

Characterization and identification of *Catharanthus roseus* epigenetic-related genes that in response to peanut witches'-broom phytoplasma-mediated infection

Hsin-I Tseng^{1,4,5}, Chan-Pin Lin^{1,4,5}, Shih-Shun Lin^{1,2,3,6}

¹Institute of Biotechnology, College of Bioresources and Agriculture, National Taiwan University, 81, Chang-Xing St., Taipei, Taiwan 106

²Genome and Systems Biology Degree Program, National Taiwan University, 1, Sec. 4, Roosevelt Rd., Taipei, Taiwan 106

³Agricultural Biotechnology Research Center, Academia Sinica, 128, Sec. 2, Academia Rd., Taipei, Taiwan 115

⁴Department of Plant Pathology and Microbiology, College of Bioresources and Agriculture, National Taiwan University, 1 Sec. 4, Roosevelt Rd., Taipei, Taiwan 106

⁵These authors contributed equally to this work

⁶Corresponding author; E-mail: linss01@ntu.edu.tw; Tel; +886 2 33666023; Fax: +886 2 33666001

ABSTRACT

Tseng, H.I., Lin, C.P., and Lin S.S. 2014. Characterization and identification of *Catharanthus roseus* epigenetic-related genes that in response to peanut witches'-broom phytoplasma-mediated infection. Plant Pathol. Bull. 23:67-77.

The high-throughput transcriptome profile of *Catharanthus roseus* in the ContigViews database provides useful information on the interaction between phytoplasma and plant. In this study, several post-transcriptional gene silencing (PTGS)- and RNA-directed DNA methylation (RdDM)-related genes were identified in *C. roseus*, which might be involved in peanut witches'-broom (PnWB) phytoplasma-mediated leafy flower formation. The evolutionary relationships and expression profiles of these genes in healthy flowers (HF) and Stage 4 (S4) PnWB-infected leafy flowers were also characterized. Notably, *ARGONAUTE 5* (*CrAGO5*) and *RNA-DEPENDENT RNA POLYMERASE 1* (*CrRDR1*) in *C. roseus* were significantly up-regulated in S4 leafy flowers, whereas other PTGS- and RdDM-related genes maintained normal expression levels. These results suggested that CrAGO5 and CrRDR1 might involve in PnWB-mediated leafy flower formation and plant defense.

Keywords: Next-generation sequencing, *Catharanthus roseus*, peanut witches'-broom phytoplasma, post-transcriptional gene silencing, RNA-directed DNA methylation

Peanut witches'-broom (PnWB) phytoplasma mediates virescence and phyllody symptoms (herein referred to as leafy flower) in *Catharanthus roseus* (1). Our previous study demonstrated that the gene expression tendency of *C. roseus* was to convert the reproductive stage to the vegetative stage during PnWB infection (2). In addition, epigenetic regulations, which are controlled by post-transcriptional gene silencing (PTGS) (3, 4) and RNA-directed DNA methylation (RdDM) (5, 6), might also be involved in leafy flower formation.

In the PTGS pathway, 21- or 22-nt microRNAs (miRNAs) were generated by the DICER-LIKE 1 (DCL1) and HYPONASTIC LEAVES 1 (HYL1) complex in the nucleus. Subsequently, the double-stranded miRNA was transferred by HASTY (HST) from the nucleus to the cytoplasm and loaded into ARGONAUTE 1 (AGO1) for mRNA recognition and cleavage (4). DCL2 and DCL4 are responsible for short-interfering RNA (siRNA) biogenesis in the cytoplasm, which has been demonstrated for plant defense mechanisms (7). DCL3 produces 24-nt siRNAs, which play an important role in mediating RdDM (8).

Total RNA was extracted from 10 healthy flowers (HF) and Stage 4 (S4) PnWB-infected leafy flowers using TRIzol reagent (Invitrogen) according to the manufacturer's protocol. The deep sequencing of small RNAs of HF and S4 PnWB-infected leafy flowers was analyzed using SOLiD 4 system (Life Technologies) by the next-generation sequencing (NGS) core facility of Techcomm at National Taiwan University (9). The small RNA profiles showed that 24-nt small RNAs were significantly decreased (approximately 62% reduction) in the S4 sample, whereas the 21- and 22-nt small RNA levels were similar in the HF and S4 samples, suggesting that PnWB specifically interferes with 24-nt small RNA biogenesis (Fig. 1A).

The whole transcriptome database of *C. roseus* were published and built on ContigViews database (www.contigviews.bioagri.ntu.edu.tw) (2). We further identified epigenetic-related genes of *C. roseus*, such as DCLs and AGOs, from the ContigViews database. Due to the large protein size of DCLs (approximately 2,000 amino

acids), each DCL ortholog of *C. roseus* was divided into 2 contig sequences. DDS2790 belongs to the 5' end of *CrDCL1*, and DDS11937 belongs to the 3' end of *CrDCL1*. DDS44869 belongs to the 5' end of *CrDCL2*, whereas DDS13971 belongs to the 3' end of *CrDCL2*. DDS11443 belongs to the 5' end of *CrDCL3*, whereas DDS16373 belongs to the 3' end of *CrDCL3*. Finally, DDS112374 belongs to the 5' end of *CrDCL4*, whereas DDS3193 belongs to the 3' end of *CrDCL4*. To compare the amino acid sequences of the CrDCLs with Arabidopsis orthologs, 2 contig sequences were linked based on their alignment with Arabidopsis ortholog sequences. The missing amino acid residues in the junction area are denoted with an "X" in Figure 2A. CrDCL2 and CrDCL4 are represented as partial sequences because the C-terminus of CrDCL2 and a few amino acids of the N-terminus of CrDCL4 were missing (Fig. 2A). An alignment comparison of the DCLs of *C. roseus* and Arabidopsis showed that DEXDc, HELICc, PAZ and 2 RIBOc domains were highly conserved (Fig. 2A). In addition, CrDCL1 and AtDCL1 showed 78.5% amino acid sequence similarity between *C. roseus* and Arabidopsis, whereas the other DCLs showed 49.7 to 60.8% similarity (Fig. 1B). Moreover, the phylogenetic tree showed that 4 CrDCLs were classified into 4 groups with AtDCLs based on their sequence homologies (Fig. 1C).

Ten AGOs with specific biological functions exist in Arabidopsis. AtAGO1 (AT1G48410) is the most important protein and responds to most miRNA and siRNA-mediated mRNA cleavage or translational inhibition (10). In the *C. roseus* transcriptome database, only *CrAGO1* (DDS12000), *CrAGO4* (DDS13330), *CrAGO5* (DDS58064), *CrAGO7* (DDS11464, and DDS47116), *CrAGO10.1* (DDS42136) and *CrAGO10.2* (DDS42372) were identified, indicating that these genes were expressed in flower tissues of HF and S4. Note that DDS11464 belongs to the 5' end of *CrAGO7*, whereas DDS47116 belongs to the 3' end of *CrAGO7* (Fig. 3A). In addition, AOG7 is represented as a partial sequence because its N-terminus was missing (Fig. 3A). Two CrAGO10 orthologs exist in *C. roseus* and have 86.5 to 74.9% similarity with AtAGO10 (AT5G43810) (Fig. 1B).

In addition, CrAGO1 has 74.5% similarity with AtAGO1, whereas other CrAGOs show approximately 60% similarity with Arabidopsis AGOs (Fig. 1B). Like DCLs, CrAGOs show high conservation in their PAZ, and PIWI domains (Fig. 3A). The phylogenetic tree shows that AGO1 and AGO10 can be classified into the same group (Fig. 1D). Indeed, AGO10 competes with AGO1 for miR165/166 binding in shoot apical meristem development, suggesting that functional redundancy exists between AGO1 and AGO10 (11). In addition, AGO4 and AGO6 have been classified into the same group, and both have been shown to be functionally redundant in RdDM (8, 12).

RNA-DEPENDENT RNA POLYMERASE (RDR) plays a role in synthesis of double-stranded RNA for DCL2, DCL3 and DCL4 cleavage to produce 22-, 24- and 21-nt siRNAs, respectively (5-7). In the *C. roseus* database, *CrRDR1* (DDS16060), *CrRDR2* (DDS47054) and *CrRDR6* (DDS12885) were identified as *AtRDR* orthologs of Arabidopsis and showed 53.3 to 77.4% similarity with Arabidopsis genes (Figs. 1B and 4A). The phylogenetic tree of RDRs includes 3 groups, as shown in Figure 1E. Other PTGS-related genes, such as *CrHYLI* (DDS45028) (Fig. 2B), *CrHENI* (DDS46990) (Fig. 3B), *CrWEX* (DDS13870) (Fig. 3C), *CrHST* (DDS41673) (Fig. 4B), *CrSDE3* (DDS51099) (Fig. 4C), *CrSGS3.1* (DDS20269) and *CrSGS3.2* (DDS42452) (Fig. 4D), were identified in this study.

RNA polymerase IV and V are important components in plant 24-nt siRNA biogenesis and the RdDM mechanism plants (15, 16). In Arabidopsis, *AtRDR1* can be induced by salicylic acid (SA), and *NaRDR1* also can be induced by jasmonic acid in *Nicotiana attenuata* (16). Our previous study also indicated that *CrPRI* and other SA response genes were also induced in response to S4 leafy flower (2); therefore, we presume that increasing *CrRDR1* expression triggers the defense response.

We conclude that this high-throughput genomics approach provides an advantage for studying the molecular biology of non-model organisms. In this study, the epigenetic-related genes of *C. roseus* were identified and characterized. We found that PTGS and RdDM maintained

(8). CrNRPD1 (DDS47753) and CrNRPE1 (DDS16069) are the orthologs of Arabidopsis RNA polymerase IV and V large subunits, respectively. They showed 57.8% and 31.1% similarity to AtNRPD1 and AtNRPE1, respectively (Figs. 1B and 5A). Moreover, CrNRPD2 (DDS13297) is an ortholog of the second large subunit of RNA polymerase IV (AtNRPD2; AT3G23780) in Arabidopsis, with 75.9% similarity (Figs. 1B and 5B). The other RdDM-related genes, such as *CrDRD1* (DDS58), *CrDMS3* (DDS15300), *CrSPT5L* (DDS41978), *CrIDN2* (DDS19408) and *CrSUVH9* (DDS45142), were identified in the *C. roseus* database (Fig. 6) and showed 44.6 to 65.8% similarity with their Arabidopsis orthologs (Fig. 1B).

We compared the epigenetic-related gene expression levels in the HF and S4 samples based on the fragments per kilobase of transcript per million mapped reads (FPKM). The data indicated that most PTGS- and RdDM-related genes did not show different expression levels between HF and S4 (Fig. 1F), suggesting the epigenetic regulation have to keep the stabilization. However, CrAGO5 and CrRDR1 showed significantly increased expression levels during PnWB infection (Fig. 1F). The AGO phylogenetic tree indicated that AGO5 is closely related to AGO1 and AGO10 (Fig. 1D). Moreover, recent studies have indicated that AGO5 might have the same function as AGO1 to play a role in megagametogenesis during reproductive development (13, 14).

In addition, recent studies have indicated that RDR1 functions in anti-viral silencing in Arabidopsis and other normal gene expression during PnWB infection. However, CrAGO5 and CrRDR1 might involve in PnWB-mediated leafy flower formation and plant defense.

ACKNOWLEDGMENT

This study was supported by a grant from the National Science Council (NSC 102-2313-B-002-068-MY3) and National Taiwan University (103R7602B2).

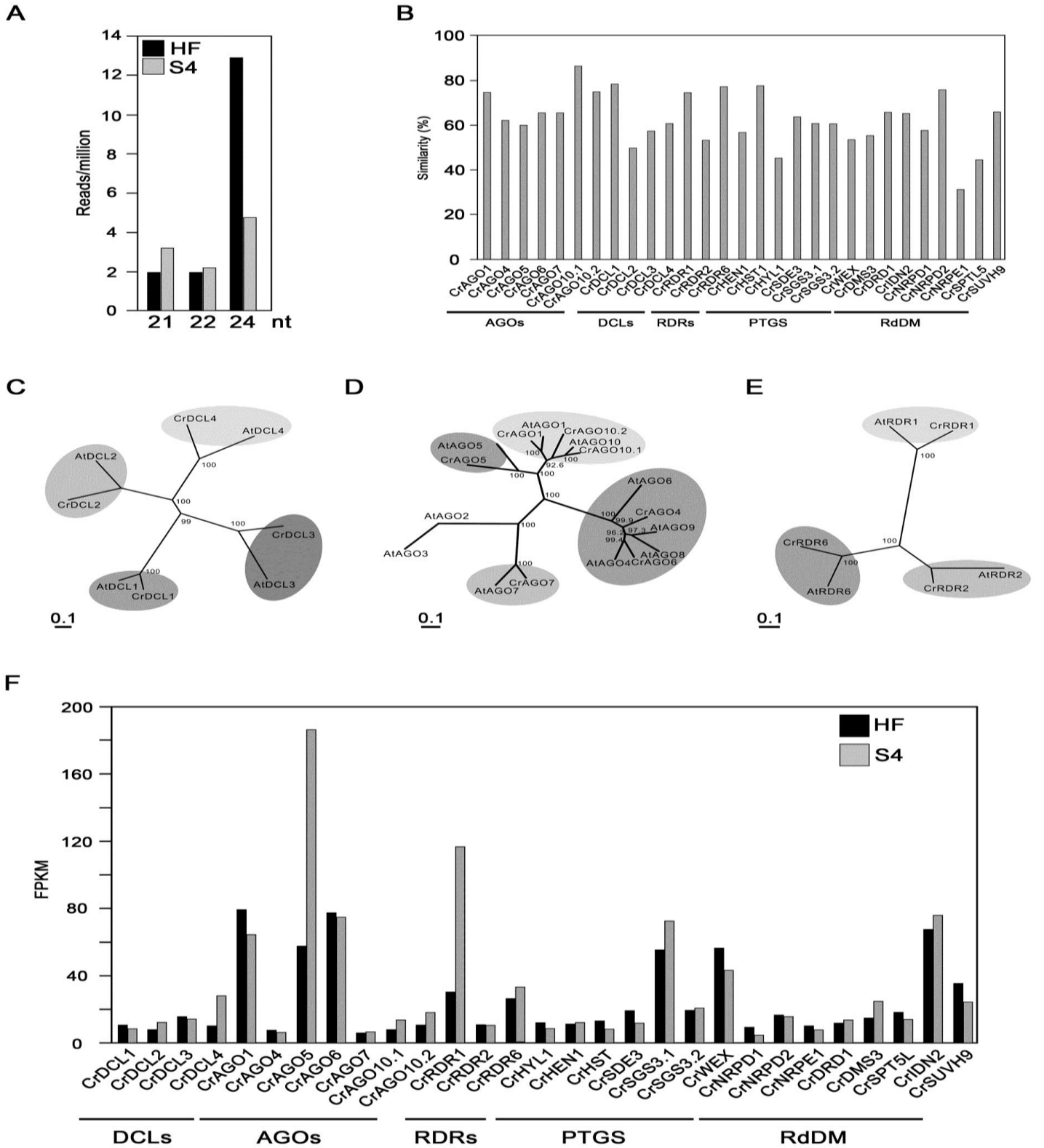


Figure 1. Small RNA and epigenetic-related gene expression levels in healthy flowers (HF) and Stage 4 (S4) peanut witches'-broom phytoplasma-infected leafy flowers of *Catharanthus roseus*. (A) The small RNA length distribution profiles of HF and S4. Total reads represents the amount of the corresponding small RNA. (B) The amino acid sequence similarity of epigenetic-related genes in Arabidopsis and *C. roseus*. (C) The phylogenetic tree of DICER-like (DCL) between Arabidopsis and *C. roseus*. (D) The phylogenetic tree of ARGONAUTE (AGO) between Arabidopsis and *C. roseus*. (E) The phylogenetic tree of RNA-DEPENDENT RNA POLYMERASE (RDR) between Arabidopsis and *C. roseus*. (F) Comparison of epigenetic-related gene expression levels in *C. roseus*, as determined using next-generation sequencing (NGS). The fragments per kilobase of transcript per million mapped reads (FPKM) value was used to indicate the gene expression levels that were determined using the NGS of HF and S4 samples.

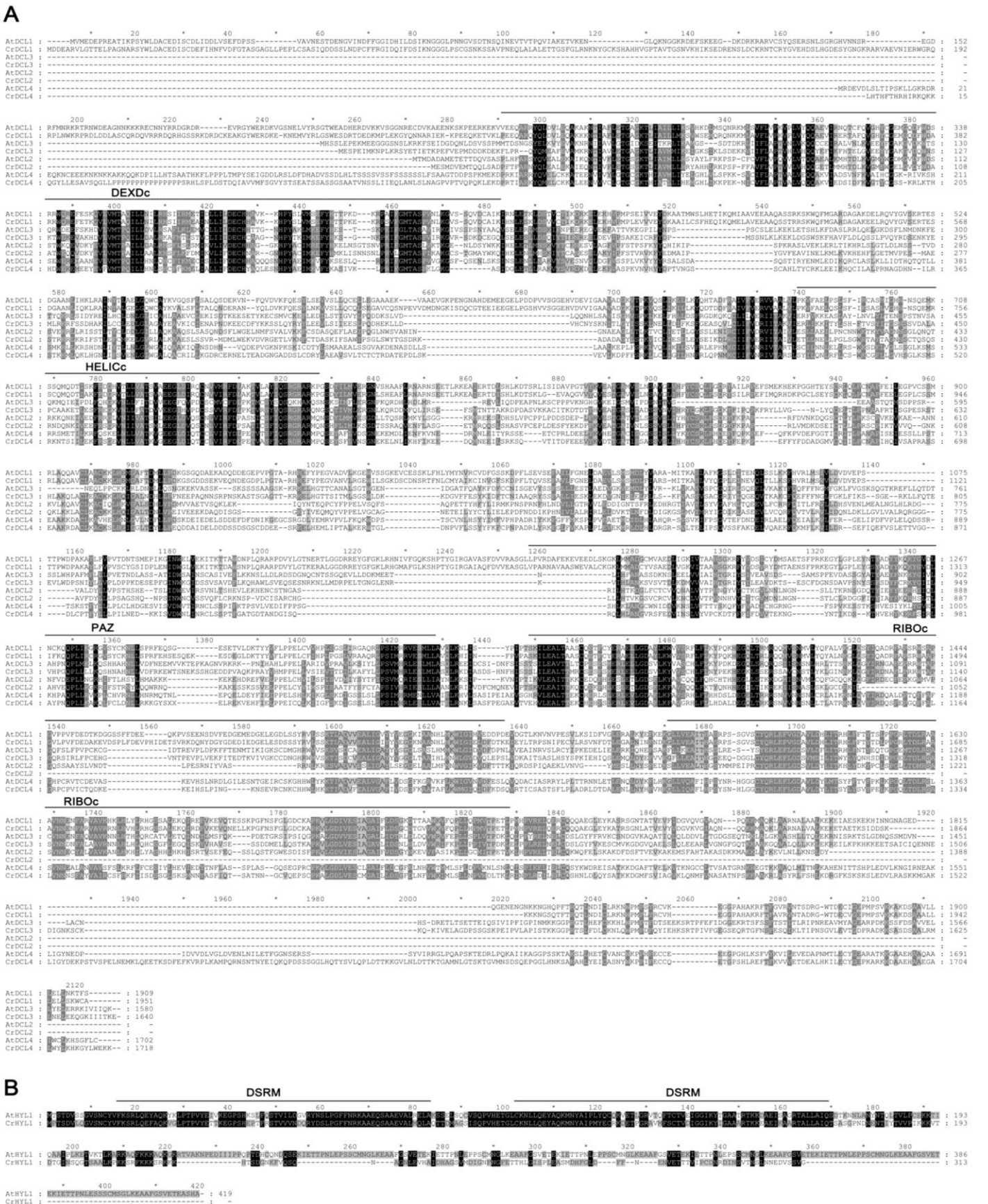
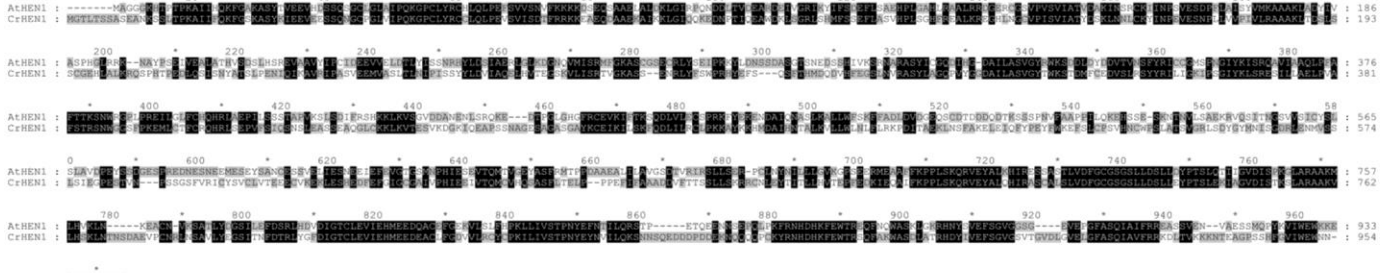


Figure 2. The amino acid sequence alignment of DICER-LIKE (DCL) and HYPONASTIC LEAVES 1 (HYL1). (A) The DCL alignments of Arabidopsis and *Catharanthus roseus*. The undetected amino acids are marked with an “X”. DEAD-like helicases domain (DEXDc). Helicase superfamily c-terminal domain (HELICc). Piwi Argonaut and Zwiille domain (PAZ). Prokaryotic ribonuclease III (RIBOc). (B) The HYL1 alignment of Arabidopsis and *C. roseus*. The specific protein domains are labeled on the sequence alignment. Double-stranded RNA binding motif (DSRM).

A



B



C

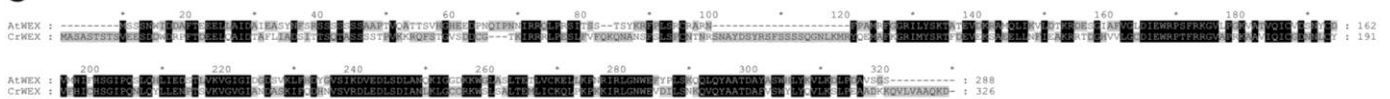


Figure 3. The amino acid sequence alignment of ARGONAUTE (AGO), HUA ENHANCER 1(HEN1) and WERNER SYNDROME-LIKE EXONUCLEASE (WEX). (A) The AGO alignment between *Arabidopsis* and *Catharanthus roseus*. The undetected amino acids are marked with an “X”. Piwi Argonaute and Zwiille domain (PAZ), P-element induced wimpy testis (PIWI). (B) The HEN1 alignment between *Arabidopsis* and *C. roseus*. (C) The WEX alignment between *Arabidopsis* and *C. roseus*. The specific protein domains are labeled on the sequence alignment.

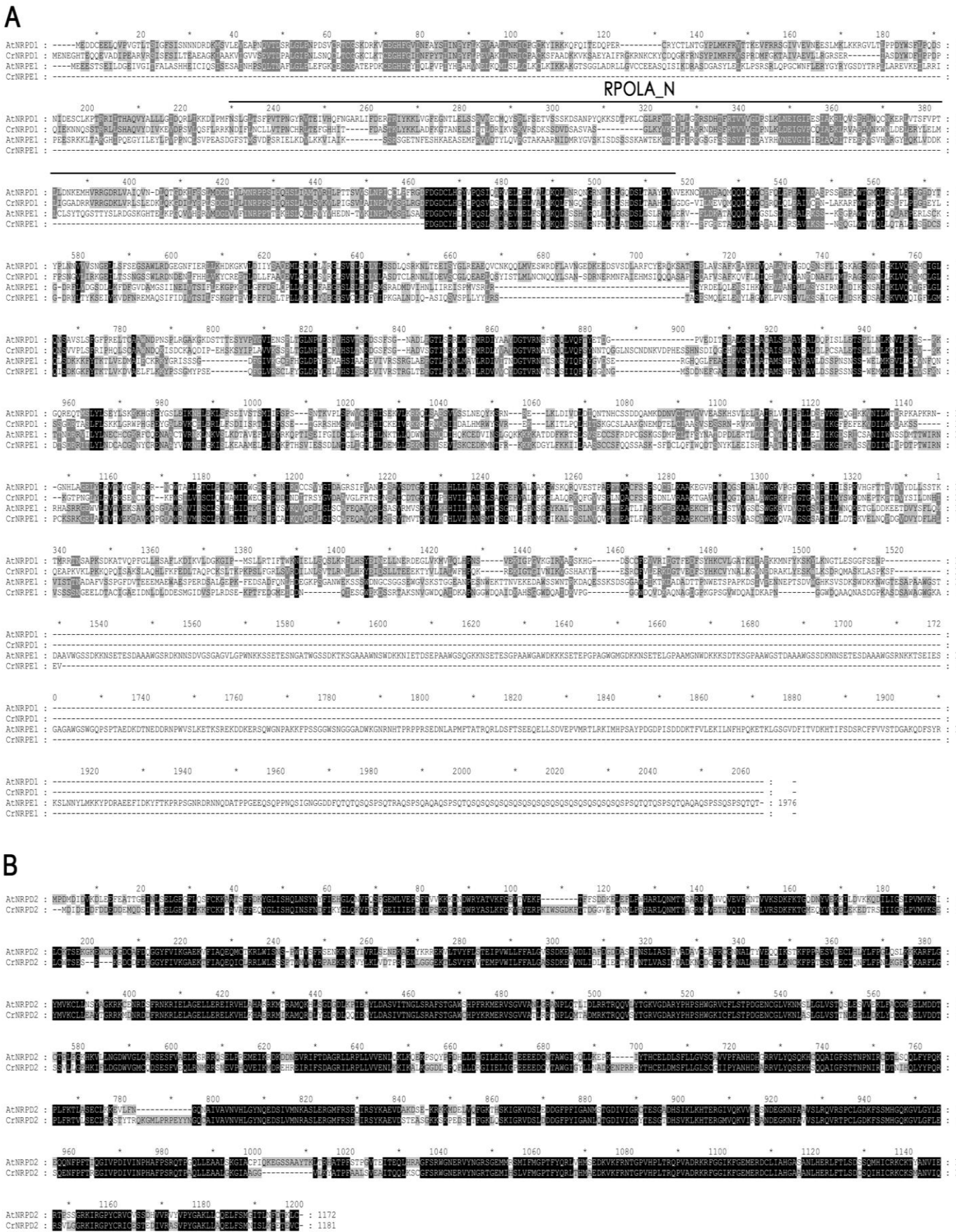


Figure 5. The amino acid sequence alignment of NUCLEAR RNA POLYMERASE D1 and E1 (NRPD1/E1), and NRPD2. (A) The NRPD1/E1 alignment between Arabidopsis and *Catharanthus roseus*. RNA polymerase I subunit A N-terminus (RPOA_N). (B) The NRPD2 alignment between Arabidopsis and *C. roseus*. The specific protein domains are labeled on the sequence alignment.

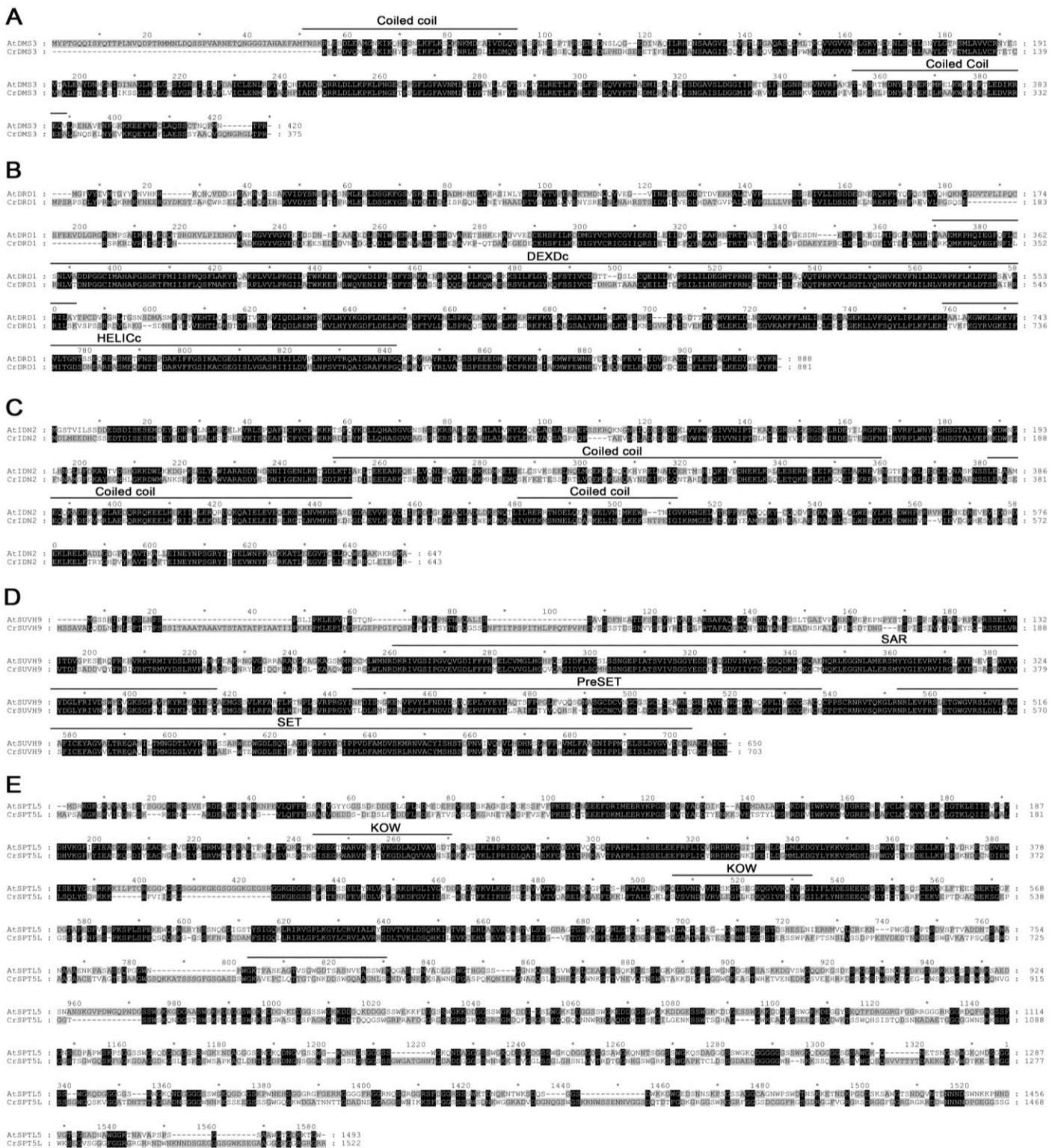


Figure 6. The amino acid sequence alignment of DEFECTIVE IN MERISTEM SILENCING 3 (DMS3), DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1), INVOLVED IN DE NOVO 2 (IDN2), SU(VAR)3-9 HOMOLOG 9 (SUVH9) and SUPPRESSOR OF TY INSERTION 5-LIKE (SPT5L). (A) The DMS3 alignment between Arabidopsis and *Catharanthus roseus*. Coiled coil domain (Coiled coil). (B) The DRD1 alignment between Arabidopsis and *C. roseus*. DEAD-like helicases domain (DEXDc). Helicase superfamily c-terminal domain (HELICc). (C) The IDN2 alignment between Arabidopsis and *C. roseus*. (D) The SUVH9 alignment between Arabidopsis and *C. roseus*. Sar1p-like members of the Ras-family of small GTPases (SAR). N-terminal to some SET domains (PreSET). Su(var)3-9, Enhancer-of-zeste, Trithorax domain (SET). (E) The SPT5L alignment between Arabidopsis and *C. roseus*. The specific protein domains are labeled on the sequence alignment. Kyprides, Ouzounis, Woese motif (KOW).

LITERATURE CITED

1. Chen, W.Y., and Lin, C.P. 2011. Characterization of *Catharanthus roseus* genes regulated differentially by peanut witches' broom phytoplasma infection. *J. Phytopath.* 159:505-510.
2. Liu, L.Y.D., Tseng, H.I., Lin, C.P., Lin, Y.Y., Huang, Y.H., Huang, C.K., Chang, T.H., and Lin, S.S. 2014. High-throughput transcriptome analysis of the leafy flower transition of *Catharanthus roseus* induced by peanut witches'-broom phytoplasma infection. *Plant Cell Physiol.* doi:10.1093/pcp/pcu029.
3. Vaucheret, H. 2006. Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes Dev.* 20:750-771.
4. Vaucheret, H., Vazquez, F., Cr  t  , P., and Bartel, D.P. 2004. The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. *Genes Dev.* 18:1187-1197.
5. Law, J.A., and Jacobsen, S.E. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11:204-220.
6. Zhang, H., and Zhu, J.K. 2011. RNA-directed DNA methylation. *Curr. Opin. Plant Biol.* 14:142-147.
7. Bouch  , N., Laressergues, D., Gascioli, V., and Vaucheret, H. 2006. An antagonistic function for *Arabidopsis* DCL2 in development and a new function for DCL4 in generating viral siRNAs. *EMBO J.* 25:3347-3356.
8. Daxinger, L., Kanno, T., Bucher, E., van der Winden, J., Naumann, U., Matzke, A.J., and Matzke, M. 2009. A stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of DNA methylation. *EMBO J.* 28:48-57.
9. Tesng, H.I. 2011. High-throughput transcriptome and small RNA analysis for studying phytoplasma infection on *Cantharanthus roseus* using Next Generation Sequencing. Master thesis. National Taiwan University.
10. Baumberger, N., and Baulcombe, D.C. 2005. *Arabidopsis* ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proc. Natl. Acad. Sci. USA* 102:11928-11933.
11. Zhu, H., Hu, F., Wang, R., Zhou, X., Sze, S.H., Liou, L.W., Barefoot, A., Dickman, M., and Zhang, X. 2011. *Arabidopsis* Argonaute10 specifically sequesters miR166/165 to regulate shoot apical meristem development. *Cell* 145:242-256.
12. Zheng, X., Zhu, J., Kapoor, A., and Zhu, J.K. 2007. Role of *Arabidopsis* AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. *EMBO J.* 26:1691-1671.
13. Tucker, M.R., Okada, T., Hu, Y., Scholefield, A., Taylor, J.M., and Koltunow, A.M. 2012. Somatic small RNA pathways promote the mitotic events of megagametogenesis during female reproductive development in *Arabidopsis*. *Development* 139:1399-1404.
14. Borges, F., Pereira, P.A., Slotkin, R.K., Martienssen, R.A., and Becker, J.D. 2011. MicroRNA activity in the *Arabidopsis* male germline. *J. Exp. Bot.* 62:1611-1620.
15. Liao, Y.W., Sun, Z.H., Zhou, Y.H., Shi, K., Li, X., Zhang, G.Q., Xia, X.J., Chen, Z.X., and Yu, J.Q. 2013. The role of hydrogen peroxide and nitric oxide in the induction of plant-encoded RNA-dependent RNA polymerase 1 in the basal defense against Tobacco mosaic virus. *PLoS One* 8:e76090.
16. Hunter, L.J., Westwood, J.H., Heath, G., Macaulay, K., Smith, A.G., Macfarlane, S.A., Palukaitis, P., and Carr, J.P. 2013. Regulation of RNA-dependent RNA polymerase 1 and isochorismate synthase gene expression in *Arabidopsis*. *PLoS One* 8:e66530.

摘要

曾欣怡^{1,4,5} 林長平^{1,4,5} 林詩舜^{1,2,3,6}. 2014. 日日春基因靜默機制相關基因參與花生簇葉病菌質體誘發花器葉化機制之探討. 植病會刊 23:67-77. (¹臺灣大學生物科技學研究所; ²臺灣大學基因體暨系統生物學學程; ³中央研究院農業生物技術研究中心; ⁴臺灣大學植物病理暨微生物學系; ⁵共同第一作者; ⁶聯絡作者, 電子信箱: linss01@ntu.edu.tw)

ContigViews 資料庫平台提供龐大的日日春植物 (*Catharanthus roseus*) 轉錄體基因資訊。為了研究表觀遺傳對於花生簇葉病植物菌質體 (peanut witches'-broom phytoplasma; PnWB) 誘發日日春花器葉化的影響。其參與在後轉錄時期基因靜默 (post-transcriptional gene silencing; PTGS) 及核醣核酸引發之去氧核醣核酸甲基化 (RNA-directed DNA methylation; RdDM) 相關基因已由此資料庫中選殖出來, 並與阿拉伯芥同源基因相互比較。這些相關基因的演化關係及在健康花 (HF) 與 PnWB 誘發第四型 (S4) 葉狀花的表現量都被詳細的研究。研究結果顯示, 日日春 *ARGONAUTE 5* (*CrAGO5*) 及 *RNA-DEPENDENT RNA POLYMERASE 1* (*CrRDR1*) 基因在 S4 葉狀花中表現量有明顯的增加。然而, 大部份的 PTGS 及 RdDM 相關基因在 HF 及 S4 樣本中的表現量都維持恆定, 顯示表觀遺傳的基因表現必須維持恆定。這些證據推測 *CrAGO5* 與 *CrRDR1* 可能參與在 PnWB 誘導之葉狀花及植物抗病反應。

關鍵字: 次世代基因定序、日日春、花生簇葉病植物菌質體、後轉錄時期基因靜默、核醣核酸引發之去氧核醣核酸甲基化