PCP
PLANT & CELL PHYSIOLOGY

Regular Paper

High-Throughput Transcriptome Analysis of the Leafy Flower Transition of *Catharanthus roseus* Induced by Peanut Witches'-Broom Phytoplasma Infection

Li-Yu Daisy Liu[1,8], Hsin-I Tseng[2,8], Chan-Pin Lin[2,3,8], Yen-Yu Lin[2,3], Yuan-Hung Huang[4], Chien-Kang Huang[5], Tean-Hsu Chang[2,3] and Shih-Shun Lin[2,6,7,*]

[1]Department of Agronomy, National Taiwan University, 1, Sec. 4, Roosevelt Rd., Taipei, Taiwan, 106
[2]Institute of Biotechnology, College of Bioresources and Agriculture, National Taiwan University, 81, Chang-Xing ST., Taipei, Taiwan 106
[3]Department of Plant Pathology and Microbiology, College of Bioresources and Agriculture, National Taiwan University, 1, Sec. 4, Roosevelt Rd., Taipei, Taiwan 106
[4]Department of Computer Science and Information Engineering, National Taiwan University, 1, Sec. 4, Roosevelt Rd., Taipei, Taiwan, 106
[5]Department of Engineering Science and Ocean Engineering, National Taiwan University, 1, Sec. 4, Roosevelt Rd., Taipei, Taiwan, 106
[6]Genome and Systems Biology Degree Program, National Taiwan University, 1, Sec. 4, Roosevelt Rd., Taipei, Taiwan 106
[7]Agricultural Biotechnology Research Center, Academia Sinica, 128, Sec. 2, Academia Rd., Taipei, Taiwan 115
[8]These authors contributed equally to this work.
*Corresponding author: E-mail, linss01@ntu.edu.tw; Fax, +886-2-33666001.
(Received January 1, 2014; Accepted January 29, 2014)

Peanut witches'-broom (PnWB) phytoplasma are obligate bacteria that cause leafy flower symptoms in *Catharanthus roseus*. The PnWB-mediated leafy flower transitions were studied to understand the mechanisms underlying the pathogen–host interaction; however, our understanding is limited because of the lack of information on the *C. roseus* genome. In this study, the whole-transcriptome profiles from healthy flowers (HFs) and stage 4 (S4) PnWB-infected leafy flowers of *C. roseus* were investigated using next-generation sequencing (NGS). More than 60,000 contigs were generated using a de novo assembly approach, and 34.2% of the contigs (20,711 genes) were annotated as putative genes through name-calling, open reading frame determination and gene ontology analyses. Furthermore, a customized microarray based on this sequence information was designed and used to analyze samples further at various stages of PnWB infection. In the NGS profile, 87.8% of the genes showed expression levels that were consistent with those in the microarray profiles, suggesting that accurate gene expression levels can be detected using NGS. The data revealed that defense-related and flowering gene expression levels were altered in S4 PnWB-infected leafy flowers, indicating that the immunity and reproductive stages of *C. roseus* were compromised. The network analysis suggested that the expression levels of >1,000 candidate genes were highly associated with *CrSVP1/2* and *CrFT* expression, which might be crucial in the leafy flower transition. In conclusion, this study provides a new perspective for understanding plant pathology and the mechanisms underlying the leafy flowering transition caused by host–pathogen interactions through analyzing bioinformatics data obtained using a powerful, rapid high-throughput technique.

Keywords: *Catharanthus roseus* • Microarray • Leafy flower • Next-generation sequencing • Peanut witches'-broom phytoplasma • Transcriptome.

Abbreviations: AFR, *AUXIN RESPONSE FACTOR*; *AG*, *AGAMOUS*; *AGL19*, *AGAMOUS-like 19*; *AP1*, *APETALA1*; *AS1*, *ASYMMETRIC LEAVES1*; CDS, coding sequence; DL, diseased leaf; *FT*, *FLOWER LOCUS T*; GO, gene ontology; HD/ZIP, homeodomain/leucine zipper; HF, healthy flower; HL, healthy leaf; *KAN1*, *KANADI1*; KEGG, Kyoto Encyclopedia of Genes and Genomes; *KNOX*, *KNOTTED1*-like homeobox; ncRNA, non-coding RNA; NGS, next-generation sequencing; ORF, open reading frame; *PHB*, *PHABULOSA*, *PHV*, *PHAVOLUTA*, *PI*, *PISTILLATA*; PnWB, peanut witches'-broom; *PR1*, *PATHOGENESIS-RELATED GENE 1*; *REV*, *REVOLUTA*; *ROT3*, *ROTUNDIFOLIA3*; S4, stage 4; SAM, shoot apical meristem; SEM, scanning electron microscopy; *SEP3*, *SEPALLATA3*; *SOC1*, *SUPPRESSOR OF OVEREXPRESSION OF CO1*; *SVP*, *SHORT VEGETATIVE PHASE*; TIA, terpenoid indole alkaloid; UTR, untranslated region; *ZWI*, *ZWICHEL*.

The raw transcriptome reads reported in this paper have been submitted to the NCBI Short Read Archive under accession numbers SRR1144633 (healthy flowers of *C. roseus*) and SRR114464 (S4 leafy flowers of *C. roseus*). The transcriptome contig sequences are available in the ContigViews database (www.contigviews.bioagri.ntu.edu.tw)

Plant Cell Physiol. 55(5): 942–957 (2014) doi:10.1093/pcp/pcu029, available online at www.pcp.oxfordjournals.org

Introduction

Incomplete genomic information is a limiting factor for further investigating the mechanisms of unique phenomena in non-model organisms. After the publication of a next-generation sequencing (NGS) study by Margulies et al. (2005), many high-throughput approaches based on this concept have been developed in various areas that include de novo transcriptome analysis (Mardis 2008, Morozova and Marra 2008). De novo transcriptome analysis with NGS has revolutionized the study of gene expression profiles in non-model organisms under various conditions or treatments. For instance, gene expression levels have been analyzed in a wide range of vegetative and reproductive tissues in *Phalaenopsis aphrodite* (Su et al. 2011a). Approximately 42,000 contigs with an average length of 875 bp were obtained after de novo assembly, and the expression profiles provided crucial genetic information for *P. aphrodite* development studies (Su et al. 2011a). Moreover, the medicinal *Catharanthus roseus* produces >150 terpenoid indole alkaloids (TIAs), such as vinblastine and vincristine, which are major commercial anticancer molecules (Van Moerkercke et al. 2013). The omics approaches were designed based on transcriptomic and metabolomic profiles to identify the unknown enzymes needed to engineer TIA pathways (Van Moerkercke et al. 2013). These studies indicated that systems biology studies in non-model organisms can be conducted by NGS.

Catharanthus roseus was also used as a host for an obligate phytoplasma infection in plant pathology research (Yang 1983). Various species of phytoplasma, such as peanut witches'-broom (PnWB) and aster yellow witches'-broom, cause leafy flower symptoms in *C. roseus* and Arabidopsis, respectively (Chen and Lin 2011, MacLean et al. 2011). Several studies have indicated that the phytoplasma-mediated virescence and phyllody (herein referred to as leafy flower) might be the result of abnormal endogenous phytohormone signaling (Davey and Van Staden 1981). For instance, cytokinin levels decreased in leaves but increased in floral tissues in phytoplasma-infected *C. roseus* (Lee et al. 2000b). Furthermore, Hoshi et al. (2009) discovered the *TENGU* gene, encoding a virulence factor in phytoplasma that can induce proliferation and dwarfism in *TENGU*-expressing Arabidopsis by interfering with auxin-related gene expression. Also, indole-3-butyric acid induced recovery of the phytoplasma-mediated internode-shortening symptoms in *C. roseus*, suggesting that anti-phytoplasma activity caused by auxin is involved (Curković Perica 2008). These results suggested that auxin signaling and hormone balance play roles in phytoplasma-mediated host symptom development.

The ABC model is widely used to explain floral organ formation. The A-class genes, *APETALA1* (*AP1*) and *AP2*, confer sepal identity in Arabidopsis, whereas the B-class genes, *AP3* and *PISTILLATA* (*PI*), promote petal development (Ferrándiz et al. 2000, Wollmann et al. 2010). The C-class gene *AGAMOUS* (AG) specifies stamen and carpel identity (Wollmann et al. 2010). During the flowering stage transition, the *SUPPRESSOR OF OVEREXPRESSION OF CO1* (*SOC1*), which contains a MADS-box domain, plays a positive regulatory role in multiple flowering pathways, including the vernalization, photoperiod, autonomous, gibberellin and aging pathways in Arabidopsis (Lee et al. 2000a, Samach et al. 2000, Moon et al. 2003, Wang et al. 2009a, Srikanth and Schmid 2011). In addition, SHORT VEGETATIVE PHASE (SVP) is a flowering repressor that suppresses the expression of the *FLOWER LOCUS T* (*FT*) gene (Li et al. 2008).

Pracros et al. (2006) demonstrated that floral abnormalities are associated with the down-regulation of the floral development genes *LeWUSCHEL*, *LeCLAVATA1* and *LeDEFICIENS* upon stolbur phytoplasma infection in tomato. The expression levels of the hydrangea ortholog B-class and C-class flower genes were found to be repressed when plants were infected with Japanese hydrangea phyllody phytoplasma (Kitamura et al. 2009). Su et al. (2011b) revealed that *C. roseus* infected with PnWB exhibited down-regulation of the Arabidopsis *MADS-BOX* ortholog genes *CrAG*, *SEPALLATA3* (*CrSEP3*), *CrAP3* and *CrPI*. The expression levels of several flowering-related genes were altered by phytoplasma; however, differential display and subtraction cloning approaches were used extensively in these studies, and a thorough understanding of the mechanisms underlying the leafy flower transition during phytoplasma infection was difficult to obtain.

During leaf development, *KNOTTED1*-like homeobox (*KNOX*) genes, including *KNAT1* and *KANT2*, are expressed in shoot apical meristems (SAMs) and involved in generating leaf organs (Ori et al. 2000). *KNOX* genes negatively regulate gibberellin in the meristem, thus keeping cells from expanding and differentiating (Sakamoto et al. 2001). In leaf shape development, the *PHABULOSA* (*PHB*), *PHAVOLUTA* (*PHV*) and *REVOLUTA* (*REV*) genes, which encode Class III homeodomain/leucine zipper (HD/ZIP) transcription factors, are expressed in the adaxial leaf domain and are required for specifying adaxial identity (McConnell et al. 2001, Prigge et al. 2005). *ASYMMETRIC LEAVES1* and 2 (*AS1* and 2) facilitate the specification of leaf adaxial identity. *AS1* encodes a protein that contains an R2-R3 MYB-domain, whereas *AS2* encodes a plant-specific leucine zipper protein that associates with *AS1* (Xu et al. 2003). *KANADI1* (*KAN1*) negatively regulates *AS2* expression, and the loss of this gene family causes upward curled leaves and precocious abaxial trichomes (Wu et al. 2008). Abaxial fate on the abaxial side of the leaf requires the *AUXIN RESPONSE FACTOR* (*AFR*) family genes (Pekker et al. 2005, Kelley et al. 2012). *ZWICHEL* (*ZWI*) encodes a kinesin motor molecule involved in trichome branching (Folkers et al. 2002). *ROTUNDIFOLIA3* (*ROT3*) is involved in determining the sizes of leaf cells (Kim et al. 1998).

In this study, whole-transcriptome profiles of *C. roseus* were analyzed using NGS, resulting in the identification of 20,711 genes. These sequences were used to design a customized microarray for verifying the gene expression levels during a time-course of PnWB infection. The gene expression accuracy in the NGS profiles was evaluated by comparing the NGS data with the microarray profiles. In addition, an entire leafy flower transition network was generated by connecting ABC model genes, and interesting genes involved in the PnWB-mediated leafy flower transition were discovered. The mechanism of the leafy flower transition in *C. roseus* during PnWB infection was investigated.

Results

PnWB-mediated leafy flower symptoms in *C. roseus*

PnWB-infected *C. roseus* flowers exhibit various floral abnormalities, including discoloration, virescence and phyllody, depending on the number of days after infection (Su et al. 2011b). **Fig. 1A** shows five categories of floral abnormalities in PnWB-infected *C. roseus* plants. Stage 1 (S1) flowers show abnormal petals with partial virescence compared with healthy flowers; stage 2 (S2) flowers show smaller and abnormal virescent petals. Phyllody malformations develop in stage 3 (S3) flowers, which are fully virescent and have partially pink petals. A leaf vein pattern was observed in the stage 4 (S4) flowers, which also had phyllodic petals. Finally, fully phyllodic petals characterize stage 5 (S5) flowers. The corolla tube was shorter in infected floral organs than in healthy organs, and it shortened as the infection developed (S1 to S5). A swollen ovary was observed in S3 and S4, and a cracked swollen ovary was occasionally observed. Moreover, the flowers developed whorl-like structures and small leaves in S5.

Scanning electron microscopy (SEM) images revealed that the surface of the healthy flower (HF) petals contained various flocculent structures. However, the morphology changed when *C. roseus* was infected with PnWB (**Fig. 1B**). The flocculent structure of the S1 petal developed an explanted structure, and its cell shape became polygonal (**Fig. 1B**). In the petals of the S2 and S3 flowers, the epidermal cells possessed an abnormal shape accompanied by a hairstreak structure on the surface (**Fig. 1B**). Moreover, stomata and guard cells appeared on the leaf-like petals at S4 and S5. Finally, the surface of the S5 petal was identical to that of a true leaf (**Fig. 1B**). The results indicated that endogenous gene regulation in PnWB-infected *C. roseus* might cause leafy flower formation. Eventually, the petals became leaves. In addition, trichomes were observed in the S4 leafy flowers that were similar to those in healthy leaves, whereas no trichome-like structures were observed in healthy flowers (**Fig. 1C**). These results suggest that PnWB infection caused the reversal of host plants from the reproductive to the vegetative stage.

Whole-transcriptome analysis and annotations

To obtain a comprehensive gene expression profile of *C. roseus*, the transcriptomes of HF and S4 flowers were analyzed using NGS (**Fig. 2A**). We obtained 27,494,876 raw HF reads and 25,995,408 raw S4 reads, and 81.3% of the reads were paired (**Table 1**). These paired-end reads from the HF and S4 samples were mixed for de novo assembly, and the results showed that 85.9% of the paired-end reads were successfully assembled as 60,580 contigs (**Table 1**).

The 60,580 contigs were analyzed using the Basic Local Alignment Search Tool (BLAST) with an *Arabidopsis thaliana* coding sequence (CDS) database (TAIR 10) and an EMBL CDS database for name-calling and open reading frame (ORF) annotation (**Fig. 2B**). According to the BLAST results, the best hit for the reference ORF sequence was used for alignment with the corresponding predicted amino acid sequence of the contig. The rule-based predictor in the ContigViews NGS platform (http://www.contigviews.bioagri.ntu.edu.tw) predicted the 'Start' and 'Stop' codons for each contig sequence according to the alignment result and the start and stop codon positions of the reference sequence to help curate the data and facilitate collaboration between researchers. The rule-based predictor can complete ORF prediction in a few hours, depending on the number of contigs (data not shown). The automatic ORF analysis resulted in three categories: complete ORFs, partial ORFs and undefined contigs (**Fig. 2B**). After the ORF prediction, the researcher must conduct a final evaluation to rule out prediction errors. The final results were integrated into ContigViews to develop a *C. roseus* transcriptome database.

The ORFs of the annotated genes were further analyzed with the gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Interpro databases by Blast2GO software to annotate the GO terms, catalytic pathways and domains for each gene (**Fig. 2A**). The GO terms indicated that many of the annotated genes were involved in the response to stress and abiotic/biotic stimulus (**Supplementary Fig. S1**). Moreover, the gene set enrichment analysis (GSEA) results indicated that many genes were involved in photosynthesis, chloroplast development and energy metabolic processing (**Table 2**; **Supplementary Table S1**). Therefore, these results suggested that PnWB infection triggers the host defense response and initiates leafy development gene expression. Interestingly, the genes related to energy metabolic processing were significantly highlighted by these assays, suggesting that obligate phytoplasma require the host energy metabolic system for survival. The GO, KEGG and Interpro domain prediction results of *C. roseus* genes were integrated in the ContigViews database.

Annotated genes and undefined contigs

The three categories of ORF annotations, complete ORFs, partial ORFs and undefined contigs, are summarized in **Fig. 3A**. A contig with a complete ORF contains a start and stop codon and can be translated into a complete protein sequence. In addition, most of the complete contigs contained 5′- and 3′-untranslated regions (UTRs) (**Fig. 3A**, panel i). A contig with a partial ORF lacks a start codon, a stop codon or both codons, and it has an incomplete protein sequence similar to that of the reference (**Fig. 3A**, panel ii). The ORF annotation of

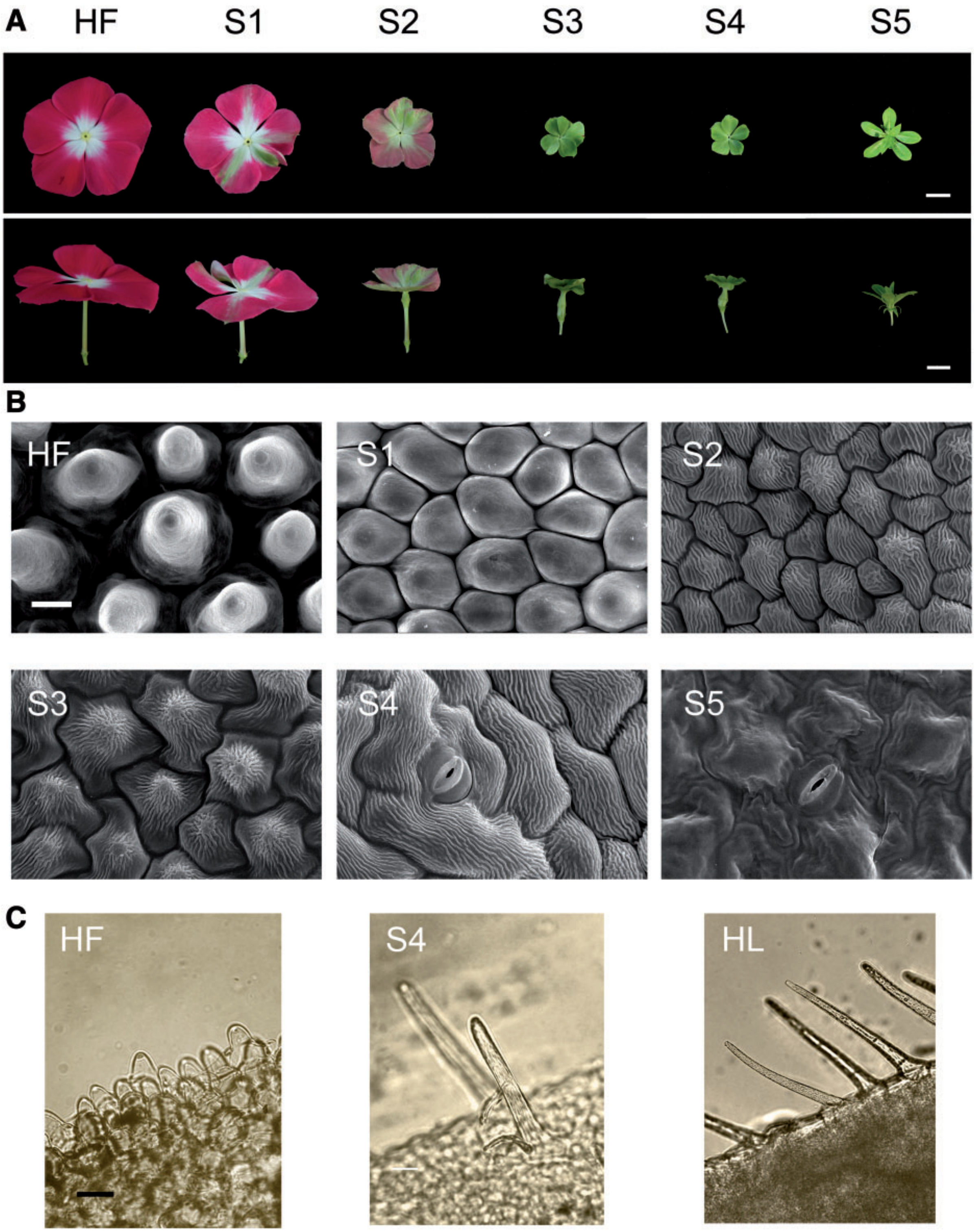

Fig. 1 Five categories of peanut witches'-broom (PnWB)-infected leafy flower symptoms compared with healthy flowers of *Catharanthus roseus*. (A) The flower morphology of PnWB-infected *C. roseus* was classified into five categories. S1 (stage 1), abnormal petals and partial virescence; S2 (stage 2), small and virescent petals; S3 (stage 3), fully virescent and partially pink petals; S4 (stage 4), phyllodic petals with leaf vein patterning; and S5 (stage 5), fully phyllodic petals. HF, healthy flower. Scale bar = 1 cm. (B) Adaxial epidermis of the HF and PnWB-infected (S1–S5) leafy flower petals. In addition to the shape and pattern changes of the epidermal cells within each sample, the existence of stomata within S4 and S5 is shown. The photos were taken at ×1,6000 magnification using SEM. Scale bar = 10 µm. (C) Trichomes of healthy leaves (HLs) and S4 PnWB-infected leafy flowers from epidermal cells are shown. The epidermal cells of the HF petals do not contain trichomes. The number on the panel indicates the number of stomata in the ×400 magnified microscopic field. Scale bar = 20 µm.

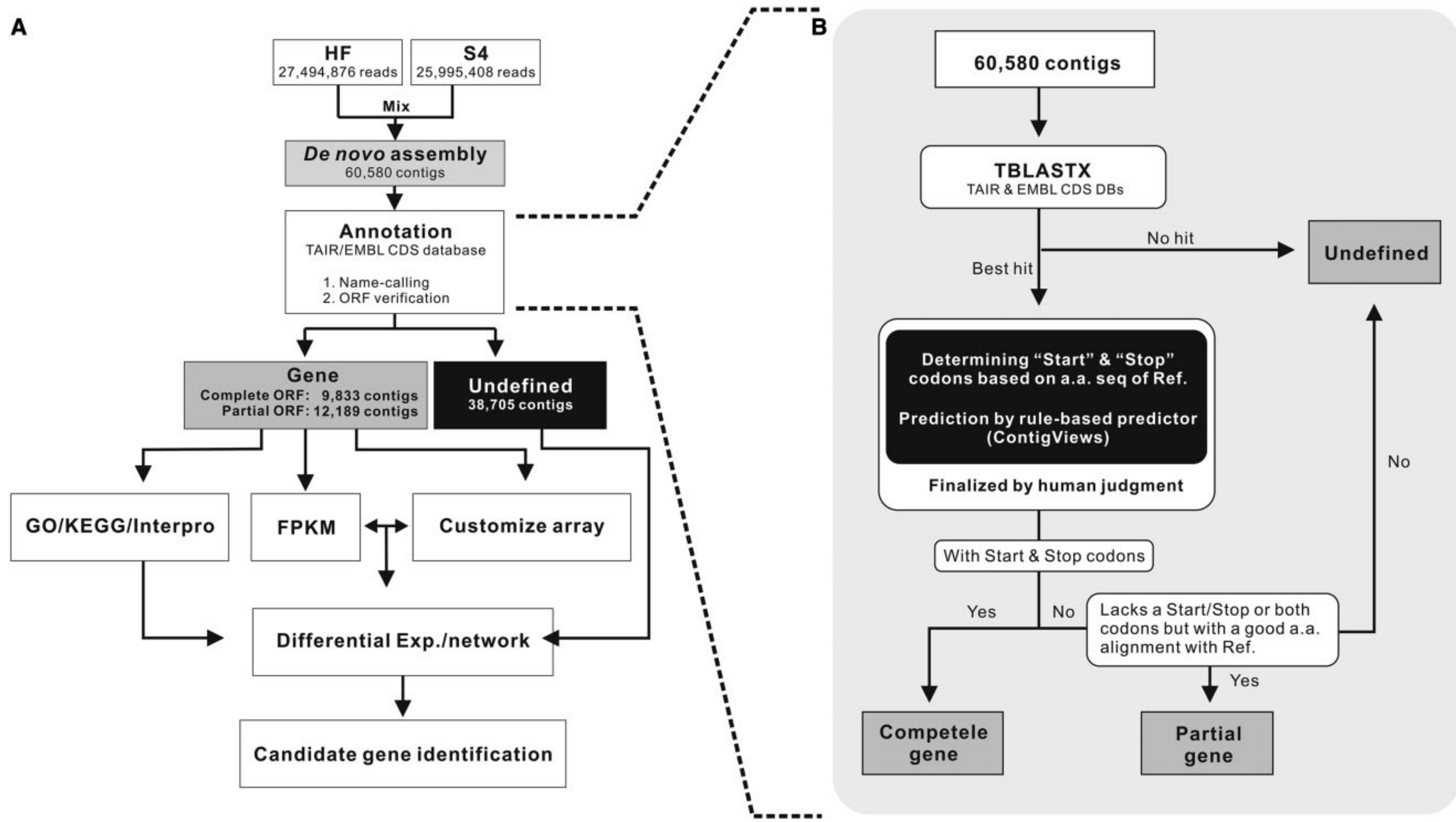

Fig. 2 The NGS analysis workflow for the whole transcriptome. (A) Whole-transcriptome analysis pipeline. Healthy flower (HF) and stage 4 (S4) peanut witches'-broom (PnWB)-infected leafy flower samples from *Catharanthus roseus* were selected for whole-transcriptome analysis using deep sequencing. Fragments per kilobase of transcript per million mapped reads (FPKM). Gene ontology (GO). Kyoto Encyclopedia of Genes and Genomes (KEGG). Interpro domain database. (B) The 60,580 de novo assembled *C. roseus* contigs sequences were compared with the TAIR and EMBL coding sequence (CDS) databases through BLAST for open reading frame (ORF) annotation.

Table 1 Summary of NGS results of the *C. roseus* transcriptome

Total reads	53,490,284 reads
Total reads length	4,703,301,221 nt
Paired reads	43,507,258 reads
Paired rate	81.3%
De novo assembly	60,580 contigs
De novo assembly successful rate (%)	85.9%
N_{50} (nt)	1,196 nt

the *C. roseus* transcriptome resulted in 9,833 complete contigs (16.15%) and 12,189 partial contigs (19.96%) (**Figs. 2A**, **3B**). All the annotated genes of *C. roseus* are available in the *C. roseus* section of the ContigViews database.

According to the ORF annotation results, 7% of the contigs contained a sequence portion that was similar to the protein sequence of the reference; however, internal stop codons or non-identical amino acids existed upstream and downstream of the similar regions, suggesting that these contigs belong to non-coding RNAs (ncRNAs) or misassembled sequences (**Fig. 3A**, panel iii). Moreover, 56.89% of the contigs did not

Table 2 Significant gene ontology (GO) terms from the gene set enrichment analysis for the 3,573 peanut witches'-broom (PnWB)-responsive genes (FDR $\leq 10^{-26}$)

Accession[a]	Type[b]	Term	FDR[c]
GO:0015979	P	Photosynthesis	9.70E-42
GO:0019684	P	Photosynthesis, light reaction	6.90E-38
GO:0008152	P	Metabolic process	1.70E-33
GO:0050896	P	Response to stimulus	1.50E-32
GO:0019748	P	Secondary metabolic process	4.70E-32
GO:0005975	P	Carbohydrate metabolic process	4.40E-31
GO:0006091	P	Generation of precursor metabolites and energy	3.90E-29
GO:0042221	P	Response to chemical stimulus	3.90E-29
GO:0051186	P	Cofactor metabolic process	3.20E-27
GO:0044262	P	Cellular carbohydrate metabolic process	6.20E-27
GO:0009657	P	Plastid organization	8.20E-27
GO:0006066	P	Alcohol metabolic process	2.20E-26
GO:0008299	P	Isoprenoid biosynthetic process	2.20E-26
GO:0034641	P	Cellular nitrogen compound metabolic process	2.30E-26
GO:0006720	P	Isoprenoid metabolic process	2.90E-26
GO:0044434	C	Chloroplast part	1.40E-52
GO:0044435	C	Plastid part	6.90E-52
GO:0009507	C	Chloroplast	2.10E-47
GO:0009536	C	Plastid	2.10E-47
GO:0009579	C	Thylakoid	2.60E-46
GO:0031976	C	Plastid thylakoid	1.10E-43
GO:0009534	C	Chloroplast thylakoid	1.10E-43
GO:0031984	C	Organelle subcompartment	1.60E-43
GO:0044436	C	Thylakoid part	4.90E-41
GO:0034357	C	Photosynthetic membrane	2.20E-38
GO:0009535	C	Chloroplast thylakoid membrane	1.80E-37
GO:0055035	C	Plastid thylakoid membrane	2.20E-37
GO:0042651	C	Thylakoid membrane	4.80E-36
GO:0016020	C	Membrane	1.30E-34
GO:0009570	C	Chloroplast stroma	7.70E-30
GO:0009532	C	Plastid stroma	1.20E-29
GO:0009526	C	Plastid envelope	1.10E-26
GO:0009941	C	Chloroplast envelope	3.60E-26

[a] Accession: gene ontology (GO) ID.
[b] Type: biological processes (P); cellular components (C); molecular functions (M).
[c] False discovery rate.

have any orthologs in the TAIR or EMBL databases and were classified as non-hit contigs (**Fig. 3A**, panel iv). These novel genes might be undiscovered or unreported and can be subjected to motif and domain predictions in the future. We classified ncRNAs, misassembled sequences and non-hit contigs as undefined contigs (**Fig. 3A**, panels iii and iv). The ORF annotation results marked the contig sequences in the ContigViews database. The 63.89% undefined contigs include ncRNAs and novel genes that might be critical for leafy flower development. The distribution of the fold changes for differential expression of undefined contigs between HF and S4 flowers is shown in **Fig. 3C**. Approximately 22,497 undefined contigs showed a 2-fold differential expression, suggesting that these contigs might play important biological roles in the PnWB-mediated leafy flower transition.

Gene sequence comparison between *C. roseus* and Arabidopsis

To understand the sequence similarities between *C. roseus* and Arabidopsis, the 9,833 complete contigs from *C. roseus* were compared with their best-matched CDS from Arabidopsis using BLAST. The distributions of the CDS lengths are shown in **Fig. 4A**. The average nucleotide identity was approximately 70% (**Fig. 4B**). The lengths of the matched CDS between *C. roseus* and Arabidopsis are highly correlated ($R = 0.964$) (**Fig. 4C**). These results indicated that gene homology exists between *C. roseus* and Arabidopsis.

The scatter plot of the coverage rate vs. the $\log_{10}$ read counts of complete genes indicated that the sequenced reads were well distributed across the transcriptome, such that the coverage rate was uniformly distributed at approximately 1.0× regardless of the number of read counts obtained for complete genes (**Fig. 4D**). For the partial genes, the coverage was lower (generally <0.5×), and fewer read counts were obtained than for the complete genes (**Fig. 4E**). This result implies that the contigs corresponding to partial genes could only be partially assembled due to the lack of sufficient sequence information.

NGS vs. customized microarray

In this study, 20,711 annotated genes, including complete and partial contigs, were selected for probe design for a *C. roseus*-customized microarray. The gene expression profiles of the HF, S3, S4, S5 and diseased leaf (DL) samples were analyzed using microarrays. First, we identified 3,573 genes that showed >50% coverage in the reference sequence and at least a 2-fold change in expression level between the S4 and

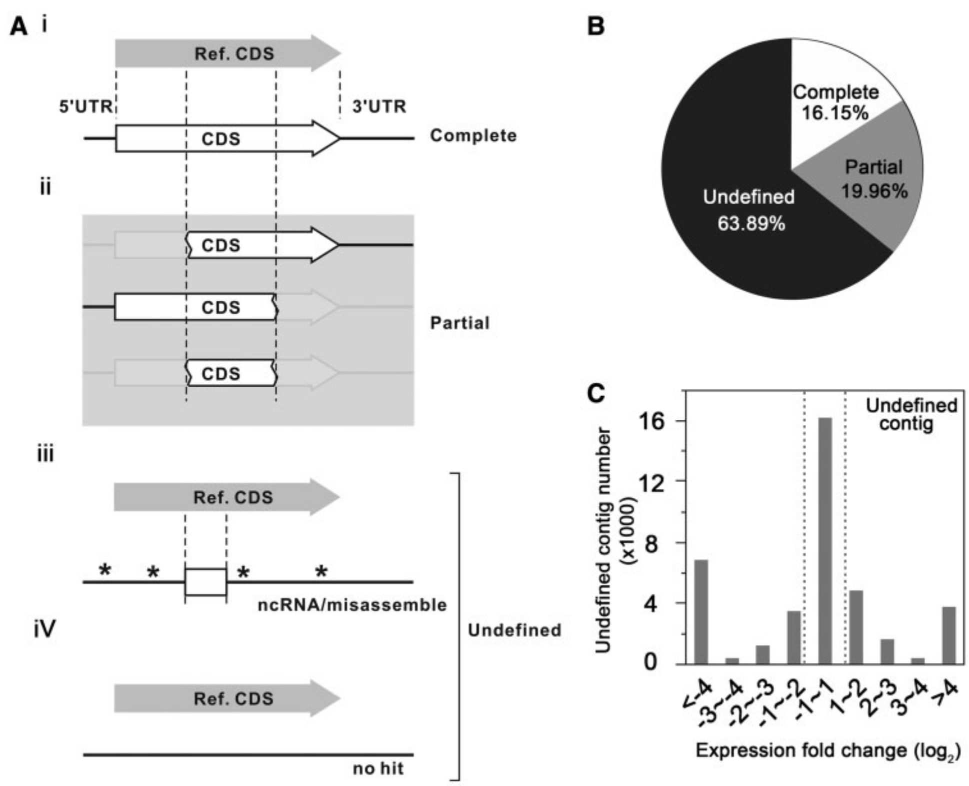

Fig. 3 The characterization of open reading frame (ORF) annotations. (A) The five ORF annotations of the assembled contigs were (i) a complete contig sequence with 5′- and 3′-untranslated regions (UTRs) and a complete ORF; (ii) a partial contig sequence lacking the 5′, 3′, or 5′ and 3′ parts of the coding sequence; and (iii) an ncRNA or misassembled contig with a partial amino acid sequence of the reference CDS but with one or more premature stop codon(s). The asterisk indicates the stop codon. (iv) The undefined contigs did not yield any BLAST results from the CDS databases. (B) Pie chart showing the percentages of complete genes, partial genes and undefined contigs. (C) Fold change distribution between HF and S4 expression levels for undefined contigs. The dotted lines indicate the threshold of a 2-fold change.

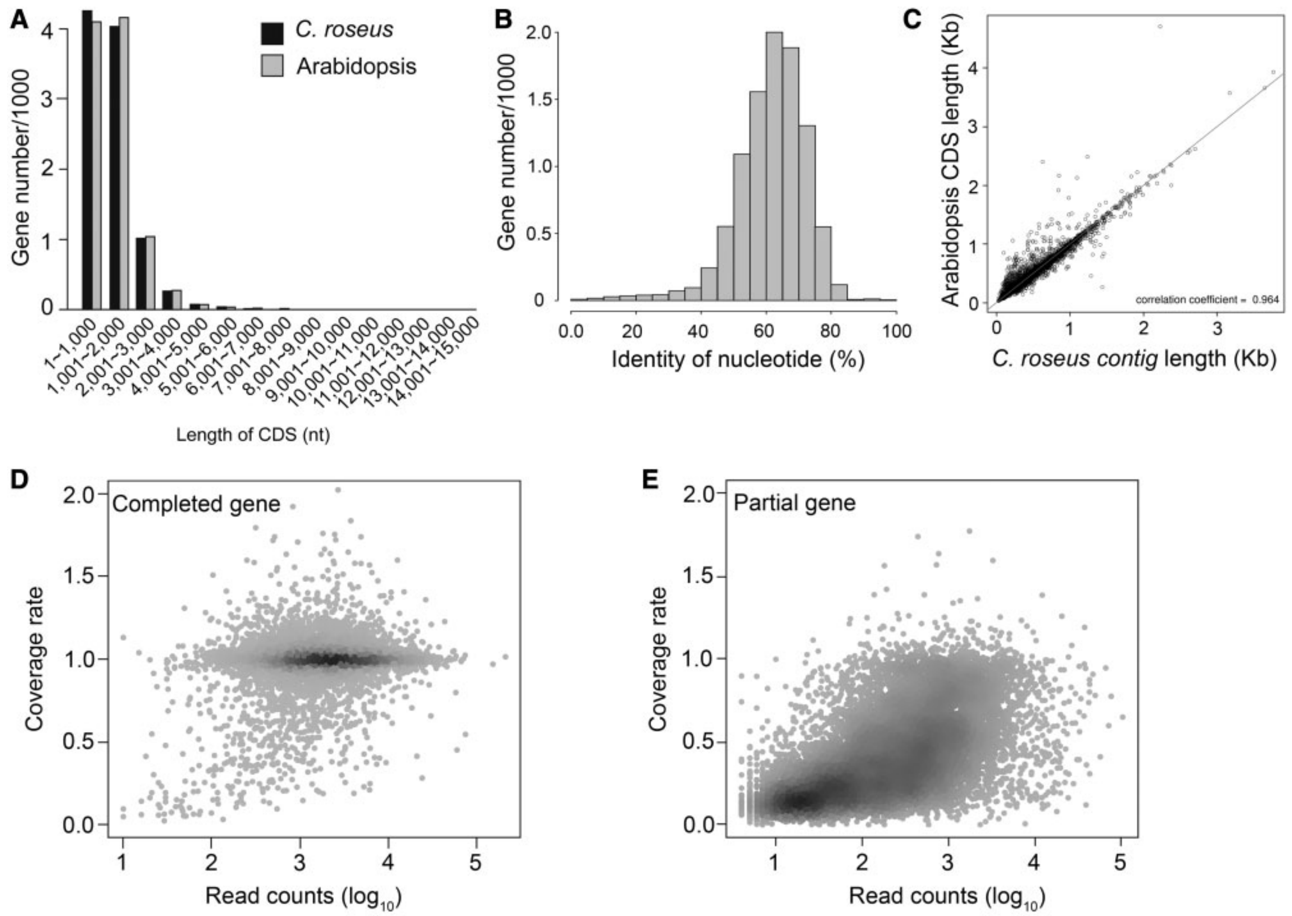

Fig. 4 Statistical analysis of gene similarity between *C. roseus* and Arabidopsis. The length distribution comparison (A) and the nucleotide identity (B) of the coding sequences (CDS) of Arabidopsis and *C. roseus* are shown. (C) Scatter plot of the CDS length of *C. roseus* vs. that of Arabidopsis. (D) Scatter plot of the average coverage to the log_{10} read counts for all completed genes. (E) Scatter plot of the average coverage to the log_{10} read counts for all partial genes.

HF samples in the NGS and microarray profiles, which were considered for expression fold change analysis (**Supplementary Table S2**). The data indicated that 56.03% of the genes (2,002 probes) in the NGS and microarray results exhibited up-regulation in S4, whereas 31.74% of the genes (1,134 probes) exhibited down-regulation in the S4 sample in both profiles (**Fig. 5A, B**). Only 12.23% (428 probes) of the genes generated conflicting results between the NGS and the microarray (**Fig. 5C, D**). These results indicated that 87.77% of the NGS and the microarray results were generally consistent.

Subsequently, we analyzed the fold changes in the gene expression of S3, S4, S5 and DL compared with the HF microarray data. The clustering results indicated that the S3 and S4 samples had similar gene expression patterns, and the gene expression patterns of the S5 and DL samples were identical (**Fig. 5E**). These results suggested that the S5 PnWB-infected leafy flowers switched to leaf identity.

Catharanthus roseus defense pathways

Because PnWB is a plant pathogen, several defense genes in *C. roseus* were identified and analyzed in this study. Two *PATHOGENESIS-RELATED GENE 1* ortholog genes, *CrPR1.1* (DDS58728) and *CrPR1.2* (DDS58885), were identified in *C. roseus* and exhibited up-regulation in the S4 PnWB-infected leafy flowers in the NGS profile (**Fig. 6A, B**). In the microarray profile, the expression levels of *CrPR1.1* and *CrPR1.2* also increased during PnWB infection. However, the expression levels of *CrPR1.2* were lower in the S5 and DL samples (**Fig. 6A, B**). The *CrSID2* (DDS51036) gene, whose ortholog in Arabidopsis plays a role in salicylic acid biosynthesis, was down-regulated during PnWB infection, whereas the flagellin receptor ortholog, *CrFLS2* (DDS46996), was up-regulated in all stages of the leafy flowers, but not the DL sample (**Fig. 6C, D**). These results indicated that *C. roseus* sensed the pathogen infection but that the pathogen counteracted the host's innate immunity pathway. The expression profiles of other defense-related genes in *C. roseus* also exhibited counteracting patterns (**Supplementary Fig. S2**).

Flowering-related genes in *C. roseus*

We hypothesized that PnWB alters the floral transition from the reproductive stage to the vegetative stage. Therefore, the

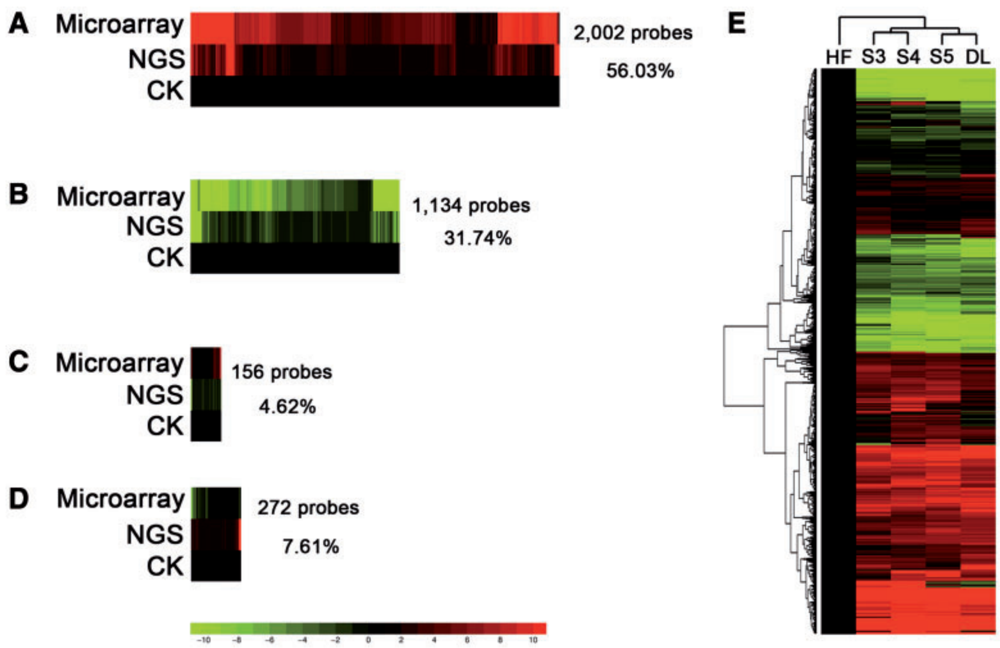

Fig. 5 Comparison of the gene expression patterns between next-generation sequencing (NGS) and the customized microarray. (A) Genes whose expression was up-regulated according to NGS and the microarray. (B) Genes whose expression was down-regulated according to NGS and the microarray. (C) Genes whose expression was down-regulated according to NGS but up-regulated in the microarray. (D) Genes whose expression was up-regulated according to NGS but down-regulated in the microarray. The relative expression pattern of stage 4 (S4) PnWB-infected leafy flowers was compared with that of healthy flowers (HFs). (E) The relative average expression patterns of flowers at the five stages of infection compared with HFs as determined by the customized microarrays.

expression levels of *C. roseus* floral-related genes that are orthologs of Arabidopsis genes were analyzed. Thirty-three of the 60 flowering-related genes of Arabidopsis showed high homology to *C. reoseus* orthologs (**Supplementary Table S3**). The *C. roseus* orthologs with significant differential expression between the HF and S4 profiles were selected for further analysis. Two *SVP* orthologs of *C. roseus*, *CrSVP1* (DDS51524) and *CrSVP2* (DDS57797), were up-regulated during PnWB infection (**Fig. 6E, F**). Both the NGS and microarray expression profiles for *CrFT* (DDS15319) showed down-regulation (**Fig. 6G**), suggesting that the flowering pathway was repressed during PnWB infection. In addition, two *SOC1* orthologs of *C. roseus*, *CrSOC1.1* (DDS15707) and *CrSOC1.2* (DDS52499), showed up-regulation in the S4 samples by NGS; however, only *CrSOC1.1* expression substantially increased in the various stages of leafy flower development and DL (**Fig. 6H, I**). *Catharanthus roseus* contains two *AP1* orthologs, *CrAP1.1* (DDS57600) and *CrAP1.2* (DDS57473), one *AP2* ortholog, *CrAP2* (DDS52139), and two *AP3* orthologs, *CrAP3.1* (DDS51937) and *CrAP3.2* (DDS30261). With the exception of *CrAP1.2*, the *CrAP1/2/3* genes were repressed during PnWB infection (**Fig. 6J–N**). Moreover, *CrPI* (DDS53164) was down-regulated during PnWB infection (**Fig. 6O**).

CrAG (DDS19760) did not exhibit substantially different expression levels between HF and various stages of leafy flowers, but it was expressed at a low level in the DL sample (**Fig. 6P**). However, an *AGAMOUS-like 19* (AGL19) ortholog, *CrAGL19* (DDS57785), which plays a flower activator role in Arabidopsis (Schönrock et al. 2006, Kim et al. 2013), was highly expressed during PnWB infection (**Fig. 6Q**). *CrAGL11* (DDS52103), which is specifically expressed in ovules in Arabidopsis (Rounsley et al. 1995), exhibited down-regulation under PnWB infection (**Fig. 6R**). In addition, *CrGA1.1* (DDS34392), an ortholog in the gibberellin pathway, and *CrCDF1* (DDS13408), an ortholog in the photoperiodic pathway, were down-regulated in PnWB-infected *C. roseus* (**Fig. 6S, T**). These results suggested that the expression of the flowering genes increased to counteract the leaf development triggered by the pathogen, resulting in abnormal leafy phenotypes (**Fig. 1A**). The expression patterns of other orthologs related to flowering pathways are shown in **Supplementary Figure S3**.

Leaf development genes in *C. roseus*

In contrast, two *C. roseus KNOX* genes, *CrKNAT1* (DDS16838) and *CrKNAT6* (DDS57620), whose Arabidopsis orthologs promote leaf development, were up-regulated in the S4 samples (**Fig. 6U, V**). In addition, *CrZWI* (DDS12250), whose Arabidopsis ortholog is involved in trichome development, was up-regulated in the S4 samples (**Fig. 6X**). However, *CrROT3* (DDS50896), whose Arabidopsis ortholog is involved in determining the size of the leaf cells, was down-regulated in the S4 samples (**Fig. 6W**). In summary, most of the leaf development genes were turned on in PnWB-trigged leafy flower tissues. The expression patterns of other orthologs

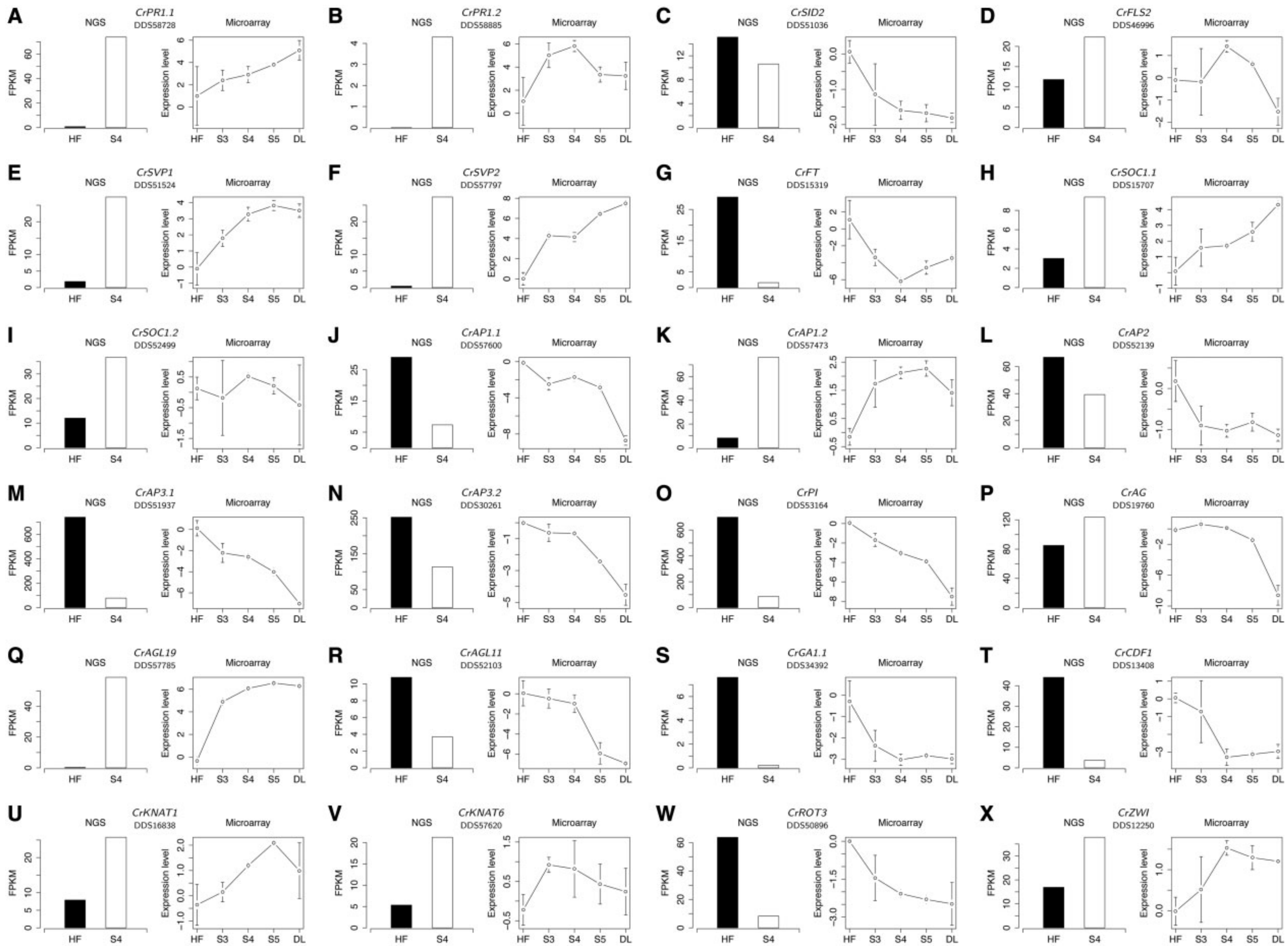

Fig. 6 Comparison of defense-, flowering- and leaf development-related gene expression levels in *C. roseus*, as determined using next-generation sequencing (NGS) and microarrays. Defense-related genes are shown in A–D. Flowering-related genes are shown in E–T. Leaf development genes are shown in U–X. The fragments per kilobase of transcript per million mapped reads (FPKM) value was used to indicate the gene expression levels that were determined by NGS in healthy flowers (HFs) and stage 4 (S4) PnWB-infected leafy flowers (left panel). The gene expression levels in HF, S3, S4, S5 and diseased leaf (DL) samples were detected using a microarray (right panel). The bars represent SDs ($n = 3$).

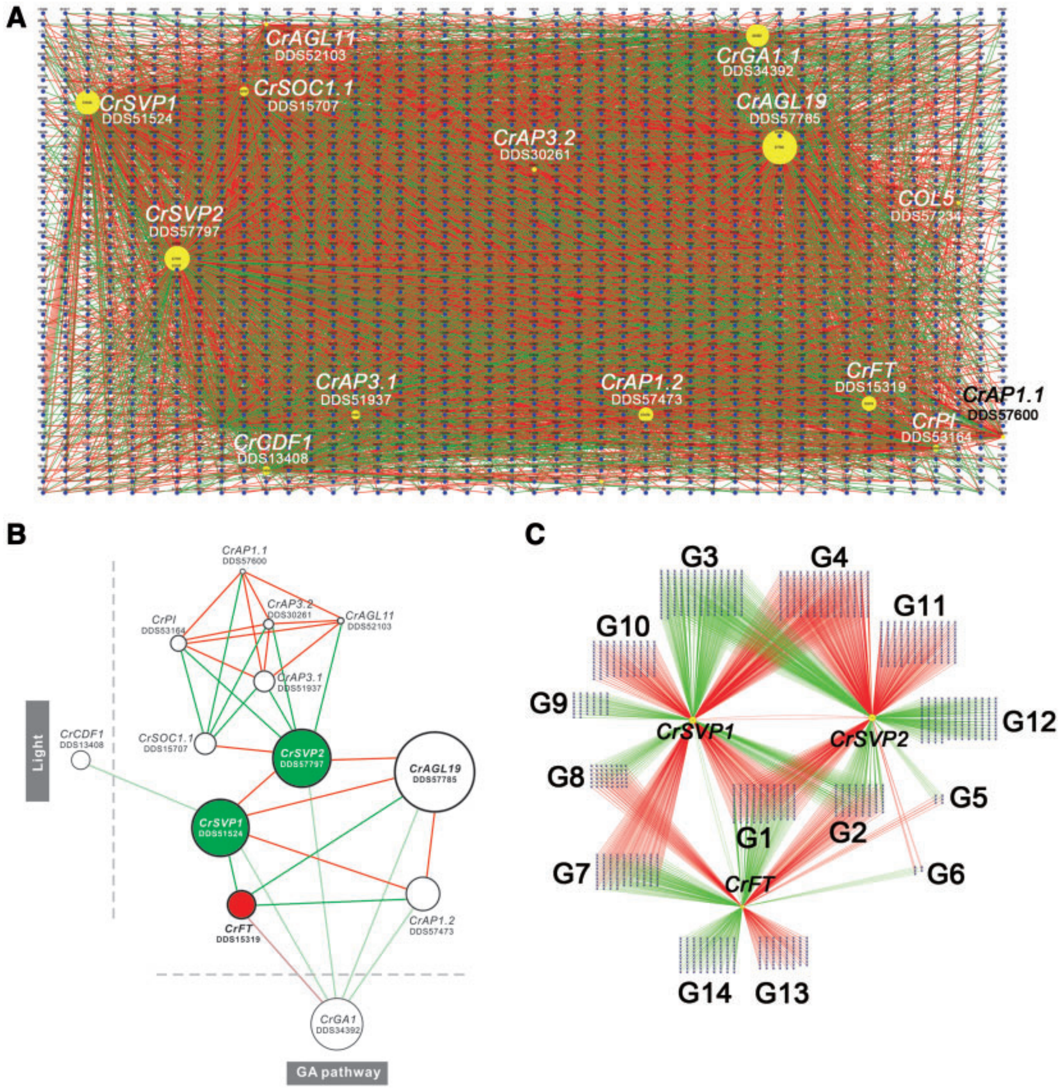

Fig. 7 Gene-to-gene networks in flowering pathways. (A) Flowering pathway networks. The key flowering genes are indicated on the map by yellow circles. The size of the circle represents the number of genes connected with that gene. (B) Close-up view of the flowering pathway. The size of the circle represents the number of genes connected with that gene. (C) Relationship gene grouping based on the *CrSVP1*, *CrSVP2* and *CrFT* genes. Fourteen groups are indicated on the relationship map. A positive relationship between two genes is indicated by a red line. A negative relationship between two genes is indicated by a green line.

related to leaf development pathways are shown in **Supplementary Fig. S4**.

Leafy flower transition networks

To identify the major genes that are involved in or correlated with PnWB-mediated leafy flower development, 3,573 genes that had $\geq$2-fold changes between the S4 and HF expression ratios with a coverage rate of >50% (contig sequence of *C. roseus*/CDS sequence of Arabidopsis) were selected for further network analysis (**Supplementary Tables S2, S4**). Fourteen flower-related genes (**Supplementary Table S5**) selected from the 3,573 gene list were used as queries to generate a 14 $\times$ 3,573 relationship map (**Fig. 7A**). In the correlation network, the yellow circles represent *C. roseus* flowering genes; the size of the circle represents the outdegree (the numbers of connected nodes) of the gene. The red and green lines represent the positive and negative relationships, respectively (**Fig. 7**). The data indicated that *CrAGL19*, *CrSVP1*, *CrSVP2*, *CrGA1.1*, *CrFT* and *CrAP1.2*, which have >500 connected nodes, might play essential roles in the PnWB-mediated leafy flower transition (**Fig. 7A**).

A zoomed-in map of the connected nodes containing 14 major flower-related genes is shown in **Fig. 7B**. The map indicates that *CrSVP1*, *CrSVP2*, *CrSOC1.1* and *CrAP1.2* have positive relationships with each other but negative relationships with *CrFT*, *CrAGL11*, *CrAP3* and *CrPI* (**Fig. 7B**). In contrast, *CrAP1.1*, *CrAP3.1*, *CrAP3.2*, *CrAGL11* and *CrPI* have positive relationships with each other (**Fig. 7B**). In Arabidopsis, *CDF1* (AT3G47500) and *AtGA1* (A4G02780) are involved in the photoperiod, gibberellin and ethylene pathways, respectively. In *C. roseus*, *CrCDF1* and *CrGA1.1* exhibited negative relationships with *CrSVP1*, *CrSVP2*, *CrAGL19* and *CrAP1.2*, whereas only the gibberellin pathway exhibited a positive relationship with *CrFT* (**Fig. 7B**). Interestingly, *CrAGL19* and *CrGA1* had 1,246 and 828 connected nodes in the $14 \times 3,573$ relationship map, respectively, implying that these genes play critical roles in the leafy flower transition, and they are candidate target genes.

Identifying candidate genes involved in the PnWB-mediated leafy flower transition

The $14 \times 3,573$ relationship map suggested the existence of multiple undiscovered gene-to-gene networks. The *CrSVP1*, *CrSVP2* and *CrFT* genes were first analyzed using 3,573 genes to generate the $3 \times 3,573$ relationship map to identify potential genes that might be involved in the PnWB-mediated leafy flower transition (**Fig. 7C**). Based on the relationships, 14 groups (G1–G14) with various relationships to the *CrSVP1*, *CrSVP2* and *CrFT* genes were identified (**Fig. 7C**; **Supplementary Table S6**). Each blue spot represents a candidate gene in each group. G1 had a positive relationship with *CrSVP1/2* but a negative relationship with *CrFT*, whereas G2 showed a negative relationship with *CrSVP1/2* but a positive relationship with *CrFT* (**Fig. 7C**). Moreover, G3 and G4 showed negative and positive relationships, respectively, with *CrSVP1* and *CrSVP2* (**Fig. 7C**). These results strongly suggested that *CrSVP1* and *CrSVP2* have similar functions but are antagonistic to *CrFT*. In contrast, G7 (104 genes) and G8 (48 genes) exhibited opposite relationships with *CrSVP1* and *CrFT*, whereas <10 gene relationships (G5 and G6) existed between *CrSVP2* and *CrFT* (**Fig. 7C**; **Supplementary Table S6**). Moreover, G9–G14 had only a single positive or negative relationship with *CrSVP1*, *CrSVP2* or *CrFT* genes. These data implied that *CrSVP1* plays a major role in antagonizing *CrFT* function. In this study, 1,391 genes from G1 to G14 were summarized and are listed in **Supplementary Table S6**.

Next, 1,391 genes were selected to generate a $1,391 \times 3,573$ relationship map. The relationships with >50 nodes of genes and a correlation coefficient >0.95 were identified as potentially the most critical genes involved in the PnWB-mediated leafy flower transition. The list was trimmed down to 132 genes in the 14 groups (**Supplementary Table S7**). These genes are uncharacterized; however, most of them were predicted to exhibit expression changes in flowering or vegetative stage development microarray experiments by TAIR (data not shown).

Discussion

In this study, we developed a pipeline for whole-transcriptome analysis in a non-model organism based on NGS and customized microarray profiles. The *C. roseus* and PnWB interaction increases the interest in discovering the mechanism of the PnWB-mediated leafy flower transition. Through the mixed reads de novo assembly approach, the longest 60,580 *C. roseus* contigs were identified from 253 million transcriptome reads. Analyzing the efficiency of these contigs and identifying the crucial genes involved in the PnWB-mediated leafy flower transition are major areas of interest.

Quantitative transcription level

Biases were generated during deep sequencing due to the fragmentation of RNA/cDNA or amplification artifacts that occurred during library preparation (Wang et al. 2009b). These biases affect the accuracy of the quantification of RNA expression analyzed via NGS. Therefore, validation of RNA expression is a necessary procedure. The customized microarray provides an advantage in allowing the efficient verification of NGS expression profiles. In this study, an 87.77% identity was demonstrated between the NGS and microarray results, suggesting that the whole-transcriptome sequence is appropriate for performing quantitative comparisons of RNA expression.

The importance of ORF identification

We designed algorithms in the ContigViews web server for gene and ORF region identification. The annotation concept was relayed on the TAIR and EMBL CDS databases to determine the ORF regions of the contig sequences. Following this procedure, highly reliable orthologs of *C. roseus* and their ORF regions were predicted, and the algorithms excluded 70% of the non-gene sequences from the 60,580 contigs. The *C. roseus* database contains 22,022 annotated genes, which enables researchers to focus immediately on gene expression patterns and functions. The information on distinct genes suggests the completeness of the genetic map in the current database.

The scatter plot results indicated that a low expression level of a gene results in a partially assembled ORF contig. These partial ORF contigs can be completed with another batch of transcriptome data, which may be derived from different treatments or tissues. In addition, some partial ORF contigs might belong to a gene that is missing certain connecting reads, resulting in two or more injunction fragments. This phenomenon was observed for *CrEMF2*, which is an ortholog of *AtEMF2* (AT5G51230), a component of the Polycomb group (PcG). *CrEMF2* was divided into *CrEMF2.1* (DDS13840) and *CrEMF2.2* (DDS20205) because some NGS reads were missing from the central region of the gene (**Supplementary Table S8**). *CrEMF2.1* should be located at the 5′ end of the *CrEMF2* gene, whereas *CrEMF2.2* is located at the 3′ end (data not shown). However, this problem can be solved through mixed assembly of long-read sequences, which were generated by 454 sequencing (Su et al. 2011a). Because there is 70% nucleotide identity between *C. roseus* and Arabidopsis, the Arabidopsis CDS sequence can be used as a reference to identify lost junctions of fragments and link them by a scaffolding approach.

Furthermore, 63.89% of the undefined contigs exist in the database; we assumed that numerous novel genes are in this group. These novel genes might be essential for PnWB-mediated leafy flower transitions. The domain and expression pattern identification of the undefined contigs will facilitate the identification of these genes. In summary, these different approaches will make the databases more complete at the ORF level.

PnWB counteracts the defense and flowering pathways, but triggers the leaf development pathway in *C. roseus*

The obtained transcriptome profiles revealed up-regulation of the defense-related *CrPR1.1* and *CrPR1.2* genes in PnWB-infected leafy flowers. In addition, the Arabidopsis ortholog *CrFLS2* was up-regulated during PnWB infection. These results indicated that the innate immunity pathway of *C. roseus* was triggered. However, PnWB repressed the expression of *CrSID2*, which is an Arabidopsis ortholog involved in salicylic acid biosynthesis. Similar counteraction effects were also detected during the leafy flower transition. Negative regulator genes, such as *CrSVP1* and *CrSVP2*, were generally up-regulated under PnWB infection, whereas positive regulator genes were generally down-regulated. However, *CrSOC1.1*, *CrSOC1.2*, *CrAGL19* and *CrAP1.2*, which are positive regulators of flowering in Arabidopsis, were up-regulated. In addition, several important leaf development genes, such as *CrKNAT1*, *CrKNAT2*, *CrAS2*, *CrREV*, *CrROT3*, *CrZWI* and *CrKAN2*, exhibited differential expression levels between healthy and PnWB-infected *C. roseus*. In summary, the counteraction effects observed between reproductive and vegetative stages triggered in *C. roseus* by PnWB result in the leafy flower transition.

Identifying genes that might play a critical role in the leafy flower transition

We propose a selection procedure based on the relationship network for identifying candidate genes. In this study, 132 genes in the 14 groups with a correlation coefficient >0.95 and relationships with >50 nodes were selected for further analysis (**Supplementary Table S7**). The network results suggested multiple candidates that might be involved in PnWB-mediated leafy flower formation. Most of the gene groups that have positive relationships with *CrSVP1* and *CrSVP2* were suggested to play roles in leaf or chloroplast development. In contrast, the gene groups that have a positive relationship with *CrFT* were suggested to play a role in flower development, including genes involved in the development of pollen and petals as well as gametogenesis.

CrGA1 presents 828 nodes corresponding to relationship with other genes involved in the PnWB-mediated leafy flower transition. In Arabidopsis, *GA1* regulates an early step in the gibberellin biosynthesis pathway and is involved in floral development (Silverstone et al. 1997, Brock et al. 2012). The *GA1* mutant (*ga1-3*) shows a dwarfed, late flowering phenotype (Brock et al. 2012). Therefore, the results of network analysis implied that *GA1* is a candidate gene to be further studied in relation to the PnWB-mediated leafy flower transition.

Furthermore, *CrAGL19* has significant relationships (1,246 nodes) with other genes involved in the PnWB-mediated leafy flower transition pathway. AGL19 is another flowering activator in the Arabidopsis vernalization pathway that acts independently of the SVP/FLC-regulated flowering pathway (Alexandre and Hennig 2008). *AGL19* expression has been shown to be repressed by PcG group repressive complex 2 (PRC2), which includes the MSI1, CLF/SWN, EMF2 and FIE proteins. This repression occurs through enriched trimethylation of Lys27 on histone H3 (H2K27me3) in *AGL19* chromatin (Kim et al. 2013). Cold treatment relieves *AGL19* expression by PRC2 repression and activates *LFY* and *AP1*, resulting in flowering in Arabidopsis (Kim et al. 2013). Interestingly, a long ncRNA, *COLDAIR*, is induced by vernalization and associates with PRC2 to form a complex to repress *FLC* expression. This repression occurs through enhanced H3K27me3 in the *FLC* chromatin, resulting in flowering in Arabidopsis (Heo and Sung 2011). *AGL19* was highly expressed in S4 leafy flowers, suggesting that the degree of H3K27me3 in the chromatin might be reduced. However, the *CrMSI1* (DDS115692), *CrMSI2* (DDS15361), *CrCLF* (DDS13072), *CrSWN* (DDS11461), *CrEMF2.1* (DDS13840), *CrEMF2.2* (DDS20205) and *CrFIE* (DDS30886) orthologs did not show differential expression between the HF and S4 samples (**Supplementary Table S8**). We hypothesize that unidentified ncRNA of *C. roseus* might decrease during PnWB infection.

Working model for the PnWB-mediated leafy flower transition

We propose a model for the PnWB-mediated leafy flower transition, as demonstrated in **Fig. 8**. In the *C. roseus* floral meristem, gene expression is predicted to promote a switch from the vegetative stage to the reproductive stage. However, in PnWB-infected *C. roseus*, the pathogen represses certain flowering genes, such as *CrFT*, *CrAP* genes, *CrPI*, *CrAGL11*, *CrAG* genes and *CrCDF1*, but turns on leaf development genes, such as *CrKNAT* genes, *CrZWI*, *CrAS2*, *CrREV*, *CrATHB-15* genes, *CrKAN2* and *CrFZR2*. Moreover, the flowering repressors *CrSVP1* and *CrSVP2* are highly expressed and are triggered by PnWB to suppress the reproductive stage. These alterations switch the development of the host to the vegetative stage. However, the original developmental direction in the floral meristem of *C. roseus* is from vegetative to flowering, and certain unrepressed flowering genes, such as *CrSOC1.1*, *CrAP1.2*, *CrAG* and *CrAGL19*, were still highly expressed in the PnWB-infected floral meristem. Therefore, two groups of genes (reproductive and vegetative) were abnormally co-expressed in the floral meristem, resulting in leafy flower formation.

Conclusion

In this study, an NGS and microarray pipeline was developed to study the transcriptome expression profiles of a non-model

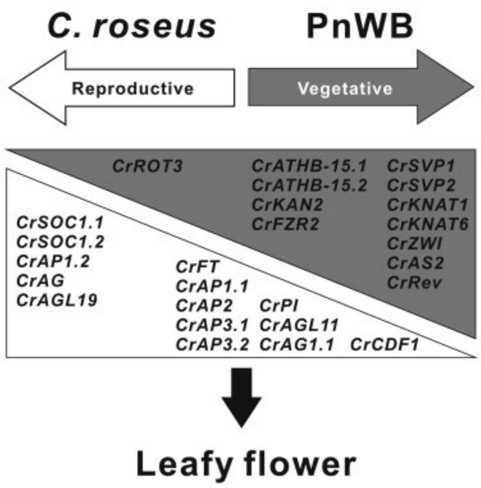

Fig. 8 Model for leafy flower formation during peanut witches'-broom (PnWB) phytoplasma infection. The arrows represent the direction of host plant development. The triangles represent gene expression levels. The gray color represents vegetative development. The white color represents reproductive development. The black arrow represents the leafy flower transition.

organism. Differentially expressed genes involved in developmental reprogramming and the defense response were identified using the NGS approach. This study provides support for a new strategy for understanding the disease mechanisms between hosts and pathogens.

Materials and Methods

Plant material, phytoplasma inoculation and disease development

Catharanthus roseus cv. Pacifica Punch Halo was used in this study. The seeds were directly sown on a Floralbella potting peat/vermiculite/perlite mix (2:1:1). The 3-week-old seedlings were transferred to 15 cm pots containing the peat/vermiculite/perlite mix and maintained in a greenhouse environment (16 h light/8 h darkness, 20–25°C). Two-month-old *C. roseus* plants were inoculated with phytoplasma through grafting. A PnWB phytoplasma strain obtained from Dr. I-lang Yang was used in this study (Taiwan Agriculture Research Institute). Each experimental grafting was performed 5–7 times for each assay. The virescence symptoms of the infected *C. roseus* developed 3–4 weeks post-grafting, whereas the phyllody symptoms developed 5–6 weeks post-grafting. Within 2–3 months after grafting, the virescent flowers became phyllodic. Five categories of floral malformations in PnWB-infected *C. roseus* were described.

Observation of the petal surfaces by microscopy

Trichome development on the petals of the HF, HL and five stages of leafy flowers (S1–S5) in *C. roseus* were observed. The abaxial epidermal cells were torn and rinsed using a cleaning solution (80% choral hydrate and 30% glycerol), and the trichomes were observed using a light microscope (Nikon, ELIPSE 80i) with ×400 magnification.

The epidermal cells from an HL, a petal of an HF, and abnormal flowers (S1–S5) were observed by SEM (Jeol, JSM-6510LV; Joint Center for Instruments and Researches, NTU). The petal samples were dissected into approximately 5 mm^2 pieces using a razor blade, and submitted to a series of processing procedures, including fixation, dehydration, critical point drying and sputter coating, before being analyzed. The samples were photographed at ×40, ×200, ×1,600, and ×3,500 magnifications.

Total RNA extraction and whole-transcriptome deep sequencing

Total RNA was extracted from 10 healthy flowers or S4 leafy flowers using TRIzol reagent (Invitrogen) according to the manufacturer's protocol. One sample for each condition was used for the NGS data. The whole-transcriptome profiles of the HF and S4 flowers were analyzed using an Illumina Hiseq 2000 (Genomics BioSci & Tech Co.). The paired-end reads were then assembled using the standard protocol of de novo assembly in CLC genomics Workbench 4.71.1 (CLC bio). The transcriptome expression profiles and the database were analyzed and constructed using the ContigViews web server (http://contigviews.csbb.ntu.edu.tw). The name-calling and ORF annotations were determined using the rule-based predictor of ContigViews. Blast2GO software (Blast2Go Co.) was used to analyze the GO and KEGG pathways and the domains of the annotated genes.

Statistical analysis of the NGS expression profiles

All statistical analyes were performed using R version 2.15.2. The transcriptome expression profiles were represented by the read counts that were mapped to the assembled contigs. The coverage rate for an assembled contig is defined as (the total number of mapped reads to the contig) × (the read length)/the contig length.

Customized microarray design and analysis

The sequence information of 20,710 *C. roseus* genes was used to design probes for the customized microarray (Agilent Co.). Total RNA from the HF, S3, S4 and S5 flowers, as well as the DLs, was used, and three biological replicates for each sample were included in the microarray analyses that were performed by the DNA Microarray Core Laboratory (Academia Sinica). The log_2 fold changes of the gene expression levels of S3, S4, S5 and DL compared with those of HF in the heatmap were calculated, and the genes and samples in the heatmap were reordered based on similarity. The consistency of the expression levels between the microarray and NGS data was analyzed and illustrated by comparing the log_2 fold change of the expression levels of S4 with those of HF.

Gene set enrichment analysis

GSEA for the 3,573 PnWB-reponsive genes was performed using the agriGO analysis toolkit (Du et al. 2010). The PnWB-responsive genes included complete or partial contigs with >50% coverage in the Arabidopsis reference sequence and at least a 2-fold change in expression level between the S4 and HF samples in the NGS and microarray profiles. The best-matched CDS from Arabidopsis using BLAST were submitted to the agriGO analysis toolkit, and the GO terms with false discovery rates ≤ 0.001 were deemed significant.

Network analysis

The gene expression profiles generated from the customized microarrays were subjected to network analysis to investigate PnWB-mediated leafy flower development. Only 3,573 genes with >50% coverage (equal length of identical sequences between the *C. roseus* contig and the Arabidopsis CDS/CDS sequence) and at least a 2-fold change in expression level between S4 and HF were considered. Among the 3,573 genes, the 14 flower-related genes were selected for queries in the network analysis. To form a correlation network with $14 \times 3{,}573$ nodes, the correlation coefficients for all $14 \times 3{,}573$ combinations of the 14 flower-related genes and the 3,573 genes were computed. If a correlation coefficient of ≥ 0.8 (or ≤ -0.8) was obtained, the two nodes were connected using a red (or green) line (**Fig. 7A**). The correlation network for the 14×14 nodes was generated similarly using the correlation coefficients for all the 14×14 combinations of any two of the 14 flower-related genes (**Fig. 7B**). For the $1{,}391 \times 3{,}573$ relationship map, the threshold for the correlation coefficient was set to 0.95.

Supplementary data

Supplementary data are available at PCP online.

Funding

This study was supported by the National Science Council [grant No, NSC 102-2313-B-002-068-MY3]; the National Taiwan University [101R7602B2].

Disclosures

The authors have no conflicts of interest to declare.

References

Alexandre, C.M. and Hennig, L. (2008) FLC or not FLC: the other side of vernalization. *J. Exp. Bot.* 59: 1127–1135.

Brock, M.T., Kover, P.X. and Weinig, C. (2012) Natural variation in GA1 associates with floral morphology in Arabidopsis thaliana. *New Phytol.* 195: 58–70.

Chen, W.Y. and Lin, C.P. (2011) Characterization of Catharanthus roseus genes regulated differentially by peanut witches' broom phytoplasma infection. *J. Phytopathol.* 159: 505–510.

Curković Perica, M. (2008) Auxin-treatment induces recovery of phytoplasma-infected periwinkle. *J. Appl. Microbiol.* 105: 1826–1834.

Davey, J.E. and Van Staden, J. (1981) Endogenous cytokinin levels and development of flower virescence in Catharanthus roseus infected with mycoplasmas. *Physiol. Plant Pathol.* 19: 193–200.

Du, Z., Zhou, X., Ling, Y., Zhang, Z. and Su, Z. (2010) agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* 38: W64–W70.

Ferrándiz, C., Gu, Q., Martienssen, R. and Yanofsky, M.F. (2000) Redundant regulation of meristem identity and plant architecture by FRUITFULL, APETALA1 and CAULIFLOWER. *Development* 127: 725–734.

Folkers, U., Kirik, V., Schöbinger, U., Falk, S., Krishnakumar, S., Pollock, M.A. et al. (2002) The cell morphogenesis gene ANGUSTIFOLIA encodes a CtBP/BARS-like protein and is involved in the control of the microtubule cytoskeleton. *EMBO J.* 21: 1280–1288.

Heo, J.B. and Sung, S. (2011) Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* 331: 76–79.

Hoshi, A., Oshima, K., Kakizawa, S., Ishii, Y., Ozeki, J., Hashimoto, M. et al. (2009) A unique virulence factor for proliferation and dwarfism in plants identified from a phytopathogenic bacterium. *Proc. Natl Acad. Sci. USA* 106: 6416–6421.

Kelley, D.R., Arreola, A., Gallagher, T.L. and Gasser, C.S. (2012) ETTIN (ARF3) physically interacts with KANADI proteins to form a functional complex essential for integument development and polarity determination in Arabidopsis. *Development* 139: 1105–1109.

Kim, G.T., Tsukaya, H. and Uchimiya, H. (1998) The ROTUNDIFOLIA3 gene of Arabidopsis thaliana encodes a new member of the cytochrome P-450 family that is required for the regulated polar elongation of leaf cells. *Genes Dev.* 12: 2381–2391.

Kim, W., Latrasse, D., Servet, C. and Zhou, D.X. (2013) Arabidopsis histone deacetylase HDA9 regulates flowering time through repression of AGL19. *Biochem. Biophys. Res. Commun.* 432: 394–398.

Kitamura, Y., Hosokawa, M., Uemachi, T. and Yazawa, S. (2009) Selection of ABC genes for candidate genes of morphological changes in hydrangea floral organs induced by phytoplasma infection. *Sci. Hortic.* 122: 603–609.

Lee, H., Suh, S.S., Park, E., Cho, E., Ahn, J.H., Kim, S.G. et al. (2000a) The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in Arabidopsis. *Genes Dev.* 14: 2366–2376.

Lee, I.M., Davis, R.E. and Gundersen-Rindal, D.E. (2000b) Phytoplasma: phytopathogenic mollicutes. *Annu. Rev. Microbiol.* 54: 221–255.

Li, D., Liu, C., Shen, L., Wu, Y., Chen, H., Robertson, M. et al. (2008) A repressor complex governs the integration of flowering signals in Arabidopsis. *Dev. Cell* 15: 110–120.

MacLean, A.M., Sugio, A., Makarova, O.V., Findlay, K.C., Grieve, V.M., Tóth, R. et al. (2011) Phytoplasma effector SAP54 induces indeterminate leaf-like flower development in Arabidopsis plants. *Plant Physiol.* 157: 831–841.

Mardis, E.R. (2008) The impact of next-generation sequencing technology on genetics. *Trends Genet.* 24: 133–141.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A. et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376–380.

McConnell, J.R., Emery, J., Eshed, Y., Bao, N., Bowman, J. and Barton, M.K. (2001) Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. *Nature* 411: 709–713.

Moon, J., Suh, S.S., Lee, H., Choi, K.R., Hong, C.B., Paek, N.C. et al. (2003) The SOC1 MADS-box gene integrates vernalization and gibberellin signals for flowering in Arabidopsis. *Plant J.* 35: 613–623.

Morozova, O. and Marra, M.A. (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92: 255–264.

Ori, N., Eshed, Y., Chuck, G., Bowman, J.L. and Hake, S. (2000) Mechanisms that control knox gene expression in the Arabidopsis shoot. *Development* 127: 5523–5532.

Pekker, I., Alvarez, J.P. and Eshed, Y. (2005) Auxin response factors mediate Arabidopsis organ asymmetry via modulation of KANADI activity. *Plant Cell* 17: 2899–2910.

Pracros, P., Renaudin, J., Eveillard, S., Mouras, A. and Hernould, M. (2006) Tomato flower abnormalities induced by stolbur phytoplasma infection are associated with changes of expression of floral development genes. *Mol. Plant Microbe Interact.* 19: 62–68.

Prigge, M.J., Otsuga, D., Alonso, J.M., Ecker, J.R., Drews, G.N. and Clark, S.E. (2005) Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in Arabidopsis development. *Plant Cell* 17: 61–76.

Rounsley, S.D., Ditta, G.S. and Yanofsky, M.F. (1995) Diverse roles for MADS box genes in Arabidopsis development. *Plant Cell* 7: 1259–1269.

Sakamoto, T., Kamiya, N., Ueguchi-Tanaka, M., Iwahori, S. and Matsuoka, M. (2001) KNOX homeodomain protein directly suppresses the expression of a gibberellin biosynthetic gene in the tobacco shoot apical meristem. *Genes Dev.* 15: 581–590.

Samach, A., Onouchi, H., Gold, S.E., Ditta, G.S., Schwarz-Sommer, Z., Yanofsky, M.F. et al. (2000) Distinct roles of CONSTANS target genes in reproductive development of Arabidopsis. *Science* 288: 1613–1616.

Schönrock, N., Bouveret, R., Leroy, O., Borghi, L., Köhler, C., Gruissem, W. et al. (2006) Polycomb-group proteins repress the floral activator AGL19 in the FLC-independent vernalization pathway. *Genes Dev.* 20: 1667–1678.

Silverstone, A.L., Chang, C., Krol, E. and Sun, T.P. (1997) Developmental regulation of the gibberellin biosynthetic gene GA1 in Arabidopsis thaliana. *Plant J.* 12: 9–19.

Srikanth, A. and Schmid, M. (2011) Regulation of flowering time: all roads lead to Rome. *Cell Mol. Life Sci.* 68: 2013–2037.

Su, C.L., Chao, Y.T., Alex Chang, Y.C., Chen, W.C., Chen, C.Y., Lee, A.Y. et al. (2011a) De novo assembly of expressed transcripts and global analysis of the Phalaenopsis aphrodite transcriptome. *Plant Cell Physiol.* 52: 1501–1514.

Su, Y.T., Chen, J.C. and Lin, C.P. (2011b) Phytoplasma-induced floral abnormalities in Catharanthus roseus are associated with phytoplasma accumulation and transcript repression of floral organ identity genes. *Mol. Plant Microbe Interact.* 24: 1502–1512.

Van Moerkercke, A., Fabris, M., Pollier, J., Baart, G.J., Rombauts, S., Hasnain, G. et al. (2013) CathaCyc, a metabolic pathway database built from Catharanthus roseus RNA-Seq data. *Plant Cell Physiol.* 54: 673–685.

Wang, J.W., Czech, B. and Weigel, D. (2009a) miR156-regulated SPL transcription factors define an endogenous flowering pathway in Arabidopsis thaliana. *Cell* 138: 738–749.

Wang, Z., Gerstein, M. and Snyder, M. (2009b) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10: 57–63.

Wollmann, H., Mica, E., Todesco, M., Long, J.A. and Weigel, D. (2010) On reconciling the interactions between APETALA2, miR172 and AGAMOUS with the ABC model of flower development. *Development* 137: 3633–3642.

Wu, G., Lin, W.C., Huang, T., Poethig, R.S., Springer, P.S. and Kerstetter, R.A. (2008) KANADI1 regulates adaxial–abaxial polarity in Arabidopsis by directly repressing the transcription of ASYMMETRIC LEAVES2. *Proc. Natl Acad. Sci. USA* 105: 16392–16397.

Xu, L., Xu, Y., Dong, A., Sun, Y., Pi, L., Xu, Y. et al. (2003) Novel as1 and as2 defects in leaf adaxial–abaxial polarity reveal the requirement for ASYMMETRIC LEAVES1 and 2 and ERECTA functions in specifying leaf adaxial identity. *Development* 130: 4097–4107.

Yang, I.L. (1983) Dodder transmission and microscopic observation of peanut witches' broom disease. *Natl Sci. Council Monthly (ROC)* 11: 821–826.