig. 3 Deletion map of 17 polymorphic genes across the *S*-locus of *Turnera subulata*. (a) The parental plants, S16L (L-morph) and F60SS (S-morph), and 12 X-ray deletion mutants were assayed for each gene. All deletion mutants are L-morph with the exception of LH1 which is a long-homostyle, and SH1 which appears to be a short-homostyle. The gene order follows that of the *S*-haplotype. The upper and lower panels are schematics of previously mapped markers *H5SP1* (H5) and *N24SP1* (N24), which lie 0.1 cM and 0.3 cM, respectively, on opposite sides of the *S*-locus. Mutants showing additive banding patterns of the parental plants possess copies of alleles from both haplotypes. Single-strand conformation polymorphism analysis was used to detect all polymorphisms except those of *TsSPH1*, *TsYUC6* and *TsBAHD*, for which PCR amplicons were run on agarose gels. (b) Vertical lines indicate the *S*-haplotype genes possessed by various deletion mutants.

an apparently greater level of expression in younger buds (Fig. 5a). *TsYUC6* showed two (alternatively spliced) transcripts (details of the alternative splicing are provided in Fig. S1). *TsBAHD* was not expressed in anthers. Only *TsSPH1* was expressed in filaments (Fig. 5b) with an apparent greater level of expression at later bud stages. Only *TsBAHD* was expressed in styles, with an apparent greater level of expression at intermediate bud ages/sizes (Fig. 5c).

Amino acid alignments

Alignment of the inferred amino acid sequence of the *TsSPH1* gene of *T. subulata* against putative homologues from *P. trichocarpa,* and *A. thaliana* showed that they shared conserved regions in common with the *Papaver rhoeas S8*-allele of *SPH1* (Fig. S2). The sequences possess a putative signal peptide predicted using SignalP 4.0 (Petersen *et al.*, 2011). We align the inferred amino acid sequence of the *TsBAHD* gene against those of *A. thaliana* (AT3G26040) and *Populus trichocarpa* (XP_002314550). The sequences possessed the functional and conserved regions of BAHD acyltransferases (Bayer *et al.*, 2004; Tuominen *et al.*, 2011; Fig. S3). The amino acid alignment of *TsYUC6* against *Yuc6* of *A. thaliana* showed 63% identity at the amino acid level (Fig. S4). *Arabidopsis thaliana* has two alternatively spliced forms of this gene AT5G25620.1 and AT5G25620.2 (417 vs 434 amino acid residues in length, respectively).

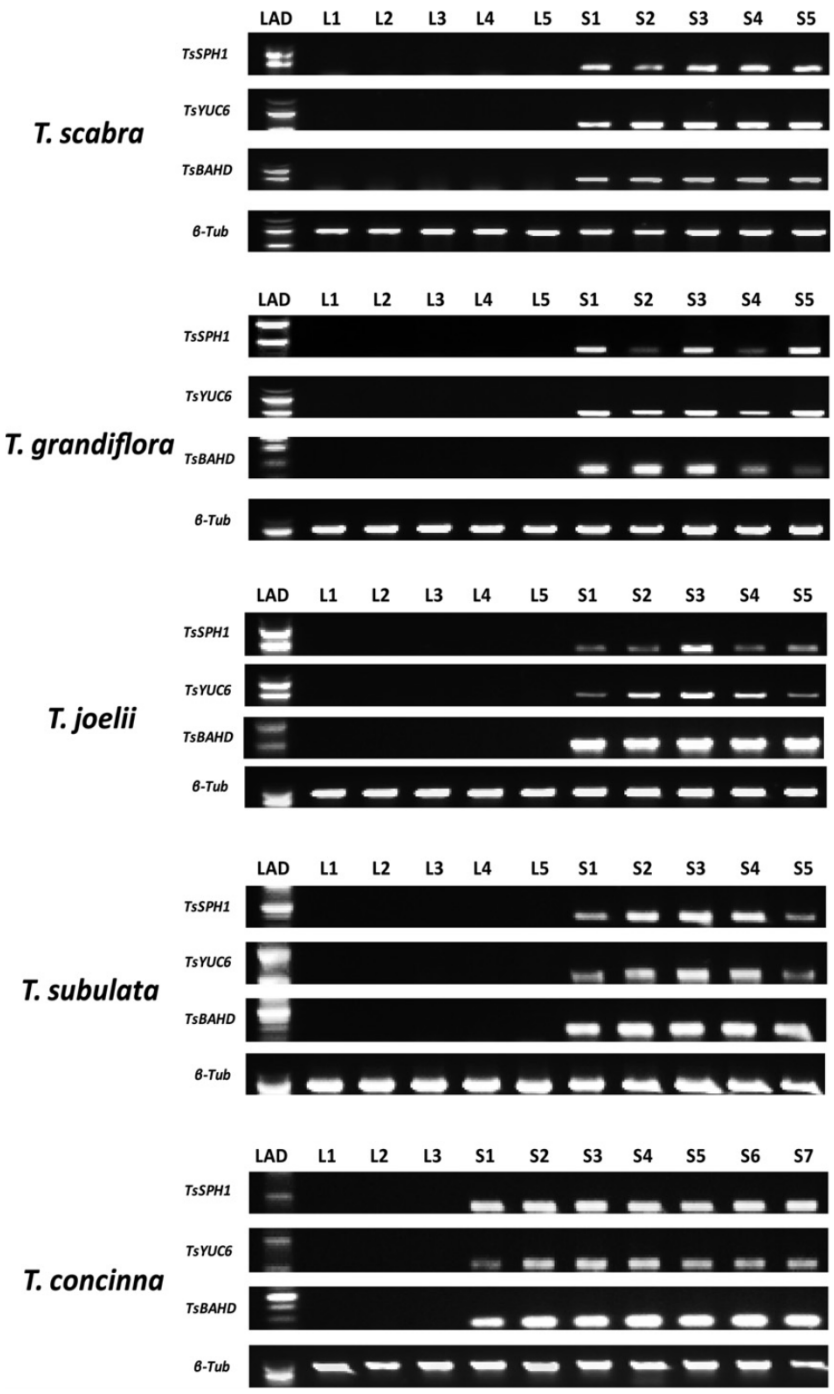

Fig. 4 Hemizygosity of *TsSPH1*, *TsYUC6* and *TsBAHD* for five *Turnera* species. PCR amplicons of the genes occur only in S-morph plants (lanes labelled S) and are absent from L-morph plants (lanes labelled L). The β-*tubulin* (β-*Tub*) control gene occurs in all samples. Five plants of each morph were assayed for each species with the exception of *T. concinna* (three L-morph and seven S-morph). LAD, smallest band for all genes is 500 bp except for β-*Tub*, which shows a 1-kb band.

Mutants and the function of genes in distyly

We analyzed two additional mutants for the presence and/or expression of three hemizygous genes. The self-compatible long-homostyle mutant, Drh, does not possess *TsBAHD*, which does occur in S-morph plants of autotetraploid *T. scabra* from the DR (Fig. 6). The Drh mutant possesses both *TsSPH1* and *TsYUC6* (Fig. 6).

We used PCR to amplify the two stamen-specific hemizygous genes (*TsSPH1* and *TsYUC6*) from DNA of a self-incompatible short-homostyle mutant of autotetraploid *T. subulata* (TrinSH), and three of its progeny (two short-homostyle and one of the S-morph). Both *TsSPH1* and *TsYUC6* amplified from short-homostyle and S-morph plants, but not from L-morph plants (Fig. 7a). RT-PCR revealed that *TsSPH1* was not expressed in filaments of the short-homostyle mutant nor short-homostyle progeny, but was expressed in the single S-morph progeny assayed and a control S-morph plant (Fig. 7b). *TsYUC6* was expressed in all short-homostyle and S-morph plants (Fig. 7b). *TsSPH1* was sequenced from the TrinSH mutant and from a normal S-morph plant of autotetraploid *T. subulata* (Table S2). The inferred amino acid sequences were identical whereas there were three nucleotide substitutions.

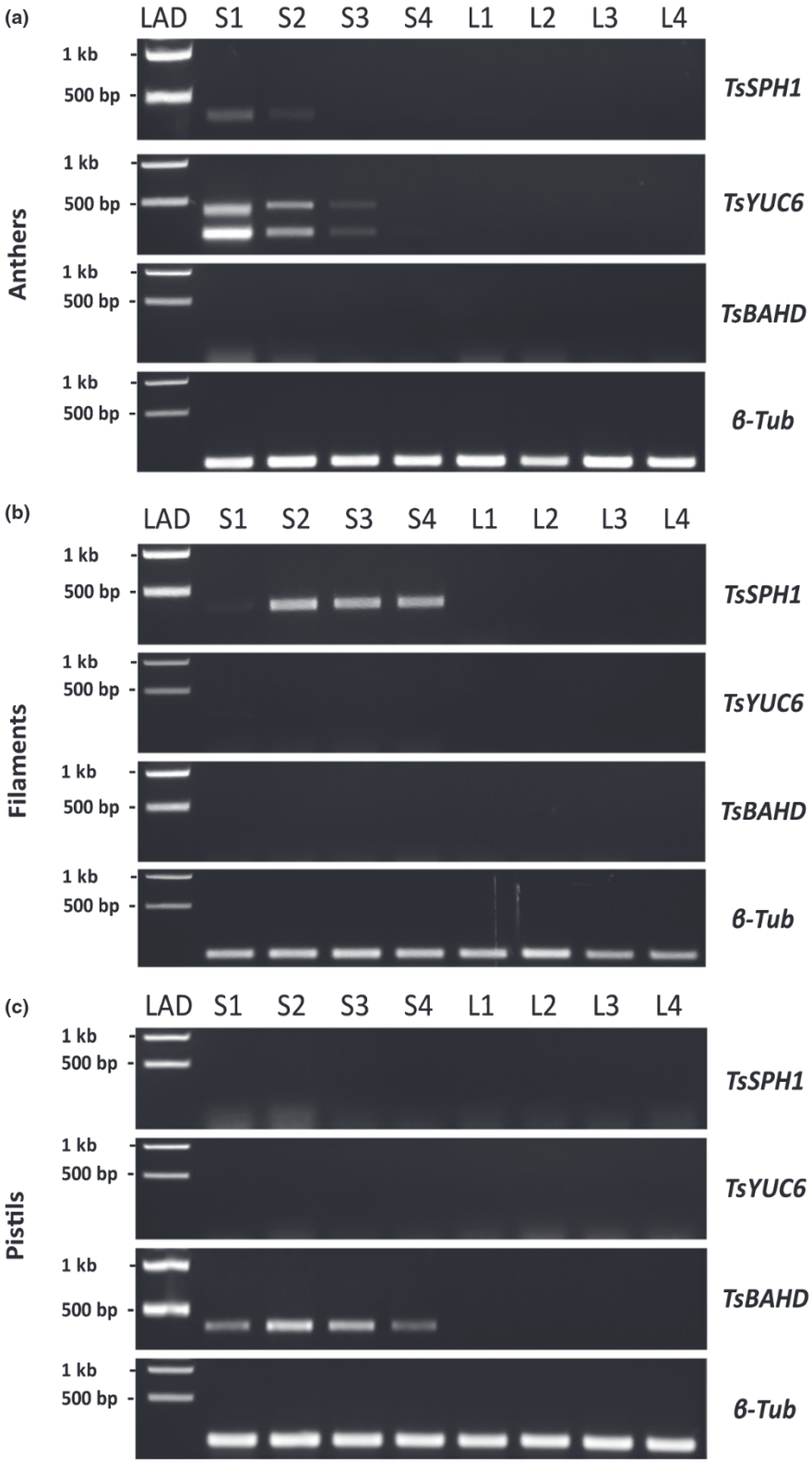

Fig. 5 Semi-quantitative reverse transcription polymerase chain reaction for *Turnera subulata* reveals differential expression of the *TsSPH1*, *TsYUC6*, and *TsBAHD* genes in different floral organs of S-morph plants and absence of expression in L-morph plants. Results for anthers, filaments and pistils, from buds of increasing age (1, 6–8 mm; 2, 10–12 mm; 3, 14–16 mm; and 4, > 18 mm), for organs from S-morph (lanes S1–S4) and L-morph plants (lanes L1–L4). (a) Anthers: *TsYUC6* appears to be most highly expressed in the smallest buds and shows evidence of two transcripts. *TsSPH1* is more highly expressed in the smallest buds whereas *TsBAHD* is not expressed. (b) Filaments: only *TsSPH1* is expressed, and it occurs in all buds but appears to be more weakly expressed in the smallest buds. (c) Pistils: only *TsBAHD* is expressed, with apparent greater expression in pistils from intermediate-sized buds. The β-*tubulin* control gene (β-*Tub*) is expressed in both morphs and all buds. LAD, 1-kb ladder.

Discussion

Our study assembled dominant and recessive haplotypes across the *S*-locus of *Turnera subulata*. The *S*-haplotype contains three genes (*TsSPH1*, *TsYUC6* and *TsBAHD*) not found in the recessive assembly or in L-morph plants. These genes appear to play a role in determining the phenotype of the S-morph. It is not clear if the remaining *S/s*-haplotype genes play any role in distyly. From deletion mapping alone, we cannot discount the possibility of a role for *TsSKIP2*, *TsSDN3* and *TsPTEN*, because the

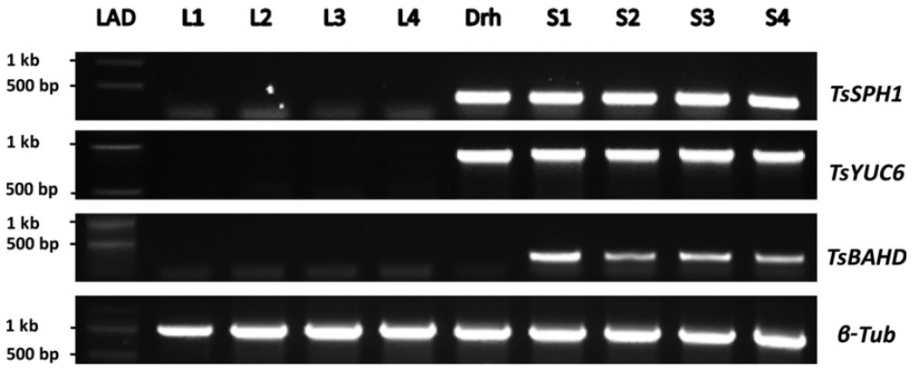

Fig. 6 Assays for presence/absence of *TsSPH1*, *TsYUC6* and *TsBAHD* in a long-homostyle mutant, Drh, and four L- and S-morph plants of autotetraploid *Turnera scabra*. L, L-morph; S, S-morph. All three genes amplify in S-morph plants but not in L-morph plants. The *TsBAHD* gene did not amplify in the long-homostyle mutant, Drh, but both *TsSPH1* and *TsYUC6* did amplify. A β-*tubulin* (β-*Tub*) control gene amplified in all samples. LAD, 100-bp ladder.

S-haplotype alleles of these genes were deleted from all 12 mutants. Interestingly, *TsSKIP2* shows homology to the pollen-specific *S*-locus F-box proteins of the S-RNase-mediated self-incompatibility systems of *Antirrhinum*, *Petunia* and *Prunus* spp. (Wang *et al.*, 2004). *TsDUF247* contains a protein domain of unknown function (DUF247) and the dominant allele is deleted from all mutants but one (mutant L22). Intriguingly, the candidate genes for the two-locus gametophytic self-incompatibility system of *Lolium perenne* are *S-DUF247* and *Z-DUF247* (Manzanares *et al.*, 2015; Thorogood *et al.*, 2017). Further investigation of these genes and perhaps other *S/s*-haplotype genes are warranted. In addition, our analysis targeted genes determining the S-morph; the genes responsible for characteristics of the L-morph require investigation. Below we compare the function and organization of the *S*-locus genes with that of other distylous species, consider possible roles for each hemizygous gene in distyly, and a possible mechanism of incompatibility.

Hemizygosity in all systems

The discovery that the *S*-locus in *Turnera* is a hemizygous region parallels that of *Primula* (Li *et al.*, 2016), *Linum* (Ushijima *et al.*, 2012, 2015) and *Fagopyrum* (Yasui *et al.*, 2016). Although these three genera represent a rather small sample in regard to the number of independent origins of heterostyly, it is beginning to appear as though such hemizygous regions are a common theme in the evolution of these complex mating systems.

Haplotype organization and size

The recessive *s*-haplotype of *T. subulata* is collinear with *Vitis vinifera* (NC_012013.3 Reference 12× Primary Assembly) from positions 16 436 957 to 16 233 322 on chromosome 7, with the exception of an additional gene in *V. vinifera*. Likewise, the *s*-haplotype showed considerable collinearity with *Populus trichocarpa* (NC_037291.1 Chromosome 7 Reference Pop_tri_v3 Primary Assembly; positions 4210 642 to 4583 579). *Populus trichocarpa* does not possess *SUMO* or *ADPRF* and has two genes not present in either *V. vinifera* or *T. subulata*. Although we have not searched exhaustively, the region also is highly collinear in other species in the Malpighiales including *Jatropha curcas* (scaffold NW_012124448.1, position *c.* 1510K), but also *Durio zibethinus* (scaffold NW_019168470, position *c.* 3840K) in the Malvales. Homologues of the hemizygous genes on the *S*-haplotype (*TsSPH1*, *TsYUC6* and *TsBAHD*) do not cluster in *V. vinifera* or *P. trichocarpa* but fall on different chromosomes. The *s*-haplotype clearly possesses the ancestral gene order and content, whereas the *S*-haplotype is a novel arrangement including three hemizygous genes. There are two inversions in the *S*-haplotype relative to the *s*-haplotype. The order and orientation of five *S*-haplotype scaffolds are not known with certainty raising the possibility of additional re-arrangements of gene order.

Putative paralogues of the hemizygous genes occur on different scaffolds of the *T. subulata* genome assembly. A paralogue showing the greatest identity (80%) to *TsBAHD* occurs on a 47.8 kb scaffold (GenBank no. MK922466) containing two additional genes encoding BAHDs; a paralogue with the greatest identity (83%) to *TsYUC6* occurs on a 193.8-kb scaffold (GenBank no. MK922465); and, a paralogue with the greatest identity (79%) to *TsSPH1* occurs on a 149.1-kb scaffold (GenBank no. MK922467) which contains four additional genes encoding SPH1-like proteins. Given that these putative paralogues occur on different scaffolds of the *T. subulata* genome, this could indicate that the hemizygous genes have moved to the *S*-haplotype in a stepwise manner, a possibility that might support Lloyd & Webb's (1992a) stepwise model for the evolution of distyly.

We consider the *S*-locus haplotypes to begin with the *TsAP2* gene and end with *TsEB1C*. They differ considerably in length with the *S*-haplotype 4.6× (*c.* 1.5 Mb) the size of the *s*-haplotype, with the difference apparently due to a greater percentage of TEs (76.6%), comparable to that of *F. esculentum* (Yasui *et al.*, 2012). The accumulation of TEs is expected to occur in regions of suppressed recombination (Uyenoyama, 2005), but the extent may vary with element type and the pattern is not universal (Kent *et al.*, 2017). The inversions of haplotypes in *T. subulata* should suppress recombination at the *S*-locus. Labonne *et al.* (2007) had detected evidence of recombination suppression across the *S*-locus in *T. subulata*, although their estimates were not statistically significant, likely due to limited power because of the distance of the flanking markers from the *S*-locus.

The abundance of TE sequences in the *S*-haplotype might provide substrates for ectopic recombination (Kent *et al.*, 2017) within the region, generating further rearrangements. Our investigation of gene order has been restricted to chromosomes derived from a single individual of *T. subulata*. Investigation of

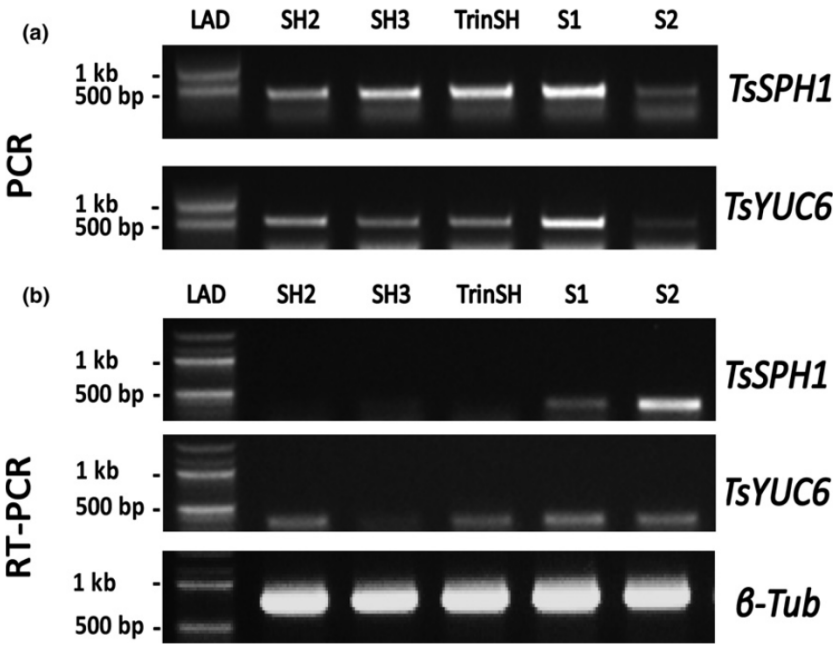

Fig. 7 Assays for the presence/absence and expression of *TsSPH1* and *TsYUC6* in a short-homostyle mutant, TrinSH, two short-homostyle and one S-morph progeny as well as a control S-morph plant of autotetraploid *Turnera subulata*. (a) PCR amplification of *TsSPH1* and *TsYUC6* from genomic DNA. Lanes SH2 and SH3 (short-homostyle progeny of TrinSH); TrinSH mutant; Lane S1, S-morph progeny; Lane S2, S-morph control plant. (b) Reverse trasncsription polymerase chain reaction (RT-PCR) of *TsSPH1*, *TsYUC6* and the β-*tubulin* control gene (β-*tub*) from stamens. Lanes are as above. The primers used for *TsYUC6* target only one isoform.

other *T. subulata* individuals and *Turnera* species would be of value. We might expect different arrangements of the *S*-haplotype in divergent lineages. Given that the three hemizygous genes occur within an approximate 230 kb region, it is possible that translocations could move the genes to other genomic locations.

Roles of the hemizygous genes in distyly

We have discovered three hemizygous genes in S-morph plants. One is expressed in filaments and anthers (*TsSPH1*), a second only in anthers (*TsYUC6*) and the third in pistils (*TsBAHD*). This is remarkably consistent with the *Primula* GPA model of inheritance of distyly (Dowrick, 1956) except that the genes occur only in S-morph plants rather than having recessive allelic counterparts.

Of the three hemizygous genes, only *TsBAHD* is expressed in *S*-morph pistils. *TsBAHD* appears to be a BAHD acyltransferase as its amino acid sequence contains the highly conserved motifs characteristic of these enzymes (D'Auria, 2006). BAHD acyltransferases have a wide range of biochemical activities, and inferring specific biochemical function by sequence homology alone has proven difficult (Beekwilder *et al.*, 2004), but tantalizingly, top BLAST hits to *TsBAHD* include genes reported to acetylate (and inactivate) brassinosteroids (BRs) (Roh *et al.*, 2012; Wang *et al.*, 2012; Choi *et al.*, 2013). Recent studies in *Primula* have revealed that the *S*-locus gene controlling style length (the *G*-locus) is *CYP734A50*, and it has high homology to *BAS1,* a gene that possesses BR inactivating activity (Neff *et al.*, 1999; Huu *et al.*, 2016). BRs are regulators of cell expansion and several lines of evidence suggest that reduced BR concentrations underlie the reduced cell expansion in the short styles of the S-morph. S-morph styles have reduced BR concentrations relative to L-morph styles and virus-induced silencing of *CYP734A50* in S-morph flowers of *P. forbesii* resulted in a long-homostyle phenotype. Significantly, exogenous BR application 'rescues' the phenotype of S-morph flowers (Huu *et al.*, 2016).

The independent evolutionary origin of heterostyly in numerous angiosperm lineages has long been viewed as a remarkable example of convergence in form and function at the phenotypic level. That the results presented herein identify *TsBAHD* as the prime candidate for controlling style length in *Turnera* and the possibility that *TsBAHD*, a distinct class of enzyme from *CYP734A50*, also might function to reduce style length by inactivating brassinosteroids, suggests for the first time, the possibility that convergence may have occurred at the biochemical/physiological level. Further research is necessary to investigate this intriguing possibility.

The expression of *TsYUC6* is restricted to S-morph anthers. S-morph stamens possess longer filaments than those of the L-morph, and S-morph pollen is larger and possesses a biochemical 'recognition' specificity distinct from that of L-morph pollen. *TsYUC6* was so-named because BLASTP searches revealed high amino acid identity to the YUCCA gene family of flavin-dependent monooxygenases (Cheng *et al.*, 2006), the top BLAST hit from *Arabidopsis* being *AtYUC6*. Members of this gene family catalyze the second step in the two-step process of auxin synthesis from L-tryptophan (the first step being catalyzed by tryptophan aminotransferases). In *Arabidopsis, AtYUC6* functions redundantly with *AtYUC2*, both being expressed in stamen tissues and pollen, and have been proposed to be pivotal to auxin signalling events in these tissues. Double knockouts of these genes are defective in late anther development (Cheng *et al.*, 2006; Cecchetti *et al.*, 2008). Auxin is known to negatively control anther dehiscence and pollen maturation, but through an independent mechanism, it positively triggers filament elongation (Cecchetti *et al.*, 2008, 2017). A recent report further demonstrated that *AtYUC2* and *AtYUC6* are expressed in sporophytic microsporocytes, and are essential for the early and late stages of pollen development (Yao *et al.*, 2018).

The processes in which *AtYUC6* is involved are consistent with *TsYUC6* being the *S*-locus component that controls male characteristics of the S-morph (pollen size, number and physiology) and notably, pollen characteristics in distyly also are sporophytically regulated consistent with the expression of *AtYUC6* in sporophytic microspores. Significant further work is needed to investigate this possibility. Whether *TsYUC6* expression patterns at the cellular level mirror that of *AtYUC6* remains to be shown. Furthermore, is there also redundancy of YUC isoform expression in the anthers of *Turnera*, and how might this impact anther and pollen development?

TsSPH1 was expressed in filaments and anthers. It is a member of the large plant-specific S-protein homologue family (SPH1) defined by the discovery of the *Papaver rhoeas* pistil incompatibility gene, *PrsS* (Foote *et al.*, 1994; Kakeda *et al.*, 1998; Ride *et al.*, 1999). *PrsS* is the female determinant of the gametophytic self-incompatibility system and is expressed in the stigma with the encoded protein secreted by stigmatic papillae. It interacts with a pollen protein (PrpS) to signal an incompatibility response involving Ca^{2+}-dependent signalling and ultimately, programmed cell death of incompatible pollen (Wilkins *et al.*, 2005; De Graaf *et al.*, 2012). There are 84 members of the SPH1 family in *A. thaliana*, the functions of these largely floral-expressed genes are unknown (Ride *et al.*, 1999; Wheeler *et al.*, 2009, 2010) although they may be involved in extracellular signalling (De Graaf *et al.*, 2012). The *TsSPH1* gene, like *PrsS*, possessed a putative signal peptide, supporting the possibility that it is exported to the apoplast.

By contrast to *PrsS*, *TsSPH1* appears to be associated with male function as it is expressed in S-morph filaments and anthers, but not in styles. It is absent from all 10 L-morph deletion mutants as well as the short-homostyle deletion mutant but occurs in the long-homostyle mutant (LH1). Although it occurs in the short-homostyle mutant of *T. subulata* (TrinSH), it is not expressed. These results suggest that *TsSPH1* plays a role in filament elongation. We are continuing to investigate the inheritance of the short-homostyle phenotype, and the cause(s) of the lack of expression of *TsSPH1*. Finally, we have sequenced *TsSPH1* homologues from four long-homostyle species of *Turnera*, all of which appear to have functional copies of the gene which could be responsible for filament elongation in these species.

A possible mechanism for heteromorphic incompatibility

Finally, a key unanswered question relates to how the *S*-genes that we have identified might be involved with the breeding barrier known as heteromorphic incompatibility – a defining feature of the heterostylous syndrome in *Turnera* and the majority of other heterostylous species. It has been established that there is substantial interplay between auxin and brassinosteroid signalling pathways in the regulation of plant growth and development (Tian *et al.*, 2018). Therefore, if our hypotheses concerning male and female organ development in the floral morphs are valid, cross-pollination with pollen that developed under elevated auxin conditions (S-morph pollen) would only be compatible with 'high BR' pistils of the L-morph. Conversely, pollen that developed under reduced auxin conditions (L-morph pollen) would only be compatible with 'low BR' pistils of the S-morph. Thus, we propose that physiological differences resulting from contrasting hormone concentrations during development may be the underlying mechanism causing biochemical incongruity between pollen and pistils following intramorph pollination, and manifested as heteromorphic incompatibility. Our proposal provides a mechanistic basis for Lloyd & Webb's (1992a) hypothesis that incompatibility in heterostylous plants results from a mismatch between pollen and stylar environments. Future investigations manipulating hormone concentrations through exogenous applications, as well as knocking out the function of the hemizygous genes, and/or transforming and expressing them in L-morph plants, would allow tests of these hypotheses.

Acknowledgements

We thank Adrian Platts and Stephen Wright for advice, scripts and server access for genome assembly, Upendra Divisetty and Kira Neller for assistance with annotation, Deanna Harris and Kyra Dougherty, for technical assistance, Spencer Barrett for comments on the manuscript. We thank Gary Leveque (Genome Quebec), Derek Pouchnik and Mark Wildung (WSU Genomic Core Lab) for BAC sequencing and assembly. Supported by NSF grant MCB170034 (to AGM) 'Assembly and Annotation of the genome of the distylous species *Turnera subulata*' and an NSERC grant (JSS).

Author contributions

AGM, PDJC, JDJL and JSS conceived various aspects of the study. PMH and AGM carried out semi-quantitative RT-PCR. JDJL, PDJC and JSS screened BAC libraries and isolated DNA from some BACs, as well as genomic DNA. They were involved in initial end-sequencing of BACs for identification and chromosome walking. HJH, PDJC and JDJL carried out deletion mapping including primer design, PCR and SSCP analyses. HJH screened species and mutants for hemizygous genes. HJH and PDJC carried out RT-PCR. AGM sequenced two BACs. HJH, PDJC, JDJL and AGM were involved in Sanger sequencing of various genes. JSS assembled and annotated the *T. subulata* genome and BACs. AGM, JSS, HJH, PMH and PDJC wrote various parts of the manuscript and/or prepared various tables and figures.

ORCID

Andrew G. McCubbin https://orcid.org/0000-0003-2811-3544
Joel S. Shore https://orcid.org/0000-0003-2198-3133

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410.

Barrett SCH. 1992. Heterostylous genetic polymorphisms: model systems for evolutionary analysis. In: Barrett SCH, ed. *Evolution and function of heterostyly.* Berlin, Germany: Springer-Verlag, 1–29.

Barrett SCH. 2002. The evolution of plant sexual diversity. *Nature Reviews Genetics* **3**: 274–284.

Barrett SCH, Shore JS. 2008. New insights on heterostyly: comparative biology, ecology and genetics. In: Franklin-Tong VE, ed. *Self-incompatibility in flowering plants – evolution, diversity and mechanisms.* Berlin, Germany: Springer-Verlag, 3–32.

Bateson W, Gregory RP. 1905. On the inheritance of heterostylism in *Primula*. *Proceedings of the Royal Society of London. Series B, Biological Sciences* **76**: 581–586.

Bayer A, Ma X, Stöckigt J. 2004. Acetyltransfer in natural product biosynthesis—functional cloning and molecular analysis of vinorine synthase. *Bioorganic & Medicinal Chemistry* **12**: 2787–2795.

Beekwilder J, Alvarez-Huerta M, Neef E, Verstappen FWA, Bouwmeester HJ, Aharoni A. 2004. Functional characterization of enzymes forming volatile esters from strawberry and banana. *Plant Physiology* **135**: 1865–1878.

Bodmer WF. 1960. The genetics of homostyly in *Primula vulgaris*. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **242**: 517–549.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**: 578–579.

Burrows BA, McCubbin AG. 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**: 53–67.

Cecchetti V, Altamura MM, Falasca G, Costantino P, Cardarelli M. 2008. Auxin regulates *Arabidopsis* anther dehiscence, pollen maturation, and filament elongation. *Plant Cell* **20**: 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**: 1194–1207.

Chaisson MJ, Tesler G. 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics* **13**: 238.

Charlesworth D. 2016. The status of supergenes in the 21st century: recombination suppression in Batesian mimicry and sex chromosomes and other adaptations. *Evolutionary Applications* **9**: 74–90.

Charlesworth D, Charlesworth B. 1979. A model for the evolution of distyly. *American Naturalist* **114**: 467–498.

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 and Development* **20**: 1790–1799.

Chevreux B, Wetter T, Suhai S. 1999. In: Wingender E, ed. *Genome sequence assembly using trace signals and additional sequence information.* Hanover, Germany: GBF-Braunschweig. 45–56.

Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE *et al.* 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods* **10**: 563–569.

Choi S, Cho YH, Kim K, Matsui M, Son SH, Kim SK, Fujioka S, Hwang I. 2013. BAT1, a putative acyltransferase, modulates brassinosteroid levels in *Arabidopsis*. *The Plant Journal* **73**: 380–391.

Crosby JL. 1949. Selection of an unfavourable gene complex. *Evolution* **3**: 212–230.

Darwin C. 1877. *The different forms of flowers on plants of the same species.* London, UK: John Murray.

D'Auria JC. 2006. Acyltransferases in plants: a good time to be BAHD. *Current Opinion in Plant Biology* **9**: 331–340.

De Graaf BH, Vatovec S, Juárez-Diaz JA, Chai L, Kooblall K, Wilkins KA, Zou H, Forbes T, Franklin FC, Franklin-Tong VE. 2012. The *Papaver* self-incompatibility pollen *S*-determinant, *PrpS*, functions in *Arabidopsis thaliana*. *Current Biology* **22**: 154–159.

DISCOVAR. 2013. Assemble genomes, find variants. [WWW document] URL https://www.broadinstitute.org/software/discovar/ [accessed 26 January 2016].

Dowrick VPJ. 1956. Heterostyly and homostyly in *Primula obconica*. *Heredity* **10**: 219–236.

Fisher RA. 1941. The theoretical consequences of polyploid inheritance for the mid style form in *Lythrum salicaria*. *Annals of Eugenics* **11**: 31–38.

Foote HCC, Ride JP, Franklin-Tong VE, Walker EA, Lawrence Z, Franklin CH. 1994. Cloning and expression of a distinctive class of self-incompatibility (*S*) gene from *Papaver rhoeas* L. *Proceedings of the National Academy of Sciences, USA* **91**: 2265–2269.

Gilmartin PM. 2015. On the origins of observations of heterostyly in *Primula*. *New Phytologist* **208**: 39–51.

Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**: 491.

Huu CN, Kappel C, Keller B, Sicard A, Takebayashi Y, Breuninger H, Nowak MD, Bäurle I, Himmelbach A, Burkart M *et al.* 2016. Presence versus absence of CYP734A50 underlies the style-length dimorphism in primroses. *eLife* **5**: e17956.

Kakeda K, Jordan ND, Conner A, Ride JP, Franklin-Tong VE, Franklin FCH. 1998. Identification of residues in a hydrophilic loop of the *Papaver rhoeas* S protein that play a crucial role in recognition of incompatible pollen. *Plant Cell* **10**: 1723–1731.

Kappel C, Huu CN, Lenhard M. 2017. A short story gets longer: recent insights into the molecular basis of heterostyly. *Journal of Experimental Botany* **68**: 5719–5730.

Kent TV, Uzunovićć J, Wright SI. 2017. Coevolution between transposable elements and recombination. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **372**: 20160458.

Labonne JDJ, Goultiaeva A, Shore JS. 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**: 673–685.

Labonne JDJ, Hilliker AJ, Shore JS. 2007. Meiotic recombination in *Turnera* (Turneraceae): extreme sexual difference in rates, but no evidence for recombination suppression associated with the distyly (*S*) locus. *Heredity* **98**: 411–418.

Labonne JDJ, Shore JS. 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**: 101–111.

Labonne JDJ, Tamari F, Shore JS. 2010. Characterization of X-ray generated floral mutants carrying deletions at the *S*-locus of distylous *Turnera subulata*. *Heredity* **105**: 235–243.

Labonne JDJ, Vaisman A, Shore JS. 2008. Construction of a first genetic map of distylous *Turnera* and a fine-scale map of the *S*-locus region. *Genome* **51**: 471–478.

Leggett RM, Clavijo BJ, Clissold L, Clark MD, Caccamo M. 2014. NextClip: an analysis and read preparation tool for Nextera Long Mate Pair libraries. *Bioinformatics* **30**: 566–568.

Lewis D, Jones DA. 1992. The genetics of heterostyly. In: Barrett SCH, ed. *Evolution and function of heterostyly.* Berlin, Germany: Springer-Verlag, 129–150.

Li J, Cocker JM, Wright J, Webster MA, McMullan M, Dyer S, Swarbreck D, Caccamo M, Oosterhout CV, Gilmartin PM. 2016. Genetic architecture and evolution of the *S* locus supergene in *Primula vulgaris*. *Nature Plants* **2**: 16188.

Li J, Webster M, Dudas B, Cook H, Manfield I, Davies B, Gilmartin PM. 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**: 1–12.

Lloyd DG, Webb CJ. 1992a. The evolution of heterostyly. In: Barrett SCH, ed. *Evolution and function of heterostyly.* Berlin, Germany: Springer-Verlag, 151–178.

Lloyd DG, Webb CJ. 1992b. The selection of heterostyly. In: Barrett SCH, ed. *Evolution and function of heterostyly.* Berlin, Germany: Springer-Verlag, 179–208.

López A, Panseri A, Poggio L, Fernandez A. 2011. Nuclear DNA content in the polyploid complex *Turnera ulmifolia* (*Turnera* L., Passifloraceae). *Plant Systematics and Evolution* **296**: 225–230.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y *et al.* 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* **1**: 1–18.

Manzanares C, Barth S, Thorogood D, Byrne SL, 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**: 870–884.

Neff MM, Nguyen SM, Malancharuvil EJ, Fujioka S, Noguchi T, Seto H, Tsubuki M, Honda T, Takatsuto S, Yoshida S *et al.* 1999. *BAS1*: a gene regulating brassinosteroid levels and light responsiveness in *Arabidopsis*. *Proceedings of the National Academy of Sciences, USA* **96**: 15316–15323.

Nowak MD, Russo G, Schlapbach R, Huu CN, Lenhard M, Conti E. 2015. The draft genome of *Primula veris* yields insights into the molecular basis of heterostyly. *Genome Biology* **16**: 12.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**: 785–786.

Piper JG, Charlesworth B, Charlesworth D. 1986. Breeding system and evolution in *Primula vulgaris* and the role of reproductive assurance. *Heredity* **56**: 207–217.

Ride JP, Davies EM, Franklin FCH, Marshall DF. 1999. Analysis of *Arabidopsis* genome sequence reveals a large new gene family in plants. *Plant Molecular Biology* **39**: 927–932.

Roh H, Jeong CW, Fujioka S, Kim YK, Lee S, Ahn JH, Choi YD, Lee JS. 2012. Genetic evidence for the reduction of brassinosteroid levels by a BAHD acyltransferase-like protein in *Arabidopsis*. *Plant Physiology* **159**: 696–709.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212.

Smit AFA, Hubley R. 2015. *RepeatModeler Open-1.0.* [WWW document] URL http://www.repeatmasker.org [accessed 16 November 2017].

Tamari F, Khosravi D, Hilliker AJ, Shore JS. 2005. Inheritance of spontaneous mutant homostyles in *Turnera subulata* × *krapovickasii* and in autotetraploid *T. scabra* (Turneraceae). *Heredity* **94**: 207–216.

The UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research* **47**: D506–D515.

Thompson MJ, Jiggins CD. 2014. Supergenes and their role in evolution. *Heredity* **113**: 1–8.

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 LK, Johnson VE, 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**: 236.

Ushijima K, Ikeda K, Nakano R, Matsubara M, Tsuda Y, Kubo Y. 2015. Genetic control of floral morph and petal pigmentation in *Linum grandiflorum* Desf., a heterostylous flax. *Horticulture Journal* **84**: 261–268.

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**: 317–331.

Uyenoyama MK. 2005. Evolution under tight linkage to mating type. *New Phytologist* **165**: 63–70.

Wang L, Dong L, Zhang Y, 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**: 929–945.

Wang M, Liu X, Wang R, Li W, Rodermel S, Yu F. 2012. Overexpression of a putative *Arabidopsis* BAHD acyltransferase causes dwarfism that can be rescued by brassinosteroid. *Journal of Experimental Botonay* **63**: 5787–5801.

Weisenfeld NI, Yin S, Sharpe T, Lau B, Hegarty R, Holmes L, Sogoloff B, Tabbaa D, Williams L, Russ C *et al.* 2014. Comprehensive variation discovery in single human genomes. *Nature Genetics* **46**: 1350–1355.

Wheeler MJ, de Graaf BH, Hadjiosif N, Perry RM, Poulter NS, Osman K, Vatovec S, Harper A, Franklin FC, Franklin-Tong VE. 2009. Identification of the pollen incompatibility determinant in *Papaver rhoeas*. *Nature* **459**: 992–995.

Wheeler MJ, Vatovec S, Franklin-Tong VE. 2010. The pollen *S*-determinant of *Papaver*: comparisons with known plant receptors and protein ligand partners. *Journal of Experimental Botany* **61**: 2015–2025.

Wilkins KA, Bosch M, Haque T, Teng N, Poulter NS, Franklin-Tong VE. 2005. Self-incompatibility-induced programmed cell death in field poppy pollen involves dramatic acidification of the incompatible pollen tube cytosol. *Plant Physiology* **167**: 766–779.

Williamson RJ, Josephs EB, Platts AE, Hazzouri KM, Haudry A, Blanchette M, Wright SI. 2014. Evidence for widespread positive and negative selection in coding and conserved noncoding regions of *Capsella grandiflora*. *PLoS Genetics* **10**: e1004622.

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**: e1007397.

Yasui Y, Hirakawa H, Ueno M, Matsui K, Katsube-Tanaka T, Yang SJ, Aii J, Sato S, Mori M. 2016. Assembly of the draft genome of buckwheat and its applications in identifying agronomically useful genes. *DNA Research* **23**: 215–224.

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**: 1–9.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Splice variants of *TsYUC6* gene of *Turnera subulata*.

Fig. S2 Inferred amino acid alignments of the *TsSPH1* gene of *T. subulata* against homologues from *A. thaliana*, *Populus trichocarpa* and *Papaver rhoeas*.

Fig. S3 Inferred amino acid alignment of the *TsBAHD* gene of *T. subulata* against homologues from *A. thaliana*, *Populus trichocarpa*.

Fig. S4 Inferred amino acid alignment of *TsYUC6* gene of *T. subulata* against *Yuc6* of *Arabidopsis thaliana*.

Table S1 List of gene-specific primers used in experiments for *Turnera* spp.

Table S2 GenBank accession numbers for *TsSPH1* sequences of *Turnera* spp.

Table S3 Summary of sequence data and GenBank accession numbers for bacterial artificial chromosome clones and genome scaffolds from *Turnera subulata*.

Table S4 Genes on the *S*-locus haplotypes of *Turnera subulata*.

Table S5 Percent of *Turnera subulata S*-locus bacterial artificial chromosome clones and scaffolds comprised of transposable element-derived sequence.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.