the
geneticssociety

ARTICLE

Global transcriptome and gene co-expression network analyses on the development of distyly in *Primula oreodoxa*

Zhongtao Zhao[1] · Zhonglai Luo[1,2] · Shuai Yuan[1] · Lina Mei[1,3] · Dianxiang Zhang[1]

Abstract

Distyly is a genetically controlled flower polymorphism that has intrigued both botanists and evolutionary biologists ever since Darwin's time. Despite extensive reports on the pollination and evolution of distylous systems, the genetic basis and mechanism of molecular regulation remain unclear. In the present study, comparative transcriptome profiling was conducted in primrose *(Primula oreodoxa)*, the prime research model for heterostyly. Thirty-six transcriptomes were sequenced for styles at different stages and corolla tube in the three morphs of *P. oreodoxa*. Large numbers of differentially expressed genes (DEGs) were detected in the transcriptomes of styles across different morphs. Several transcription factors (TFs) and phytohormone metabolism-related genes were highlighted in S-morphs. A growing number of genes showed differential expression patterns along with the development of styles, suggesting that the genetic control of distyly may be more complicated than ever expected. Analysis of co-expression networks and module-trait relationships identified modules significantly associated with style development. CYP734A50, a key S-locus gene whose products degrade brassinosteroids, was co-expressed with many genes in the module and showed significant negative association with style length. In addition, crucial TFs involved in phytohormone signaling pathways were found to be connected with CYP734A50 in the co-expression module. Our global transcriptomic analysis has identified DEGs that are potentially involved in regulation of style length in *P. oreodoxa*, and may shed light on the evolution and broad biological processes of heterostyly.

Introduction

Heterostyly is characterized by the presence of two (distyly) or three (tristyly) distinct floral morphs within a species, with reciprocal positioning of stigma and anther heights (reciprocal herkogamy), and is one of the key contributors to diversity in angiosperms. Occurring in at least 199 genera (Naiki 2012), heterostyly is considered to be a mating system that prevents selfing and promotes outcrossing, as the flowers of one morph can produce seeds only when they are pollinated by other morphs and will reject pollen from the same morph (Barrett et al. 2000; Barrett 2002; Barrett and Shore 2008; Kappel et al. 2017; Lewis and Jones 1992; Naiki 2012).

Classic genetic studies in primroses have put forward a diallelic supergene model that corresponds to several dimorphic features of the distyly, including style length, pollen size, anther position, incompatibility recognition, and other appertained traits (Kappel et al. 2017; Kurian and Richards 1997; Lewis and Jones 1992). Recent studies have revealed that the S locus in S-morph is hemizygous, which may have arisen by segmental duplication and subsequent gene loss and neofunctionalization (Li et al. 2016). Two genes in the S-locus, CYP (CYP734A50) and GLO, likely control the style length and anther position, respectively (Burrows and McCubbin 2017; Huu et al. 2016; Li et al. 2016), and the former may also control the female

✉ Zhonglai Luo
luozhongl@scbg.ac.cn

✉ Dianxiang Zhang
dx-zhang@scbg.ac.cn

1 Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, The Chinese Academy of Sciences, Guangzhou 510650, China

2 Center for Plant Ecology, Core Botanical Gardens, Chinese Academy of Sciences, Guangzhou 510650, China

3 University of Chinese Academy of Sciences, Beijing 100049, China

incompatibility (Kappel et al. 2017). It is suggested that in these species, long homostyles were caused by gene mutation, rather than recombination between S-locus haplotypes (Li et al. 2016). Apart from S-locus genes, differentially expressed genes (DEGs) between flower morphs have also been investigated in several studies (Burrows and McCubbin 2018; Li et al. 2007; Manfield et al. 2005; McCubbin et al. 2006; Nowak et al. 2015). Some of the DEGs are tightly linked to the S locus (Li et al. 2007; Manfield et al. 2005), whereas others may be downstream components in the developmental processes of floral heteromorphy (McCubbin et al. 2006). In addition, a significant number of DEGs may be regulated by the S-locus genes in *Primula vulgaris* (Burrows and McCubbin 2018). Taken together, these findings suggest that distyly is not determined solely by S-locus genes, and rather more likely involves additional genes and metabolic pathways. However, the regulatory and metabolic pathways affected by S-locus genes are still largely unknown, which need to be further investigated. Much of the information underlying distylous developmental control, e.g., genetic regulation of floral syndromes and the molecular basis of self-incompatibility, is still elusive (Kappel et al. 2017).

Therefore, to further our understanding of this phenomenon, it is necessary to investigate different genes and their expressions across various distylous species. Given the distinctions in style length and growth rate between long-styled and short-styled individuals (hereafter referred to as L-morph and S-morph, respectively) with identical genetic backgrounds, it is possible to identify genes involved in the development of distyly through tissue- and stage-specific transcriptome analysis. The investigation of gene expression in distylous species, especially those with homostylous populations, could thus provide significant insight into the molecular mechanisms of distyly.

Primula oreodoxa, occurring in subalpine habitats of western Sichuan, China, is one such distylous–homostylous species. In *P. oreodoxa*, in addition to the typical distylous L- and S-morphs, a homostylous morph (H-morph) occurs in some populations. The H-morph of this species is monomorphic, with long styles and high anthers (long homostyly) (Yuan et al. 2017). Intriguingly, unlike most heterostylous *Primula* species, the homostylous and distylous morphs of *P. oreodoxa* are highly self-compatible (Yuan et al. 2017, 2019). Based on these features, *P. oreodoxa* is ideal for investigating patterns of morph-specific gene expression and molecular mechanisms of heterostyly. In this study, we sequenced transcriptomes of styles and corolla tubes (with androecium attached) of the L-, S-, and H-morphs at different developmental stages using RNA-seq. We identified DEGs between morphs that were likely to be involved in regulation of style development. The results of this study could provide valuable information for the mining of genes involved in the broad biological processes of distyly, and may provide a genetic resource for studying the development and evolution of distyly.

Materials and methods

Plant material and RNA preparation

Five developmental stages (Fig. 1a) of L-, S-, and H-morphs of *P. oreodoxa* were investigated among several populations during the flowering season in 2017. The population of H-morph was from E'mei Mountain, and the population of L- and S-morph was from Heizhugou Mountain, both of which are located in Sichuan Province. Style length and anther position were measured to the ~0.01 mm using digital calipers (573-S, Mitutoyo, Japan). For each of the five developmental stages, 10 flowers were sampled from 10 individuals for the measurement. Our investigation showed that the growing of styles in S-morphs generally slowed down at the big floral bud stage and stopped at the early flowering period, whereas the growing of styles in L- and H-morphs generally slowed down at the early flowering period and stopped at the complete flowering stage (over 9 mm in corolla tube length) (Fig. 1b). Therefore, for RNA-seq, materials from the first three stages of each morph were employed. The three developmental stages are small floral bud (3–4 mm in corolla tube length), big floral bud (5–6 mm in corolla tube length), and early flowering stage (7–8 mm in corolla tube length). Samples of styles and corolla tubes were collected separately. For styles at each stage from each morph, three biological replicates were collected. In *P. oreodoxa*, the stamens are adnate to the corolla tube, therefore the corolla tube and androecium were sampled as a whole in this study. For each morph, three biological replicates of mixed corolla tubes/androecia from all the three stages were collected. For each sample, materials were collected from ten individuals. In total, 36 samples were collected, including 27 style samples and nine corolla tube samples. All samples were quickly stabilized using RNAStay™ (Crystalgen, Commack, NY, USA) after collection. Total RNA was extracted using HiPure Universal RNA Mini Kit (Magen, Guangzhou, China) following the manufacturer's instructions, then stored at −80 °C until use.

Transcriptome sequencing and de novo assembly

Before sequencing, RNA integrity of each sample was assessed using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA); only samples with an RNA integrity number (RIN) above 6.5 were used for downstream analysis. A total of
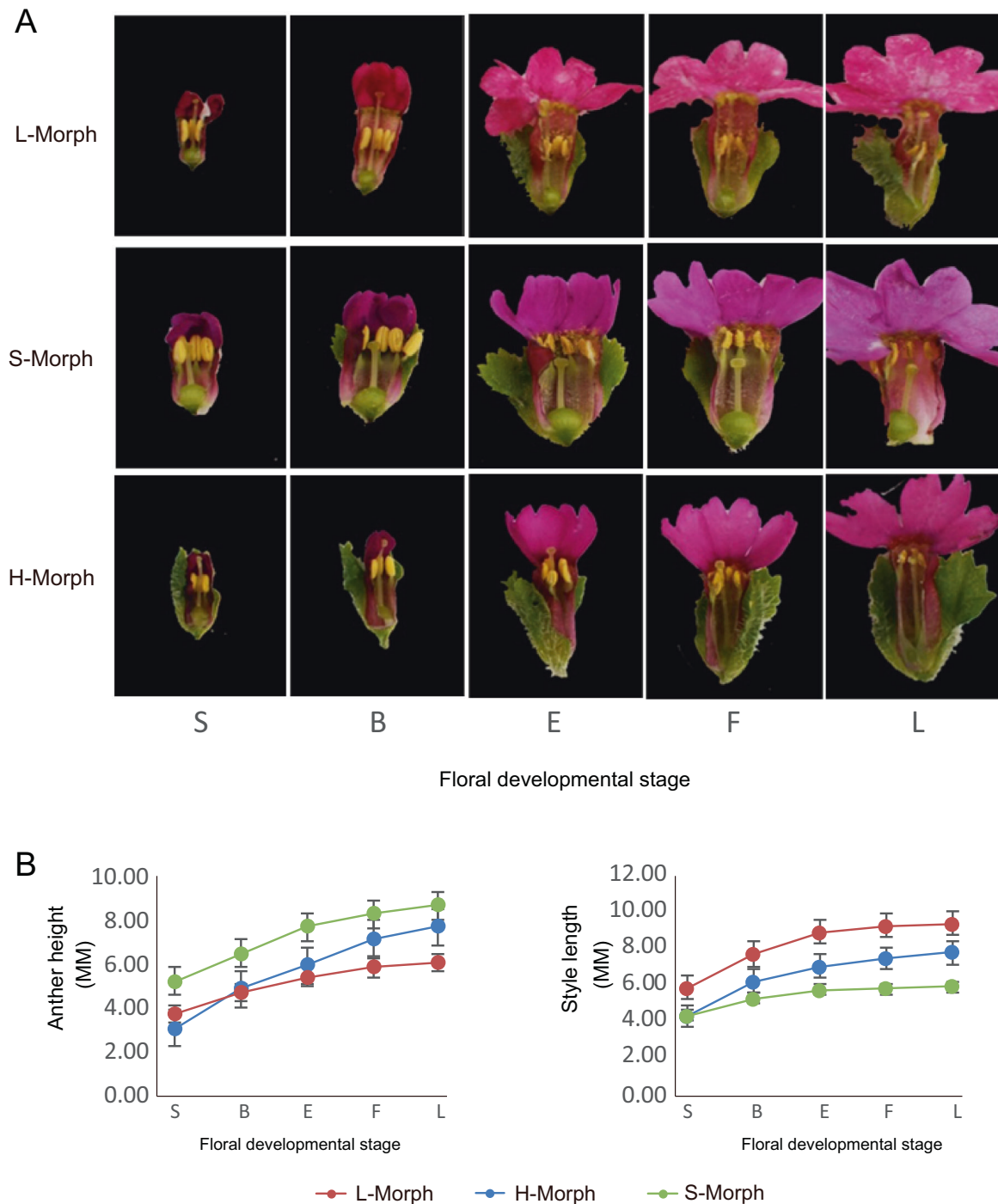
Fig. 1 Comparison of three different flower morphs representative of various developmental stages. **a** Bisected flowers at developmental stages as labeled in the L-, S-, and H-morphs. **b** Anther height and style length at each developmental stage. Mean and standard deviation are estimated with data from three independent replicates. S: small floral bud; B: big floral bud; E: early flowering stage; F: full bloom stage; L: late flowering stage

3 µg RNA per sample was used as input material for RNA sample preparations. Sequencing libraries were generated using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) in accordance with the manufacturer's recommendations. Library preparations were sequenced on an Illumina Hiseq 2000 platform and paired-end reads (raw reads) with the length of 150 bp were generated.

For assembly, raw reads in FASTQ format were first processed by removal of adapters, reads containing poly-N

(undetermined bases), and reads with low quality. Clean reads data were deposited into NCBI Sequence Read Archive (SRA) database with accession number PRJNA544868. All downstream analyses were performed on clean data with high quality. Transcriptome assembly (Unigenes) was carried out using Trinity (Grabherr et al. 2011) with min_kmer_cov set to 2, and all other parameters set to default.

Analysis of the differentially expressed unigenes

For each sample, clean reads were mapped to unigenes using Bowtie (Langmead et al. 2009). Expression abundances of the unigenes were then estimated using the RSEM program (Li and Dewey 2011). The number of fragments per kilobase of exon per million fragments mapped (FPKM) was used to estimate relative expression levels. Differential expression analysis was performed using read counts with the DESeq R package (1.10.1) (Anders and Huber 2010). The resulting p values were adjusted using the Benjamini–Hochberg method to control the false discovery rate. Genes with an adjusted p value < 0.05 were assigned as differentially expressed.

To identify genes potentially associated with phenotypic differences during style development among floral morphs in *P. oreodoxa*, we compared gene expression in styles and corolla tubes of L-, S-, and H-morphs. Pairwise comparisons were conducted between different developmental stages within morphs and between the same developmental stages of styles from different morphs, as well as for gene expression within corolla tubes (for pairwise comparisons please see additional file Table S1 and S2). In our analysis, gene expression showed at least twofold changes (up or down) in paired comparison, with adjusted $p < 0.05$, was considered as differential expression.

Gene annotation and enrichment analysis

Functional annotation of total genes was performed using the BLAST program with an E value cutoff of 1e−5 to search against the NR database of NCBI (ftp://ftp.ncbi.nih.gov/blast/db/) and the Swiss-Prot database (The UniProt, Consortium 2017). Functional domains/families were annotated by search against Pfam database (Finn et al. 2016) using the HMMER3 program (Eddy 2009) with an E value cutoff of 1e−10. Functional annotation by Gene Ontology (GO) terms was performed with AmiGO (Carbon et al. 2009). GO annotations describe gene functions at three ontology categories: Biological Process(BP); Cellular Components(CC); and Molecular Function (MF). KEGG annotation was performed using the KEGG Automatic Annotation Server (KAAS) (Moriya et al. 2007). GO enrichment analysis of DEGs was performed using the topGO packages in R, based on the Kolmogorov–Smirnov test. KOBAS (Xie et al. 2011) software was employed to test the statistical enrichment of DEGs in KEGG pathways.

Co-expression network construction

Co-expressed genes may have similar biological functions (Panteris et al. 2007). To explore genes possibly related to the style length, weighted gene co-expression network analysis was performed using the WGCNA package in R. DEGs obtained in pairwise comparisons across 27 style transcriptomes, including biological replicates, were used for network construction. Based on the criterion of approximate scale-free topology, a proper soft-thresholding power (Zhang et al. 2017) of 18 was chosen. Unigenes were hierarchically clustered according to the topological overlap matrix; the resulting gene dendrogram was then used for module detection using the dynamic tree cut method (minModuleSize = 30). Modules with similar expression profiles were merged with a height cut of 0.25. In the gene co-expression network, gene connectivity was measured by edge weight (ranging from 0 to 1), which was determined through topology overlap measure. Connectivity was defined as the sum of weights across all edges of a node. Only weights above 0.1 between any two genes were used for subsequent analysis.

Results

Transcriptome sequencing and de novo assembly

Overall, 27 transcriptomes of styles and nine transcriptomes of corolla tubes from S-, L-, and H-morphs were successfully sequenced. The clean data contain a total of approximately 1252.16 million paired-end reads, with the Q30 > 93% for all samples (Table S3). Combined assembly using total reads from all samples generated 129,872 unigenes. For these unigenes, mean length and N50 length are 851.43 and 1127 bp, respectively. Detailed assembly information is provided in Table 1. The total number of assembled unigenes may be overestimated due to the de novo assembly.

Functional annotation and classification of total unigenes

To annotate unigenes obtained in this study, unigene protein sequences were used to search against the following seven databases: NR, Swiss-Prot, GO, COG, KOG, Pfam, and KEGG. In all, 68,851 unigenes were successfully annotated in at least one database, with the NR and Swiss-prot databases contributing the highest proportions of annotations (summarized in Table S4).

A total of 5958 GO terms in three ontology categories (Biological Process, BP; Cellular Components, CC; and

Molecular Function, MF) were assigned to 37,423 unigenes. The most highly represented GO terms in the BP, CC, and MF categories are metabolic process (GO:0008152), integral component of membrane (GO:0016021), and ATP binding (GO:0005524), respectively. Overall, 16,427 unigenes were matched to 147 KEGG pathways, with the richest pathway being 'Metabolic pathways (ko01100),' followed by 'Biosynthesis of secondary metabolites (ko01110)' and 'Ribosome (ko03010)'. Many unigenes were also mapped to plant hormone metabolic pathways, including 391 genes involved in plant hormone signal transduction pathway (ko04075), 24 genes involved in brassinosteroid biosynthesis pathway (ko00905), and five genes involved in indole alkaloid biosynthesis pathway (ko00901).
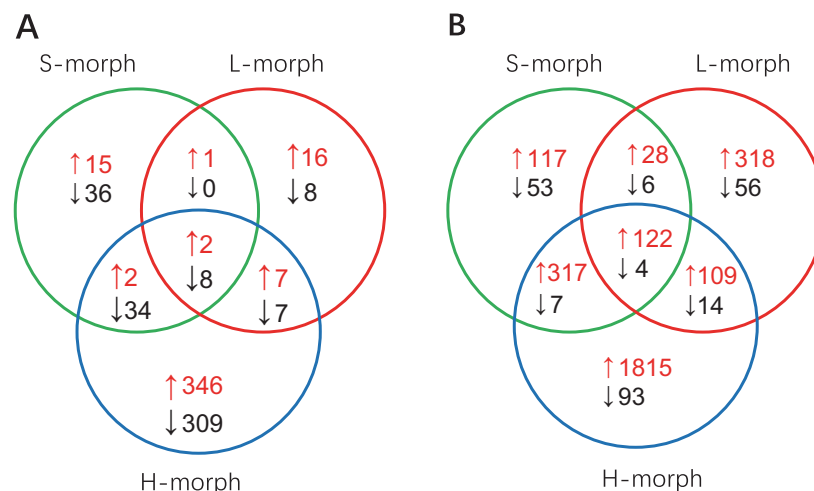
Analysis of differential expression of unigenes

A total of 31,576 unigenes were differentially expressed in at least one paired comparison (Figure S1). Overall, there are 730, 701, and 2940 unigenes differentially expressed across the different style developmental stages within the S-, L-, and H-morphs, respectively (Fig. 2).

Table 1 Summary of sequence assembly

Length range	Transcripts	Unigenes
300–500	124,407 (36.37%)	56,707 (43.66%)
500–1000	103,645 (30.30%)	41,317 (31.81%)
1000–2000	75,600 (22.10%)	21,558 (16.60%)
2000+	38,372 (11.22%)	10,290 (7.92%)
Total number	342,024	129,872
Total length	338,828,737	11,057,7294
N50 length	1408	1127
Mean length	990.66	851.43

Regardless of developmental stage, 3460 unigenes were differentially expressed during the development of styles between S-morph and L-Morph, 19227 genes were differentially expressed between S-morph and H-morph, and 15503 genes were differentially expressed between L-morph and H-morph. Many more genes were differentially expressed at early flowering stage in comparison with small floral buds and big floral buds in intermorph comparisons (Fig. 3a).

To explore genes potentially related to style development in H-morph, we identified unique DEGs in styles between H-morph and L-morph by removal of DEGs shared with S-morph vs. H-morph. Results showed that there are 1007, 1159, and 2280 unique DEGs at each developmental stage respectively in styles between H-morph and L-morph (Fig. 3b).

To further identify DEGs which may be more probably related to the control of style length, DEGs in styles at each developmental stage between S- and L-morphs (named SL DEGs) were examined at the corresponding stages in S-morph vs. H-morph. SL DEGs with significant differential expression at the corresponding stages in S-morph vs. H-morph (named core DEGs) were identified. Results showed that 733, 1380, and 1237 genes were differentially expressed at each developmental stage, respectively (Fig. 3c).

GO enrichment analysis of DEGs

To investigate functions of core DEGs at each developmental stage, we did GO enrichment analysis and enriched terms of BP were identified (Table S5). Results showed that, at both early and middle stages, the most significantly enriched GO term is "cell wall organization (GO:0071555)", followed by "external encapsulating structure organization (GO:0045229)". At early flowering stage, the most significantly enriched GO term is "plant-type cell wall organization". Although DEGs from all three

Fig. 2 Venn diagram of the numbers of DEGs in styles. **a** The number of DEGs in Big floral bud vs. Small floral bud within L-, S-, and H-morphs. **b** The number of DEGs in Early flowering stage vs Big floral bud within L-, S-, and H-morphs. Red and black arrows indicate up and downregulated genes, respectively

Fig. 3 Differential gene expression and comparative analysis of DEGs in styles at three developmental stages among three morphs. **a** Number of DEGs at different developmental stages in pairwise comparisons among three morphs. **b** Venn diagrams show common or unique DEGs at different developmental stages in H-morph vs. S-morph and H-morph vs L-morph. **c** Venn diagrams show common (core DEGs) or unique DEGs at different developmental stages in S-morph vs. L-morph and S-morph vs H-morph. Red arrows indicate upregulated genes and black arrows indicate downregulated genes
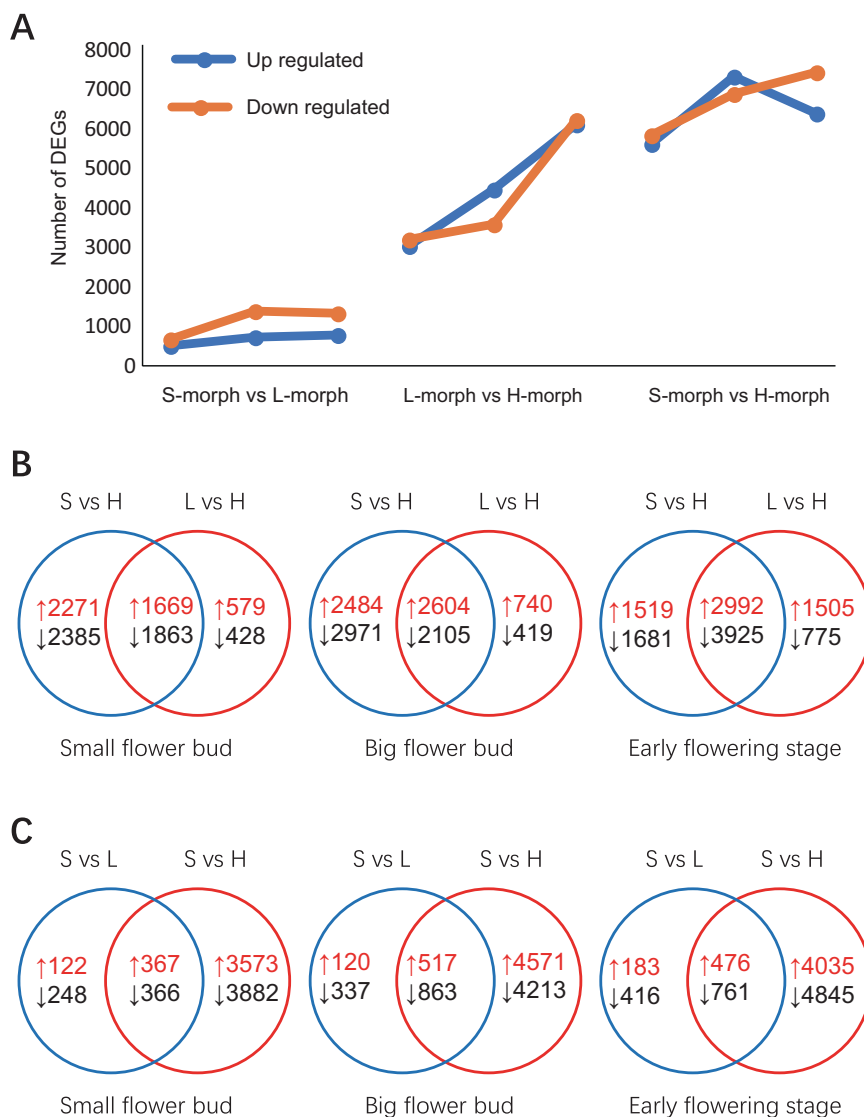
stages shared similar enriched BP terms, each of them has many unique enriched BP terms. Moreover, at the big floral bud stage, DEGs had many more top-enriched GO terms involved in phytohormone metabolic process than those at the other two stages (Figure S2).

Enriched GO terms were also examined based on total DEGs in the comparisons of H vs S and H vs L. Many enriched GO terms were overlapped in the comparisons at all three stages, whereas a large number of distinct GO terms were uniquely enriched at each stage (Figure S3). The regulated expression of genes under these biological processes may contribute to the biological characteristics of H-morph.

The expression of S-locus genes in *P. oreodoxa*

Li et al. (2016) reported five predicted genes in S locus, CCM, GLO, CYP, PUM, and KFB, all of which showed S-Morph-specific expression in *P. vulgaris*, whereas the

expression of the CYP gene was remarkably downregulated in long homostyle (Huu et al. 2016). Consistent with the report of Huu et al. (2016), our results showed that the CYP734A50 coding gene (Unigene DN49026_c0_g2, see Supplementary file 1 for sequences in S-morph and H-morph) in *P. oreodoxa* was highly expressed in the style of S-morph and was significantly reduced in the style of H-morph at each developmental stage (Table S6). In L-morph, it was expressed also at a very low level (FPKM«1). However, its low expression level in L-morph is most probably owing to the wrong mapping since that mismatches were allowed in our analysis. Interestingly, our results showed that the GLO expression was not S-morph specific because it was expressed in corolla tube/androecium of both S- and H-morphs. Transcripts of the CCM, PUM, and KFB orthologs were not detected in all 36 transcriptomes, suggesting that they are not expressed or expressed at very low level at sampling stages in this study.

Fig. 4 Heat map of DEGs related to the phytohormone metabolism. Within each row the expression data (FPKM) are normalized using Z score scaling. The blue bands indicate low gene expression and the red bands indicate high gene expression. Cluster analysis of DEGs within each phytohormone module is based on hierarchical clustering. S: small floral bud; B: big floral bud; E: early flowering stage
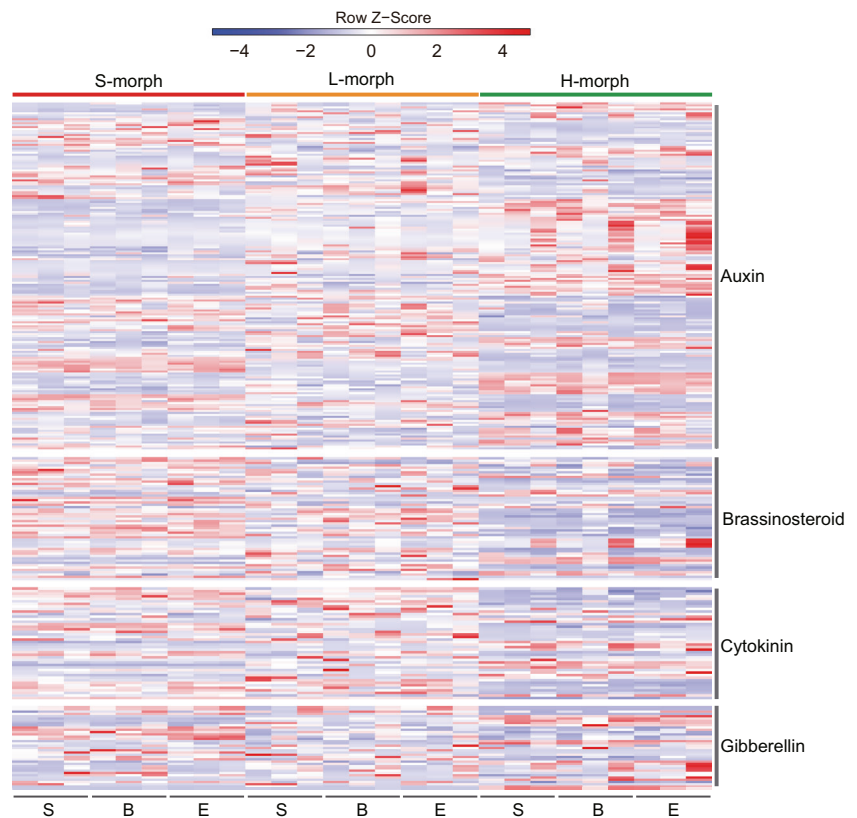
Differential expression of genes involved in phytohormone signaling pathways

Auxin, cytokinin, brassinosteroids, and gibberellins are considered to play vital roles in promoting plant growth (Depuydt and Hardtke 2011; McSteen and Zhao 2008). We examined the expression of genes related to phytohormone signaling pathways. In total, 334 genes related to the phytohormone metabolic processes were differentially expressed in at least one developmental stage of styles in pairwise comparisons among the three morphs (Fig. 4, Table S7). Many of these genes exhibited similar expression patterns in S- and L-morphs, which is distinct from that in H-morphs. Intriguingly, compared with L-morph and/or H-morph, expressions of a large number of genes responsible for the brassinosteroids metabolic process were upregulated through the growth of styles in S-morph (Fig. 4), including those genes involved in the brassinosteroid biosynthetic process (Table S7).

Differential expression of transcription factors

Transcription factors (TFs) play critical roles in the control of plant growth by regulating the gene transcription. In this study, we found that 438 TFs in styles were differentially expressed at each corresponding stage in intermorph

comparisons (Figure S4; Table S8). Interestingly, the majority of TFs also showed distinct expression patterns in H-morph in comparison with S- and L-morphs.

Transcriptional modules related to the style length

After modules with similar expression profiles were merged, 19 gene modules were identified. Subsequently, module-trait analysis was performed to detect the association between modules and style length. Two modules, II and XIX, were negatively and positively correlated with the style length, respectively (Fig. 5a). To estimate enriched biological functions for genes within the two style length correlated modules, we performed GO enrichment analysis. The module II was abundant in specific terms related to the regulation of gene expression, such as positive regulation of developmental process and multicellular organismal process (Fig. 5b). In contrast, the module XIX was abundant in specific terms related to plant development, such as tissue development and reproductive structure development (Fig. 5c). Importantly, the S-locus gene CYP734A50 that determines the style length dimorphism in primrose, was clustered into the module II. A large number of genes, including TFs and phytohormone metabolism-related genes, were predicted to be connected with CYP734A50 (Fig. 6a, b; Tables S9), implying that these genes may be involved in regulating the
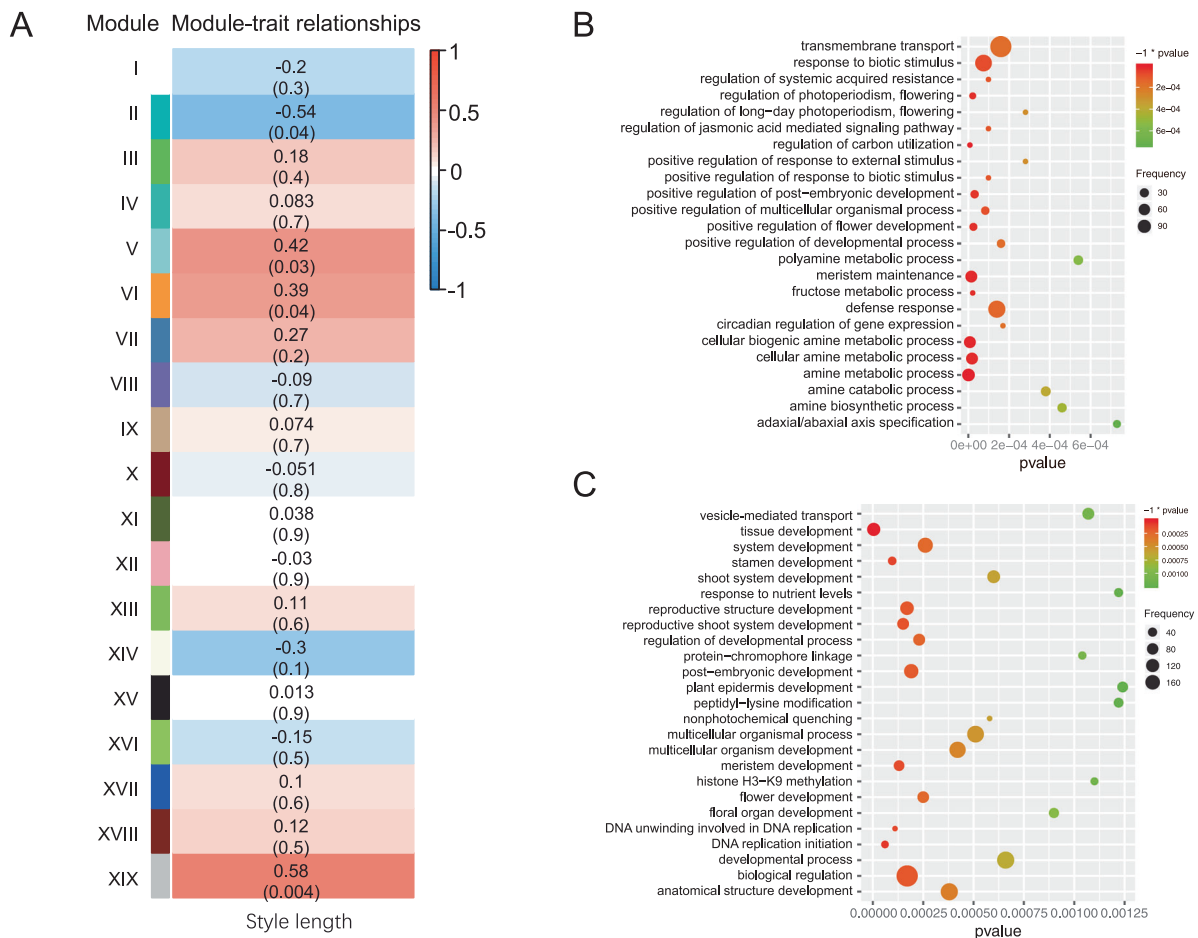
Fig. 5 Module-trait relationships and GO enrichment analysis of modules significantly correlated with the style length. **a** Relationships between style length and co-expressed gene modules. Each row corresponds to a module eigengene (correlation between a trait and a module). Each cell contains the corresponding correlation and *p* value. **b** GO (BP) enrichment analysis of module II that negatively correlated with the style length. **c** GO (BP) enrichment analysis of module XIX that positively correlated with the style length

distylous development in primrose. Interestingly, 32 of them were directly connected with CYP734A50 (Fig. 6b, Table S10).

Discussion

In this study, we performed comprehensive analysis of stage-specific transcriptomes of *P. oreodoxa*, a genetically important plant for the study of heterostyly (Yuan et al. 2019). Our results revealed that only a few genes showed significant differential expression during the style development across stages within each morph; many more intermorph DEGs were found in the transcriptomes of styles, however. Furthermore, WGCNA and module-trait analysis identified two modules associated with style length regulation, with one of positive correlation and the other one of negative correlation with the style length.

Within each morph, compared with the small floral bud, only a few DEGs were identified at the big floral bud stage,

whereas a greater number of DEGs were identified at the early flowering stage (Fig. 2). For example, in S-morphs, 98 genes were found to be differentially expressed at the big floral bud stage but 654 genes were differentially expressed at the early flowering stage. In *P. oreodoxa*, styles grow with a pronounced elongation from small floral bud stage to the big floral bud stage (Fig. 1b). At the early flowering stage, the style may also experience some other biological processes such as stigma maturity and pollen–stigma interactions, which may cause contrasting differential gene expressions compared with early stages (Zhang et al. 2017). Therefore, the elongation of styles within each morph is likely involved in only few DEGs, implying that the growth of the style (exclusive of stigma) is continuous and has no significant distinct stages. Interestingly, many more DEGs were detected in the comparison of H-morph vs. L-/S-morphs than L-morph vs. S-morph. Moreover, some groups of genes, e.g., phytohormone metabolism-related genes and TFs, showed different expression patterns in H-morph from L-morph and S-morph (Fig. 4, S4). The great difference in
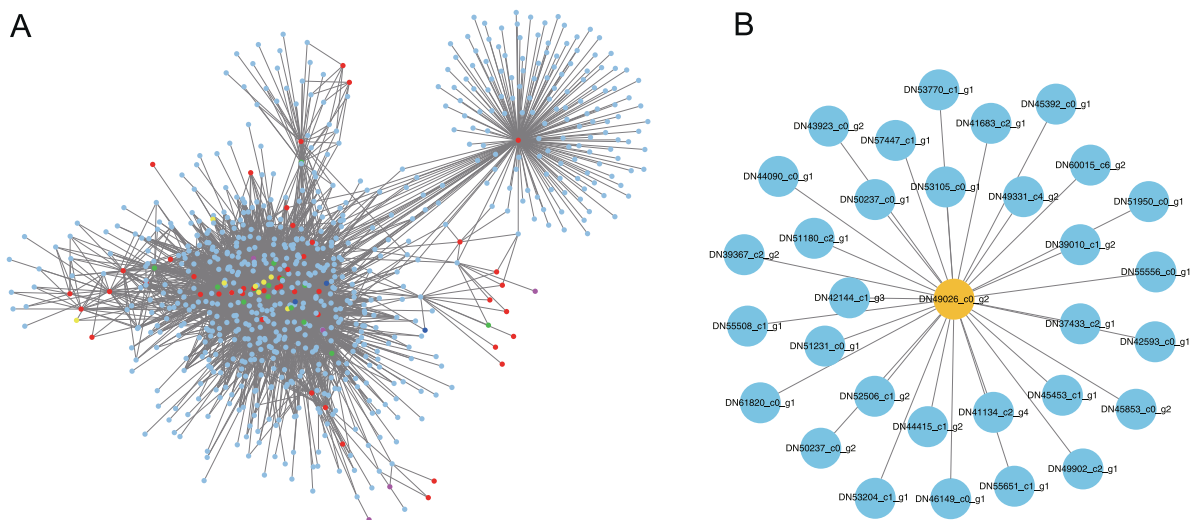
A

B

Fig. 6 Genes co-expressed with CYP734A50. **a** Genes directly or indirectly connected with CYP734A50 in the co-expression network. TFs (red) and genes related to metabolism of auxin (green), brassinosteroids (yellow), cytokinin (blue), and gibberellin (purple) are highlighted with different colors. **b** Genes directly connected with CYP734A50. Colored circles represent genes, and edges represent correlations among genes

gene expression between styles of H-morphs and those of L-/S-morphs may be owing to that the selfing H-morphs may be genetically much more different from the S-/L-morph plants than expected because of long-term inbreeding (Yuan et al. 2019). Furthermore, in H-morph population there may be many differential gene expressions in adaption to the unique environment they live in. Yet for all that, we cannot exclude the possibility that the development of H-morph styles seemed to be under different regulatory networks from those of S-/L-morphs.

During the development of flower buds, many genes showed special expression in either corolla tube/androecium or gynoecium (Figure S1). For example, the CYP734A50 was highly and continuously expressed only in the short style. In contrast, the tightly linked gene GLO, which is physically adjacent to the CYP734A50 in the S-locus (Li et al. 2016), showed special expression in corolla tube/androecia. When comparing the three stages of style development among S- and L-morphs, many more genes showed differentially expression at middle and late stages (Fig. 3), which is consistent with the expression patterns in *P. vulgaris* (Burrows and McCubbin 2018). At the early stage, the DEGs had the top-enriched GO terms involved in plant cell wall organization and phytohormone metabolic process (Figure S2). Comparing to the early stage, at the middle stage, despite those related to cell wall organization, DEGs were also significantly enriched in terms related to IABP (indoleacetic acid biosynthetic process) and the regulation of phytohormone metabolic/biosynthetic process. Many genes under these GO terms are involved in cell division/elongation and plant growth (Cheng et al. 2006; Song et al. 2009; Stepanova et al. 2008), suggesting that

different genes are involved in the style growth at different stages. Furthermore, more-distinct biological processes occurred later in development (early flowering stage), such as oxidation-reduction process and lipid metabolic process. Similarly, the transcriptome investigation in distylous *Lithospermum multiflorum* also showed distinct DEGs involved in numerous biological processes in the developing flower buds, including androecium and gynoecium (Cohen 2016). Taken together, all these findings suggested that the regulation of gene expression related to the control of distyly is much more complicated than ever expected. It is likely that the spatiotemporal expression of genes shapes the morphological features of heterostyly and determine the physiological aspects of heterostylous breeding system.

In the distylous primrose, the short morph of style is shaped by the action of the S-locus gene CYP734A50 (Huu et al. 2016). The product of CYP734A50 belongs to the cytochrome P450 CYP734A-family, which can degrade brassinosteroids (Ohnishi et al. 2006; Thornton et al. 2011) and its activity in S-morph styles would cause brassinosteroid degradation, limiting cell expansion, and style elongation (Huu et al. 2016). The reduction of castasterone (one of the major brassinosteroids) may cause the negative feedback regulation of genes involved in the BR synthesis, many of which were upregulated in the S-morph. For example, two of these genes DN56105_c4_g1 and DN49331_c4_g2 are, respectively, homologous to the *Arabidopsis thaliana* CYP90A1 and CYP90B1, which could catalyze the biosynthesis of brassinosteroid (Xu et al. 2015) and are negatively feedback controlled by brassinosteroids (Mathur et al. 1998; Salchert et al. 1998). Therefore, the presence of short styles in S-morphs may be

caused by the products of CYP734A50 acting as an antagonist of brassinosteroids instead of feedback down-regulation of brassinosteroid synthesis-related genes.

The large number of core DEGs during the early and middle development stage may be owing to the regulated expression of a cascade of genes, which may include genes regulating the expression of S-locus genes or being regulated by S-locus genes (Burrows and McCubbin 2018). Identification of gene co-expression modules associated with style development will be illuminative to understand the molecular mechanisms for the developmental control of distyly. Several modules showed significant positive association with the style length while one module (II) showed significant negative association with the style length. The gene CYP734A50 was co-expressed with many genes in module II (Fig. 6), suggesting that genes in module II are probably involved in the control of distyly development. Interestingly, several TFs were directly or indirectly associated with the CYP734A50 in the co-expression module. Many of these TFs are predicted to be related to the phytohormone signaling pathways. For example, the unigene DN42144_c1_g3 that was especially upregulated in S-morph is homologous to the *Arabidopsis* BIM1, a TF that could mediate brassinosteroid-regulated gene expression (Yin et al. 2005). Interactions between TFs and phytohormones exquisitely control the balance of cell proliferation/differentiation within the meristem to archive the proper growth and development of plant (Long and Benfey 2006). Hence, these TFs might be correlated with the transcriptionally regulation of the distylous development. The unigene DN44415_c1_g2 was predicted to code Cyclin-dependent kinase inhibitor, which regulates cell division and is crucial for plant growth (Han et al. 2005; Liu et al. 2008). Another S-morph especially upregulated unigene DN49902_c2_g1 is a homolog of *Arabidopsis* IAA9 (INDOLE-3-ACETIC ACID INDUCIBLE 9) that functions as a repressor of early auxin response genes at low auxin concentrations (Liscum and Reed 2002). In brief, we conclude that these tightly linked genes within one module may have related biological functions in the control of distyly and can be candidate genes for further study on the molecular regulatory mechanisms using CRISPR genome-editing system (Soda et al. 2018).

Accompanied with reciprocal herkogamy, heteromorphic self-incompatibility (HSI), which could limit selfing is also one of the key characteristics of heterostyly (Wedderburn and Richards 1990). The molecular mechanism of HSI is likely to be fundamentally different from that of homomorphic self-incompatibility and its self-recognition system is absent in heterostylous plants (Kappel et al. 2017; Lloyd and Webb 1992; Shimizu and Tsuchimatsu 2015). Attractively, several clues hint that in heterostyly, the style length and the female incompatibility may be controlled by the same gene, CYP734A50 (Kappel et al. 2017). However,

this model cannot explain the loss of female self-incompatibility in L-morph of *P. oreodoxa* (Yuan et al. 2019) and *P. chungensis* (Zhou et al. 2017), in which the CYP734A50 should be absent according to the hemizygous model (Li et al. 2016) and therefore the self-compatibility could not be generated from mutation of CYP734A50. The loss of female self-incompatibility in *P. oreodoxa* and *P. chungensis* may hint that polygenic modification is involved in the control of incompatibility even though the HSI is related to CYP734A50. The polygenic modification is also supported by evidences obtained from examples in *Turnera*, where the cross of different populations indicates that self-compatibility may be more likely the result of several genes, which interact with the distyly locus and result in self-compatibility in the presence of at least one S allele (Shore and Barrett 1986). Further comparative analysis of transcriptomes in *P. oreodoxa* from our work with those from self-incompatible species such as *P. vulgaris* (Burrows and McCubbin 2018) and *P. veris* (Nowak et al. 2015) will provide valuable information to understand the molecular basis of HSI.

Data archiving

Author Contributions DZ and ZL designed and supervised this study; ZZ, ZL, and SY proceeded this study and finished the manuscript; LM participated in the computational work.

Compliance with ethical standards

References

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11(10):R106

Barrett S, Shore J (2008) New insights on heterostyly: comparative biology, ecology and genetics. In: Franklin-Tong VE (ed) Self-incompatibility in flowering plants. Springer, New York, NY, p 3–32

Barrett SC (2002) The evolution of plant sexual diversity. Nat Rev Genet 3(4):274–84

Barrett SCH, Jesson LK, Baker AM (2000) The evolution and function of stylar polymorphisms in flowering plants. Ann Bot 85:253–65

Burrows B, McCubbin A (2018) Examination of S-Locus regulated differential expression in *Primula vulgaris* floral development. Plants (Basel) 7(2):pii: E38

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 Reprod 30(1):53–67

Carbon S et al. (2009) AmiGO: online access to ontology and annotation data. Bioinformatics 25(2):288–9

Cheng YF, Dai XH, Zhao YD (2006) Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in *Arabidopsis*. Genes Dev 20 (13):1790–99

Cohen JI (2016) *De novo* sequencing and comparative transcriptomics of floral development of the distylous species *Lithospermum multiflorum*. Front Plant Sci 7:1934

Depuydt S, Hardtke CS (2011) Hormone signalling crosstalk in plant growth regulation. Curr Biol 21(9):R365–73

Eddy SR (2009) A new generation of homology search tools based on probabilistic inference. Genome Inform 23(1):205–11

Finn RD et al. (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44(D1):D279–D85

Grabherr MG et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29 (7):644–U130

Han W et al. (2005) Overexpression of Arabidopsis ACK1 alters leaf morphology and retards growth and development. Biochem Biophys Res Commun 330(3):887–90

Huu CN et al. (2016) Presence versus absence of CYP734A50 underlies the style-length dimorphism in primroses. Elife 5:pii: e17956

Kappel C, Huu CN, Lenhard M (2017) A short story gets longer: recent insights into the molecular basis of heterostyly. J Exp Bot 68(21-22):5719–30

Kurian V, Richards AJ (1997) A new recombinant in the heteromorphy S supergene in *Primula*. Heredity 78:383–90

Langmead B et al. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10(3):R25

Lewis D, Jones DA (1992) The genetics of heterostyly. In: Barrett SCH (ed) Evolution and function of heterostyly. Springer, Berlin Heidelberg New York, NY, p 129–150

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323

Li J et al. (2007) Identification and characterization of pin and thrum alleles of two genes that co-segregate with the Primula S locus. Plant J 51(1):18–31

Li J et al. (2016) Genetic architecture and evolution of the S locus supergene in *Primula vulgaris*. Nat Plants 2(12):16188

Liu J et al. (2008) Targeted degradation of the cyclin-dependent kinase inhibitor ICK4/KRP6 by RING-type E3 ligases is essential for mitotic cell cycle progression during Arabidopsis gametogenesis. Plant Cell 20(6):1538–54

Liscum E, Reed JW (2002) Genetics of Aux/IAA and ARF action in plant growth and development. Plant Mol Biol 49(3-4):387–400

Lloyd DG and Webb, CJ (1992) The selection of heterostyly. In: Barrett S.C.H. (eds) Evolution and Function of Heterostyly. Monographs on Theoretical and Applied Genetics, vol 15. Springer, Berlin, Heidelberg, 179–207.

Long TA, Benfey PN (2006) Transcription factors and hormones: new insights into plant cell differentiation. Curr Opin Cell Biol 18 (6):710–4

Manfield IW et al. (2005) Molecular characterization of DNA sequences from the *Primula vulgaris* S-locus. J Exp Bot 56 (414):1177–88

Mathur J et al. (1998) Transcription of the *Arabidopsis* CPD gene, encoding a steroidogenic cytochrome P450, is negatively controlled by brassinosteroids. Plant J 14(5):593–602

McCubbin AG, Lee C, Hetrick A (2006) Identification of genes showing differential expression between morphs in developing flowers of *Primula vulgaris*. Sex Plant Reprod 19(2):63–72

McSteen P, Zhao Y (2008) Plant hormones and signaling: common themes and new developments. Dev Cell 14(4):467–73

Moriya Y et al. (2007) KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res 35:W182–5. Web Server issue

Naiki A (2012) Heterostyly and the possibility of its breakdown by polyploidization. Plant Species Biol 27(1):3–29

Nowak MD et al. (2015) The draft genome of *Primula veris* yields insights into the molecular basis of heterostyly. Genome Biol 16:12

Ohnishi T et al. (2006) Tomato cytochrome P450 CYP734A7 functions in brassinosteroid catabolism. Phytochemistry 67(17):1895–906

Panteris E et al. (2007) Mining pathway signatures from microarray data and relevant biological knowledge. J Biomed Inform 40 (6):698–706

Salchert K et al. (1998) Control of cell elongation and stress responses by steroid hormones and carbon catabolic repression in plants. Philos Trans R Soc Lond B Biol Sci 353(1374):1517–20

Shimizu KK, Tsuchimatsu T (2015) Evolution of selfing: recurrent patterns in molecular adaptation. Annu Rev Ecol Evol Syst 46:593–622

Shore JS, Barrett SCH (1986) Genetic modifications of dimorphic incompatibility in the *Turnera ulmifolia* L. Complex (Turneraceae). Can J Genet Cytol 28(5):796–807

Soda N, Verma L, Giri J (2018) CRISPR-Cas9 based plant genome editing: Significance, opportunities and recent advances. Plant Physiol Biochem 131:2–11

Song L et al. (2009) Membrane steroid-binding protein 1 (MSBP1) negatively regulates brassinosteroid signaling by enhancing the endocytosis of BAK1. Cell Res 19(7):864–76

Stepanova AN et al. (2008) TAA1-mediated auxin biosynthesis is essential for hormone crosstalk and plant development. Cell 133 (1):177–91

The UniProt, Consortium (2017) UniProt: the universal protein knowledgebase. Nucleic Acids Res 45(D1):D158–D69

Thornton LE, Peng H, Neff MM (2011) Rice CYP734A cytochrome P450s inactivate brassinosteroids in *Arabidopsis*. Planta 234(6):1151–62

Wedderburn F, Richards AJ (1990) Variation in within-Morph Incompatibility Inhibition Sites in Heteromorphic *Primula* L. New Phytol 116(1):149–62

Xie C et al. (2011) KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res 39:W316–W322

Xu J, Wang XY, Guo WZ (2015) The cytochrome P450 superfamily: Key players in plant development and defense. J Integr Agric 14 (9):1673–86

Yin Y et al. (2005) A new class of transcription factors mediates brassinosteroid-regulated gene expression in *Arabidopsis*. Cell 120(2):249–59

Yuan S et al. (2017) Ecological correlates and genetic consequences of evolutionary transitions from distyly to homostyly. Ann Bot 120 (5):775–89

Yuan S et al. (2019) Genetics of distyly and homostyly in a self-compatible *Primula*. Heredity 122(1):110–119

Zhang T et al. (2017) Time-course transcriptome analysis of compatible and incompatible pollen-stigma interactions in *Brassica napus* L. Front Plant Sci 8:682

Zhou W et al. (2017) Phylogeographic insights on the evolutionary breakdown of heterostyly. New Phytol 214(3):1368–80