In silico assisted cloning and analysis of genes involving anti-viral defense based on RNA silencing in papaya (*Carica papaya* L.)

Aroonothai Sawwa[1,2,5], Wanwisa Siriwan[3], Sittiruk Roytrakul[4] and Srimek Chowpongpang[1,2,3,4,*]

[1]*Center for Agricultural Biotechnology, Kasetsart University Kampaengsaen Campus, Nakhorn Pathom, Thailand*
[2]*Center of Excellence on Agricultural Biotechnology: AG-BIO/PERDO-CHE, Bangkok, Thailand*
[3]*Department of Plant Pathology, Faculty of Agriculture, Kasetsart University, Bangkok, Thailand*
[4]*National Center for Genetic Engineering and Biotechnology (BIOTEC), Thailand Science Park, Pathum Thani, Thailand*
[5]*Biotechnology Research and Development Office, Department of Agricultural, Pathum Thani, Thailand*

**Corresponding author: agrsmcp@gmail.com*

ABSTRACT

Papaya (*Carica papaya* L.) is planted throughout tropical and sub-tropical regions. *Papaya ringspot virus* (PRSV) seriously limits papaya production. The RNA silencing mechanism can be mediated to mount an antiviral defense. *In silico* assisted cloning was a part of bioinformatic analysis to help identifying genes obtained from the whole genome shotgun database. The aim of this study was to use the *in silico* assisted cloning and expression to analyze the *RDR6*, *DCL4* and *AGO1* genes which involved in RNA silencing in papaya. The complete coding region sequences obtained from cloning of the *CpRDR*6 and *CpAGO1* genes were 3,588 bp (KF668595.1) and 3,180 bp (KY082908) encoding for 1,196 and 1,060 amino acids, respectively. The partial *CpDCL4* consisted of 3,207 bp (KY082909) encoding for 1,068 amino acids. The phylogenetic tree analyses could clearly differentiate the RDR6, AGO1 and DCL4 with other groups and showed the same relationships among plant species. The 3D structures of CpRDR6 and CpAGO1 were successfully generated and ligand binding site residues were identified. The expression patterns were investigated in different tissues and under PRSV infected condition at indicated time points of 1, 3, 5, 7, 9, 11 and 14 days post-inoculation. The results showed that the expression level of genes involved in RNA silencing could be induced by PRSV. The aid of *in silico* combining with freely available bioinformatics software would be useful and speed up the cloning, analysis of the target gene, and study of the structure of protein that have not been reported in papaya.

Keywords: Gene cloning; RNA-dependent RNA polymerase 6; Argonaute 1; Dicer-like ribonucleases 4

INTRODUCTION

Papaya (*Carica papaya* L.) is an economic fruit of tropical and subtropical countries. It is the most important species of *Caricaceae* family. The fruit is consumed world-wide as fresh fruit and vegetable or used as processed products (Teixeira *et al.*, 2007; Pedro *et al.*, 2009). It is a cheap and rich source of vitamins. Its latex included an enzyme papain, widely used in pharmaceutical and food industries (Lines *et al*., 2002). Annual world production of *C. papaya* in 2011 was more than10 million tons, making papaya an extremely important fruit crop (Carvalho and Renner, 2012). However, the papaya crop is severely affected by *papaya ringspot virus* (PRSV) worldwide. They are transmitted in a non-persistent manner by several species of aphids (Gonsalves, 1998). In Thailand, the papaya production has been generally declining since 2001 due to the outbreak of PRSV (Janthasri and Chaiayaboon, 2006). This virus causes a major disease of papaya and cucurbits and is found in all areas of the world where they are cultivated. PRSV is grouped into the papaya infecting type P (PRSV-P) which affects both papaya and cucurbits. It belongs to the genus *Potyvirus*, a large and economically important group of plant-infecting viruses in the family *Potyviridae*. Its genome is single-stranded positive sense RNA (Gonsalves *et al.*, 2010). Generally, the DNA as well as RNA viruses such as PRSV involve in activating RNA silencing for plant defense response (Voinet, 2001). The RNA silencing mechanism was applied to prevent virus infection in many genetically modified papayas worldwide.

RNA silencing is a process to induce translation inhibition at the post-transcriptional control of gene expression in plants. It causes sequence-specific RNA degradation, when triggered by double-

stranded RNA (dsRNA). The dsRNA involves in virus infection that replicates RNA to RNA through a dsRNA by viral RdRp (Nib protein). In post-transcriptional gene silencing (PTGS) level, the gene silencing process was recognized by dsRNA that had an effect on mRNA stability. The process cleaves the dsRNA into 21-25 nucleotide-long pieces of small interfering RNAs (siRNAs) as specificity determinant (Plasterk and Ketting, 2000; Singh *et al.*, 2006; Fire, 2007). The siRNAs correspond to both sense and anti-sense strands of the target gene. These siRNAs then become associated with a protein complex called the RNA-induced silencing complex (RISC) to direct the siRNA to the target mRNA, which is cleaved and subsequently degraded (Mlotshwa *et al*, 2002; Lele, 2009; Sliva and Schnierle, 2010). RNA silencing mechanism serves as an adaptive, antiviral defense system, which is transmitted systemically in response to localized virus challenge. RNA silencing-based antiviral defense entails the production of virus-derived siRNAs which guide specific antiviral effector complexes to inactivate viral genomes. As a response to this defense system, plant viruses have elaborated a variety of counter-defensive measures to overcome the host silencing response, viruses have evolved by improving the viral suppressors of RNA silencing (VSRs) to overcome the host defense (Voinnet, 2001; Shimura and Pantaleo, 2011).

RNA silencing is a powerful tool for gene function studies and crop improvements. Understanding of the endogenous gene silencing pathways was started in Arabidopsis. Plant RNA silencing machinery consists four primary classes of proteins to achieve the sequence-specific regulation of gene expression and mount the antiviral defense. These proteins include Dicer-like ribonucleases (DCLs), Argonaute proteins (AGOs), dsRNA-binding proteins (DRBs), and RNA-dependent RNA polymerases (RDRs). Various members of the *DCL*, *AGO*, *RDR*, and *DRB* gene families in Arabidopsis play central roles in the parallel gene silencing pathways, including the microRNA (miRNA), trans-acting siRNA (tasiRNA), natural-antisense siRNA (natsiRNA), and repeat-associated siRNA (rasiRNA) / RNA-directed DNA methylation (RdDM) pathways (Eamens *et al.*, 2008; Qu *et al.*, 2008). In plant lineages, the gene families encoding the core components of RNA silencing (RDR DCL and AGO) have diversified silencing pathways that control the expression of endogenous genes and viruses (Molnar *et al.*, 2011). Plants contain RDR activities that synthesize short complementary RNAs by using viral RNAs as templates and play an important role in plant antiviral defense (Xie *et al.*, 2001). RDR have been implicated in the PTGS. A likely biochemical role of RDR in RNA silencing is to produce dsRNA that is cleaved by DCL (Dalmay *et al*., 2000). Subsequently, AGO-small-RNA complexes repress the transcription of genes, target mRNAs for site-specific cleavage or general degradation, or blocking the translation of RNA into protein (Meng *et al*., 2013).

Viruses are both initiators and targets of gene silencing related to natural defense against viruses (Ratcliff *et al.*, 1997). For the virus-induced gene silencing (VIGS) model or anti-viral defense based on RNA silencing exploits the fact that infection by many RNA viruses activates a conserved, RNA-based plant antiviral defense response. The VIGS virus replication leads to the formation of dsRNA, which is detected by the plant. This step triggers PTGS through the production of siRNA, which targets the viral RNA for degradation (Scofield and Nelson, 2009; Gronlund *et al.*, 2010). The viral titer remains low, because RDR6 amplification and DCL4 digestion produce the signal to destroy the viral transcripts. (Blevins *et al.*, 2006; Kalantidis *et al.*, 2008).

The estimated genome size of papaya is 372 Mbp (Lai *et al.*, 2006). The first commercial virus-resistant transgenic variety, SunUp has been sequenced resulting in the 3x draft genome in 2008. A total of 2.8 million whole genome shotgun (WGS) sequencing reads were generated from a female plant of SunUp. The sequence information from WGS has been used extensively for analyzing and understanding structure and functional genomics (Ming *et al*., 2008; Sharma *et al.*, 2016). Advancements in sequencing technology have made it possible to obtain more information about the DNA sequence, structure and the transcript products of the genome from different species. This information is collected and kept in DNA databases. These databases contain many genes whose functions have not yet been discovered (Passier and Doevendans, 2004). The term '*in silico*' is a word commonly used to refer to experimentation performed by computer and is related to the more commonly known biological terms such as *in vivo* and *in vitro*. The history of the '*in silico*' term is poorly defined, with several researchers claiming their roles in its origination. In a more recent book, it provides a quotation that offers a concise and cogent depiction of the potential of computational tools in chemistry, biology and pharmacology. *In silico* are computational biology tools, molecular biology terminology, and the process of DNA cloning (Ekins *et al.*, 2007; Elkins, 2011) In this study, the *in silico*

assisted cloning was employed to successfully achieve the coding region sequences of genes involved in the anti- viral defense based on RNA silencing, *RDR6*, *AGO1* and the partial of *DCL4* from *C. papaya*. Moreover, protein structures encoded by these genes were predicted. This knowledge may lead to functional study and enables analyses of transcription levels of these genes in different tissues and the patterns under the PRSV infection.

MATERIALS AND METHODS

In silico assisted genes cloning by BLAST searching of whole genome shotgun sequence of *C. papaya* and primers designing

The RDR6, AGO1 and DCL4 amino acid sequences of *Arabidopsis thaliana* in NCBI database were used as the template for probe design. The amino acid sequences were reversely translated to DNA via Sequence Manipulation Suite website (http://www.bioinformatics.org/sms2/rev_trans.html) with codon usage database of *C. papaya*. The resulting DNA sequences were used to search for the *C. papaya* whole genome contig using BLAST website (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The obtained whole genome contigs were assembled, exon regions were predicted with FGENESH program (http://linux1.softberry.com/berry.phtml?topic=fgenesh &group=programs&subgroup=gfind) and the predicted exons were allocated on *C. papaya* genome sequences using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). After manually designing primer, the qualities of primers were analyzed using OligoCalc software (http://www.basic.northwestern.edu/biotools/oligocalc.html).

Plant materials, RNA extraction and cDNA synthesis

C. papaya, Khak-Dam Srisaket cultivar seeds were incubated at 42 °C for 5 hr before planting in a green house. The papaya plants were grown until 1 month old, tissues from leaves, stems, roots, and leave stalks were collected and were frozen immediately in liquid nitrogen. The samples were stored at -80 °C. At two months of age, papayas in pots were placed in a versatile environmental test chamber (MLR-350HR, Sanyo, Japan) set to15,000 lx of the light intensity, 8 hr of light and 16 hr of dark, the temperature at 30 ^{0}C during light period and 25 ^{0}C during dark period, and 75% humidity under either long day (Chumpookam *et al.*, 2012) for 3 days before PRSV type P inoculation. The PRSV Khonkhan strain was propagated in Khak-Dam Srisaket cultivar. One gram of leaf tissue from PRSV infected plants were ground in a mortar containing 0.02 M phosphate buffer (pH 7.0) and silicon carbide abrasive (carborundum) (Siriwan *et al*., 2014). Plant leaves were ruptured with inoculum solution. The PRSV infection was compared with Mock which was dusted with phosphate buffer. The samples were collected at 1, 3, 5, 7, 9, 11 and 14 days post-inoculation (DPI). The experiment was carried out in 5 replicates. The plant leaves were collected, frozen immediately in liquid nitrogen and stored at -80 ^{0}C.

For RNA extraction, 1 g of plant tissues was ground in liquid nitrogen using mortar and pestle. Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen Inc., USA). The integrity of each total RNA sample was assessed in agarose gels (1% w/v). RNA concentration and purity were determined with spectrophotometer at the A260/ A 280 nm using a BioDropμLITE analyzer (Biodrop Ltd., UK). Then, the first stand cDNA was synthesized from 500 ng of total RNA using the RevertAid™ First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, USA).

Cloning and sequence analysis

The *CpRDR6*, *CpAGO1* and *CpDCL4* genes were amplified from cDNA using conventional PCR with primers listed in Table 1 (Zhu *et al.*, 2012). The PCR protocol was used as described in High-Fidelity DNA polymerase kit (Thermo Fisher Scientific, USA). Amplification was carried out at 94°C for 5 min, 94°C for 30 sec, 60°C for 30 sec, 72°C for 3 min repeated for 35 cycles; then 72°C, 10 min. PCR products were separated on 1 % agarose gel. A candidate DNA band was ligated into pJET1.2/ blunt cloning vector (ClonJET PCR cloning Kit, Thermo Fisher Scientific, USA). The ligations were transformed into *E. coli* DH5α competent cells and DNA was sequenced. Finally, the nucleotide sequences were submitted to the NCBI database.

In silico analysis of genes

The *CpRDR6*, *CpAGO1*, and *CpDCL4* nucleotide sequences were translated into amino acid sequences by ExPASy translate tool (http://web.expasy.org/translate/). Phylogenetic analyses were performed using the maximum likelihood (ML) method in MEGA7 program (Kumar *et al*., 2016). Branch support was assessed with 500 bootstrap replicates. Physico-chemical properties of amino acid sequences were computed using ProtParam (https://www.expasy.org/tools/protparam.html) (Gasteiger *et al.*, 2005). Three dimensional structures of genes were predicted using I-TASSER server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) (Roy *et al.*, 2010; Yang *et al.*, 2015). The EduPymol Molecular Graphics System (https://pymol.org/edu/?q=educational/)

was used to visualize the 3D structure and generate model with its template model from I-TASSER. The cleavage sites of signal peptides were predicted by SignalIP4.1 (http://www.cbs.dtu.dk/services/SignalP/) (Petersen *et al*., 2011).

Table 1 Primers used in this study.

Primer name	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')
1RDR6	ATGGGKTCWGARGGWAATATGAARA	ATCATCACCAGCAGTTCTATACCAA
2RDR6	ATGGAGTTCTTGGTGAGAGAAATTACT	TCATATTCTATCAGCAAGATACCTT
1AGO1	ATGGTGAGGAAGAGGAGAACTGAAC	AGTTTATGCAGATCCAAGTATTCAC
2AGO1	AGAGGAGTCAACCGTGCTGCTATGC	TTAACAGTAGAACATGACTCGCTT
1DCL4	GATCCCAATGATAGAATTTATAGAAA	ACTTCTCAAGAGCAGCAACATCAAT
2DCL4	TGATACAGCTTCCCAGCCAAATGCT	TTATGTATGCTTGTTAACCGAAGG
CpRDR6	ATGGGGTCTGAAGGTAATATGAAGA	TCATATTCTATCAGCAAGATACCTT
CpAGO1	ATGGTGAGGAAGAGGAGAACTGAAC	AGTTTATGCAGATCCAAGTATTCAC
CpDCL4	GATCCCAATGATAGAATTTATAGAAA	TTATGTATGCTTGTTAACCGAAGG
TBP1*	GGTAGTAGTAGTTAGGTATGTG	GGCAATCTGGTCTCACTT
qRDR6	ATTACTCCCACAAAAGCATATCGTC	TGGTTGGATGAAAAGGCTAGAAAAG
qAGO1	TGAACTTGATGCTATTCGGAAGGC	TCCCATAGCACATGATAGTGAGCA
qDCL4	ATTGCTGATGTTGTCGAGGCTCTTG	GAAACTCCAATCTCTGATAGCAGCC

Gene expression analysis

The mRNA expression level of *CpRDR6 CpAGO1* and *CpDCL4* genes in leaves, stems, roots, and leave stalks were examined using two steps-quantitative reverse transcription- polymerase chain reaction method (qRT-PCR). First strand cDNA was synthesized from 20 ng/ul total RNA and amplified through real-time PCR TBP1 (TATA binding protein 1), a housekeeping endogenous gene, was used as a reference gene. (Zhu *et al.*, 2012). Transcriptional level of *CpRDR6 CpAGO1* and *CpDCL4* genes were performed using forward and reverse primers listed in Table 1 and SYBR® Green PCR Master Mix kit (Applied Biosystems, USA) on The QuantStudio® 5 Real-Time PCR system (Applied Biosystems, USA). The reaction procedure was 95°C for 10 min followed by 40 cycles of 95°C for 15 sec, 60°C for 15 sec, 72°C for 20 sec; and a melt curve stage of 95°C for 15 sec, 60°C for 60 sec and 95°C for 15. The experiment was carried out in 3 replicates. The reaction ratios were quantified using the comparative C_t ($\Delta\Delta C_t$) method (Rao *et al*., 2013).

RESULTS AND DISCUSSION

***In silico* assisted gene cloning**

Sequences of three *A. thaliana* genes, *RDR6* (accession number NP_190519.1), *AGO1* (accession number U91995.1) and *DCL4* (accession number NM_122039) were retrieved from NCBI and their amino acid sequences were converted into DNA sequence using Sequence Manipulation Suite website with codon usage of *C. papaya*. The predicted *C. papaya* RNA-dependent RNA polymerase 6 (*Cp*RDR6) gene was matched with the whole genome shotgun sequence of *C. papaya* cultivar SunUp chromosome LG3 contig 18548 (ABIM01018523.1). This contig was used as the template for primer design. On the other hand, 5' RDR6 position which approximate length of 500 bp, was not matched with the whole genome shotgun in NCBI database. Therefore, forward primer was designed as the degenerate primer based on *A. thaliana RDR6* and *Nicotiana tabacum RDR6* genes. The DNA sequences of PCR products with 707 bp and 3,000 bp were determined and the primers specific to the complete *CpRDR6* CDS were designed as shown in Figure 1A. The completed coding *CpRDR6* contained 3,588 bp in length as shown in Figure 2.

For the *C. papaya* Argonaut 1 (*CpAGO1*) gene, the predicted exon from Sequence Manipulation Suite website was hit to contig 29130 (ABIM01029086.1) and contig 29131 (ABIM01029087.1) in the whole genome shotgun sequence of *C. papaya* cultivar SunUp chromosome. The *CpAGO1* contig 29130 and 29131 were sequenced, the PCR products were 1,800 bp and 2,484 bp. The specific primers were designed and used to amplify complete *CpAGO1* CDS (Figure 1B). The *CpAGO1* CDS with 3,180 bp in length was analyzed on agarose gel electrophoresis as shown in Figure 2.

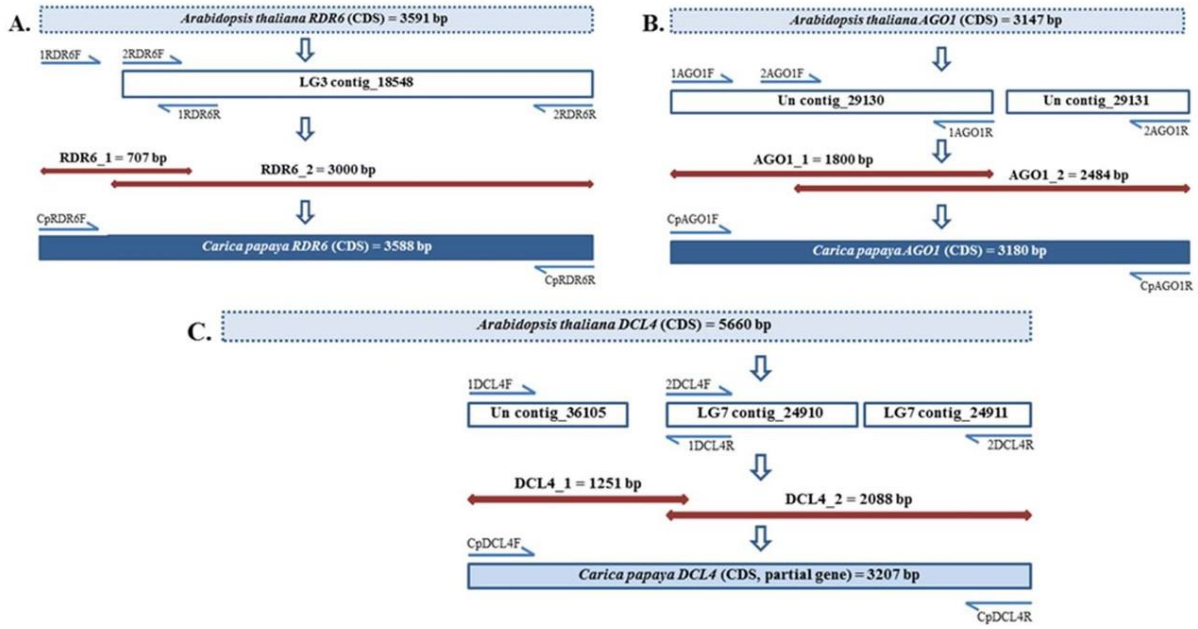

Figure 1 The schematic representation of genes prediction and cloning. Nucleotide and amino acid sequences of Arabidopsis were used as the template for gene prediction. (A) The assembled sequence of complete CDS of *CpRDR6* contained 3,588 bp, (B) the assembled sequences of *CpAGO1* contained 3,180 bp, (C) the assembled sequences of partial CDS of *CpDCL4* contained 3,207 bp.

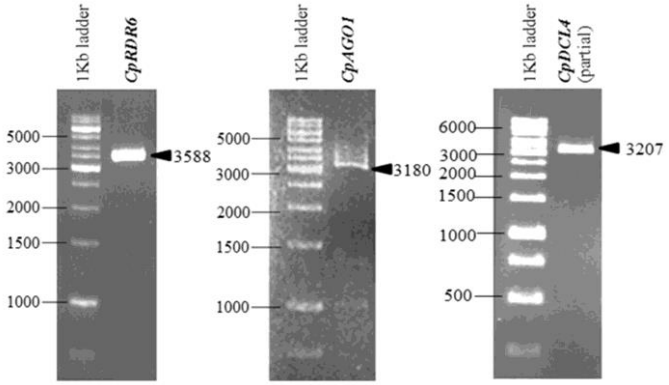

Figure 2 PCR amplification of *CpRDR6, CpAGO1*, and *CpDCL4* genes from cDNA template analyzed on 1% agarose gel.

Although the length of Dicer-like 4 (*CpDCL4*) gene in *Arabidopsis thaliana* is 5660 bp, the predicted *C. papaya* Dicer-like 4 (*CpDCL4*) gene was only 3207 bp. The predicted exons based on *A. thaliana* were matched with *C. papaya* cultivar SunUp chromosome Un contig_36105 (ABIM01036018.1), LG7 contig 24910 (ABIM01024874.1) and LG7 contig_24911 (ABIM01024875.1). Unfortunately, upstream region of CpDCL4 had not been found in *C. papaya* whole genome and could not be amplified by degenerate primer. Upstream fragment of *CpDCL4* were amplified using specific primers from contig 36105, another fragment was from contig 24910 and 24911, yielding 1,251 and 2,088 bp products (Figure 1C). After DNA sequence analyses, the specific primers CpDCL4f and CpDCL4r were used to amplify the partial CpDCL4 gene. A 3,207 bp-fragment CpDCL4 analyzed on agarose gel was observed in Figure 2.

The lack of *C. papaya* whole genome information caused an incomplete CDS of many genes including CpDCL4. However, the bioinformatically predicted *CpDCL4* is useful for further study. The 5' RACE (Rapid amplification of cDNA end) technique would be an alternative approach to complete the *CpDCL4* gene.

CpRDR6 showed 75%, 73% and 71% sequence identity with *Morus notabilis* (XM_010102964.1), *Arabidopsis lyrata* (XM_002877635.2) and *Nicotiana glutinosa* (FJ490363.1), respectively. The *CpAGO1* complete CDS showed 80%, 79% and 78% sequence identity with *Glycine max* (NM_001353328), *Malus hupehensis* (KU926967.1) and *Salvia mittiorrhiza* (KF153679.1), respectively. The partial *CpDCL4* showed 77% and 73% sequence identity with *Thebroma cacao* (LT594796.1) and *Solanum lycopersicum* (NM_001279281.2) respectively. In this study, *A. thaliana* was used as the template for gene prediction, whereas, *CpAGO1* and *CpDCL4* sequences showed no similarity with *A. thaliana*. In agreement with Lai *et al.*, (2006), high degree of apparent synteny in the papaya genome was more related to Populus than *Arabidopsis*.

Phylogenetic analysis

The phylogenetic tree was constructed based on the amino acid sequence from Genbank database using the maximum likelihood (ML) method to investigate the evolutionary relationship among CpRDR6 and other RDRs. The result revealed that CpRDR6 was more closely related to *Arabidopsis* RDR6s than other RDRs i.e. *Gossipium* and *Nicotiana*, respectively. The RDR1 and RDR2 were clustered together but distinguishable separated from RDR6 cluster. However, *A. thaliana* RDR3 was an out-group (Figure 3). Similar to *Gossipium*, GhRDR6 showed closer evolutionary relationship to various RDR6s than to other RDRs (Wang *et al*., 2012).

The amino acid sequence of CpAGO1 was clustered in the AGO1s group showing high relationship with *G. arboretum* AGO1B, *G. arboretum* AGO1-like, *N. tabacum* AGO1, *A. thailiana* AGO1 and *Oryza sativa* AGO1, respectively. This phylogenetic tree showed that the *N. attenuate* AGO10 and *N. attenuate* AGO5 were in the same group as AGO1s (Figure 4). The *Triticum aestivum* AGO1B was found to share a high degree of homology with *O. sativa* AGO1B, *T. aestivum* AGO1, *O. sativa* AGO1A, *A. thailiana* AGO1, AGO10 and AGO5 (Meng *et al*., 2013)

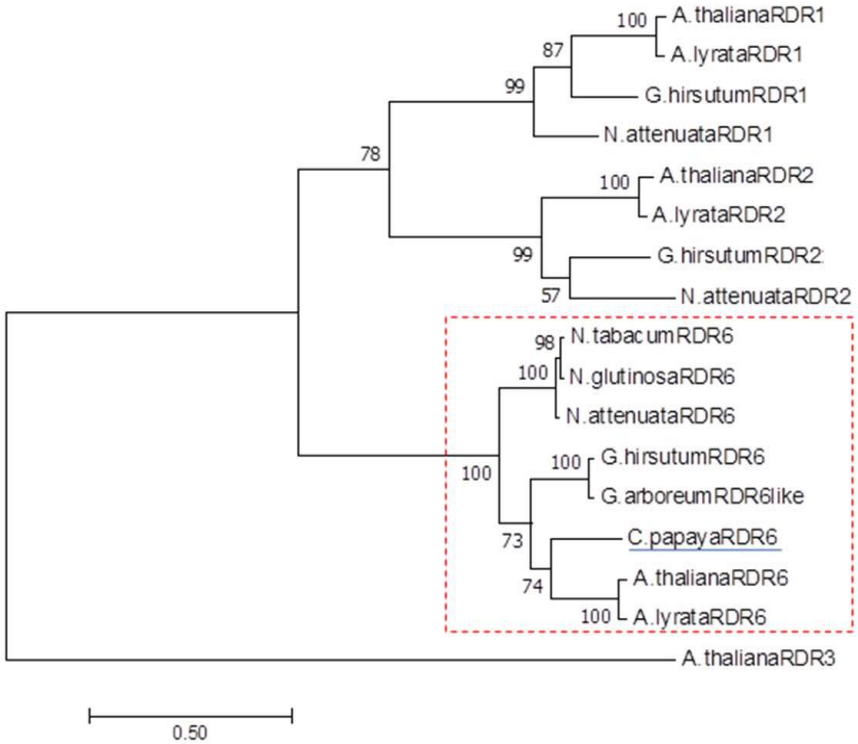

Figure 3 Phylogenetic tree of *C. papaya* RDR6 was generated based on the deduced amino acid sequences with other sequences from other plants including *A. thaliana* RDR1 (AY148431), *A. thaliana* RDR3 (O82190.2), *A. thaliana* RDR2 (NM_117183.3), *A. thaliana* RDR6 (NM_114810.3), *A. lyrata* RDR1 (EFH69089.1), *A. lyrata* RDR2 (EFH48810.1), *A. lyrata* RDR6 (EFH53940.1), *G. hirsutum* RDR1 (DQ445607.1), *G. hirsutum* RDR2 (KHG14638.1), *G. hirsutum* RDR6 (GQ254649.1), *G. arboreum* RDR6-like (KHG28504.1), *N. attenuata* RDR1 (OIT37821.1), *N. attenuata* RDR2 (OIT04938.1), *N. attenuata* RDR6 (OIT39130.1), *N. tabacum* RDR6 (ADI52625.1) and *N. glutinosa* RDR6 (ACO72600.1).

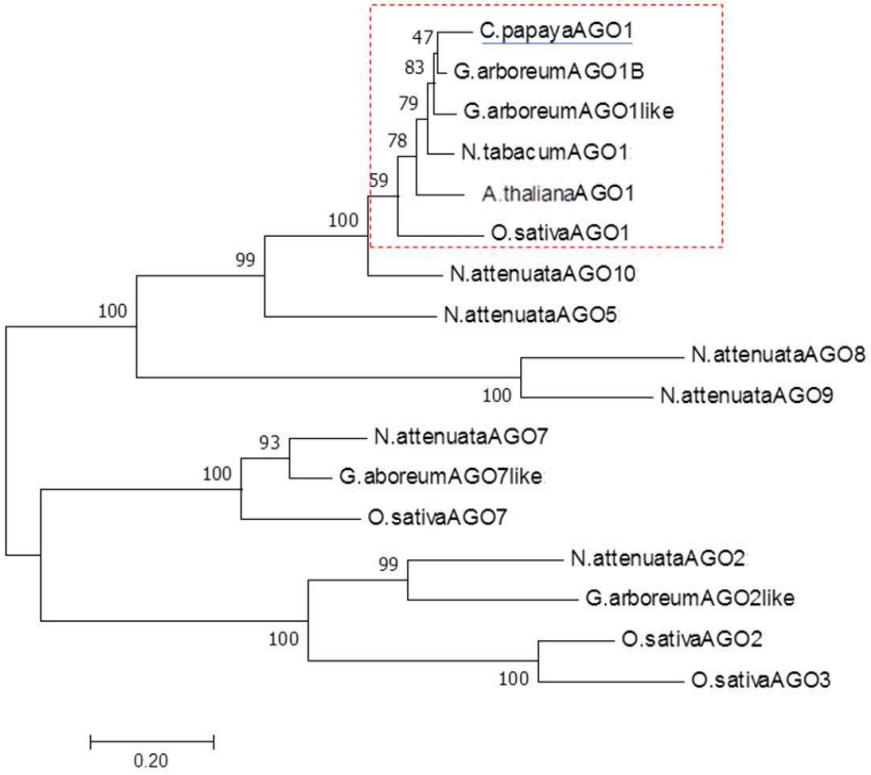

Figure 4 Phylogenetic tree of *Carica papaya* AGO1 was generated based on the deduced amino acid sequences with other sequences from other plants including *Nicotiana tabacum* AGO1 (BAJ09698.1), *Nicotiana attenuata* AGO2 (JAC96546.1), *Nicotiana attenuata* AGO5 (JAC96543.1), *Nicotiana attenuata* AGO7 (JAC96542.1), *Nicotiana attenuata* AGO8 (JAC96541.1), *Nicotiana attenuata* AGO9 (JAC96540.1), *Nicotiana attenuata* AGO10 (JAC96539.1), *Arabidopsis thaliana* AGO1 (U91995.1), *Gossypium arboreum* AGO1-like (KHG29379.1), *Gossypium arboreum* AGO1B (KHG08402.1), *Gossypium arboreum* AGO2-like (KHG20055.1), *Gossypium aboreum* AGO7-like (KHG28070.1), *Oryza sativa* AGO1 (BAD23006.1), *Oryza sativa* AGO2 (Q7XTS4.2), *Oryza sativa* AGO3 (Q7XTS3.2) and *Oryza sativa* AGO7 (ABO93307.1).

The partial CpDCL4 protein sequence was compared to DCLs from other plants. The phylogenetic tree of DCL4s showed that CpDCL4 was classified in the same DCL4 group. Generally, plants DCL4 could be grouped together, except *Physcomitrella patens* DCL4 (Bozorov *et al*., 2012). Interestingly, the highest relationship with *A. thaliana* DCL4 was investigated (Figure 5).

The *RDR6*, *AGO1* and *DCL4* are involving in RNA silencing which plays essential roles in plant development and antiviral defense through regulating gene expression at the transcriptional and post-transcriptional levels (Shao and Lu, 2014). They have been identified in various plants, such as *Arabidopsis* (Peragine *et al.*, 2004), *Nicotiana* (Yang *et al.*, 2011) and *Gossypium* (Wang *et al*., 2012). Although, RDR6, AGO1, and DCL4 gene in *C. papaya* was computationally predicted and derived from genomic sequence of WGS in NCBI. The information of cloned genes is required to study their function of genes and proteins.

***In silico* protein analysis of CpRDR6 and CpAGO1**

Since the partial- length sequences of *CpDCL4* cloning was obtained, *in silico* protein characterization of full coding sequence CpRDR6 and CpAGO1 were thus performed. The amino acid sequences of CpRDR6 and CpAGO1 were used as input data to predict their physico-chemical properties using ProtParam (Gasteiger *et al.*, 2005). The 1196- and 1060- residues of CpRDR6 and CpAGO1 were computed for their molecular weights, theoretical p*I*, amino acid compositions, atomic compositions, extinction coefficients, instability indexes, aliphatic indexes and grand average of hydropathicity as listed in Table 2.

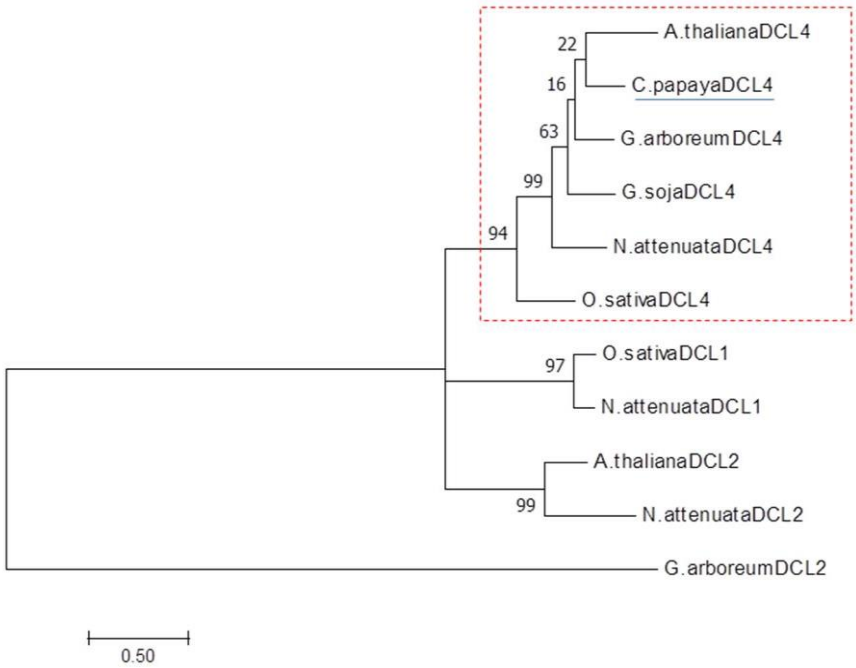

Figure 5 Phylogenetic tree of *C. papaya* DCL4 was generated based on the deduced amino acid sequences with other sequences from other plants including *Glycine soja* DCL4 (KHN22101.1), *O. sativa* DCL1 (Q8LMR2.1), *O. sativa* DCL4 (A7LFZ6.1), *G. arboretum* DCL2 (KHG08389.1), *G. arboretum* DCL4 (KHG18654.1), *A. thaliana* DCL2 (AEE73924.1), *A. thaliana DCL4* (NM_122039.5), *N. attenuate* DCL1 (OIS99363.1), *N. attenuate* DCL2 (AFD22619.1) and *N. attenuate* DCL4 (AFD22621.1).

Table 2. Predicted physico-chemical properties of full coding sequences CpRDR6 and CpAGO1.

No.	Parameter	Theoretical prediction	
		CpRDR6	**CpAGO1**
1	Molecular weight (kDa)	136.59	117.22
2	Isoelectric point (Theoretical pI)	8.30	9.48
3	Total no. of negatively charged residues (Asp + Glu)	144	98
4	Total no. of positively charged residues (Arg + Lys)	154	134
5	Extinction coefficient (M^{-1} cm^{-1}, at 280 nm)	187140	103235
6	Instability index	36.07	48.76
7	Aliphatic index	82.21	73.20
8	Grand average of hydropathicity (GRAVY)	-0.290	-0.496

Three-dimensional (3D) structures of proteins were better conserved during the course of evolution compared with their amino acid sequences. Similar amino acid sequences will give rise to similar 3D structures (National Research Council (US) Committee, 1987). Therefore, the 3D modeling and threading of these proteins were generated using I-TASSER software, an integrated platform for automated protein structure and function prediction. This platform was starting from an amino acid sequence to generate three dimensional atomic models from multiple threading alignments and iterative structural assembly simulations (Roy *et al*., 2010). The refined CpRDR6 and CpAGO1 3D structures were successfully generated and ligand binding site residues were identified. I-TASSER showed that CpRDR6 and CpAGO1 hit with 3s17A and 4olbA in the Protein Data Bank (PDB) with the C-score of 0.34 and 0.22, respectively. CpRDR6 showed the ligand binding site residues at the position Asn820, Leu865, Gly867 and Leu869. In the meantime CpAGO1 found at position Ser384, Ser385, Arg493, Arg509, Ile519, Leu523, Thr526, Cys527, Asp689, Asn690, Asn691, Tyr695, Lys699, Gln711, Cys712, Cys713, Leu714, His717, Tyr725, Asn728, Val729, Lys732, Lys736, Lys892, Arg897, His938, Ile941, Gln942, Ile944, Ser945, Arg946, Tyr975, Arg977, Cys978, Arg980, Ser983, Tyr989, Arg997 and Phe1000 (Figure 6A and 6B). The

cleavage position of the signal peptide was predicted using SignalIP4. 1 server, CpRDR6 and CpAGO1 showed cleavage residue positions at Lys27 and Gly63 with scores of 0. 237 and 0. 111, respectively. The computational tools are available for detecting signal peptides, but their abilities to locate the signal peptide cleavage site is not very significant and often less satisfactory. The SignalP was one of the best programs for predicting signal sequence with an accuracy of 78.1 % for cleavage site recognition (Zhang and Henzel, 2004). The resulting 3D figures from I-TASSER were generated to show the terminals and their signal peptide cleavage sites using EduPymol Molecular Graphics System (Figure 6C and 6D).

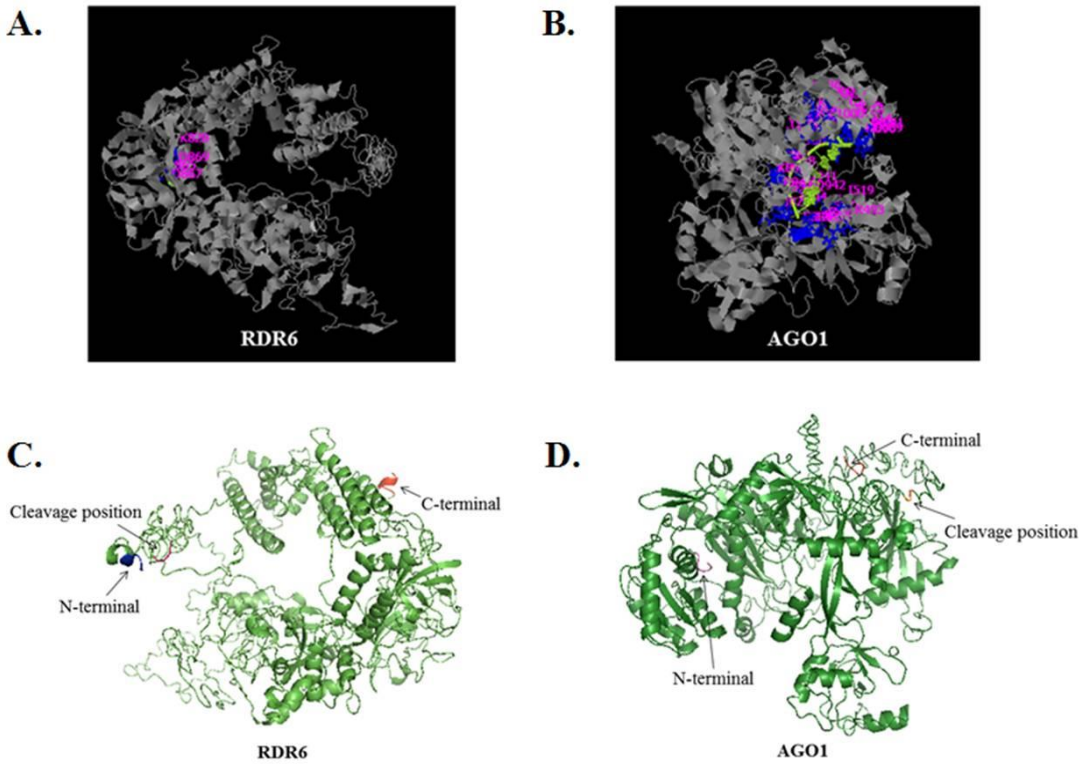

Figure 6 Three- dimensional models of CpRDR6 and CpAGO1 predicted by I- TASSER showing the ligand binding site residues of CpRDR6 (A.) and CpAGO1 (B.) presented in blue. C. and D. showed the cleavage positions of signal peptide predicted from SignalP.

In this study, the nucleotide sequences of *CpRDR6, CpAGO1*, and *CpDCL4* in *C. papaya* were therefore predicted, cloned, phylogenetically analyzed and further characterized for their protein structures with the aid of freely available bioinformatic tools. The *in silico* cloning is absolutely a robust and feasible tool for gene cloning and presents many advantages including rapid screening, low cost, high efficiency and easy operation when compared to the traditional methods (He *et al*., 2010; Baghban and Nayeri, 2016)

Gene expression analysis by qPCR

The qPCR is one of the best methods available to determine the gene expression at transcriptional level. The expression levels of *CpRDR6 CpAGO1* and *CpDCL4* were determined using mRNA isolated from leaves, stems, roots, and leave stalks of 1- month- old seedling through qPCR analysis. The transcription of three genes could be detected in all tissues, consistent with the vital roles of RNA silencing in plants. The highest gene expression level was found in leaves. On the other hand, the *CpRDR6* and *CpDCL4* mRNA abundances were lower in stems, and*CpAGO1* decreased in the leaf stalk tissues (Figure 7). Willmann *et al.*, 2011 reported that the expression level of RDR6 in *Arabidopsis* was still consistent in all tissues, whereas, the *Gossypium hirsutum* showed highest expression level of RDR6 in roots, which might regulate root growth and participate in development during the seedling stage in cotton (Wang *et al.*, 2012).

In order to investigate the relationship between expression pattern of papaya RNA silencing

genes during PRSV infection, the expression levels of *CpRDR6*, *CpAGO1*, *CpDCL4* and PRSV *coat protein* genes were analyzed using qPCR approach. After inoculating for 1 day, expression levels of all RNA silencing genes were higher in the PRSV- infected plants compared to the mock control. The coat protein gene of PRSV was detected in 3 days post-inoculation, and reached maximum at 14 DPI (Figure 8). Similarly, *Nicotiana glutinosa RDR6* was found to respond to *Cucumber mosaic virus* (CMV) but not to *Tobacco mosaic virus* (TMV) and *Potato virus Y* (PVY) (Yang *et al.*, 2011). Moreover, different responses of *A. thaliana* RDR6 against *Cauliflower mosaic* virus (CaMV), *Turnip mosaic* virus (TuMV), *Turnip yellow mosaic virus* (TYMV) and *Tobacco mosaic virus* (TMV) were reported (Lilly *et al.*, 2011). However, the interaction between RNA silencing *CpAGO1*, *CpDCL4* in papaya and PRSV infection has never been reported.

Chlorosis in leaf veins was observed at 9-day post- inoculation. The expression levels of *CpAGO1*, *CpRDR6* and *CpDCL4* in infected papaya leaves decreased. RNA virus infection was mainly affected by DCL4 and RDR6 to facilitate extensive virus- induced silencing (Blevins *et al.*, 2006). Interestingly, the *CpDCL4* and the PRSV *coat protein* expressed highly at 14 days post-inoculated (Figure 9) indicating that other virus suppressors besides RNA silencing mechanism may involve. Several viral suppressors of RNA silencing (VSRs) have been identified from almost all plant virus genera. VSRs efficiently inhibit host antiviral responses by interacting with the key components of cellular silencing machinery which included *RDRs*, *AGOs* and *DCL4s* (Burgyan and Havelda, 2011).

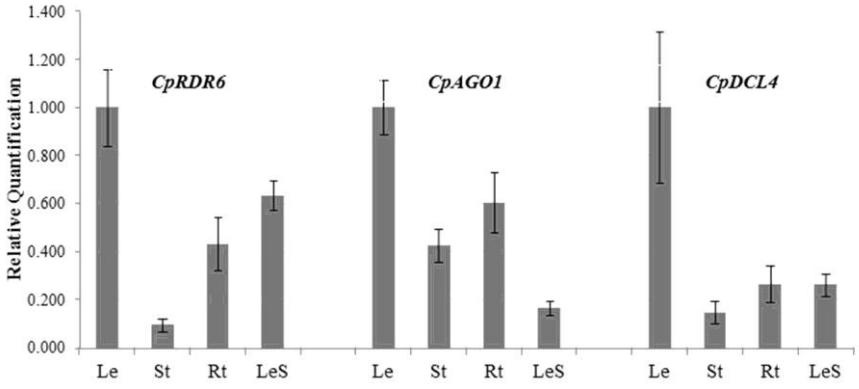

Figure 7 Expression levels of *RDR6*, *AGO1*, and *DCL4* in leaves (Le), stems (St), roots (Rt) and leaf stalks (LeS) of *C. papaya*. The expression patterns were analyzed using qRT-PCR. TATA Binding Protein 1 (TBP1) was used as an endogenous control. The reference levels in leaves were arbitrarily set to 1 and the levels in other tissues were given relative to this. Error bars represent standard deviation of mean value from triplicates for each sample.

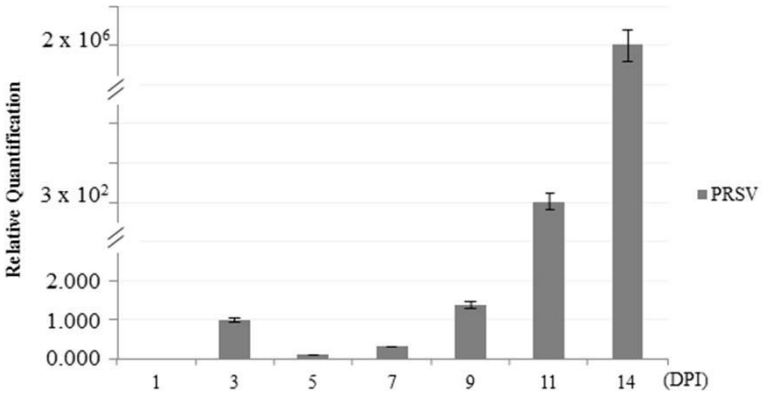

Figure 8 The PRSV coat protein gene expression levels at 0, 1, 3, 5, 7, 9, 11 and 14 days post-inoculation (DPI). The expression patterns were analyzed using qPCR. The transcript level at 3 DPI was set to 1. Error bars represent standard deviation of mean value from triplicates for each sample.

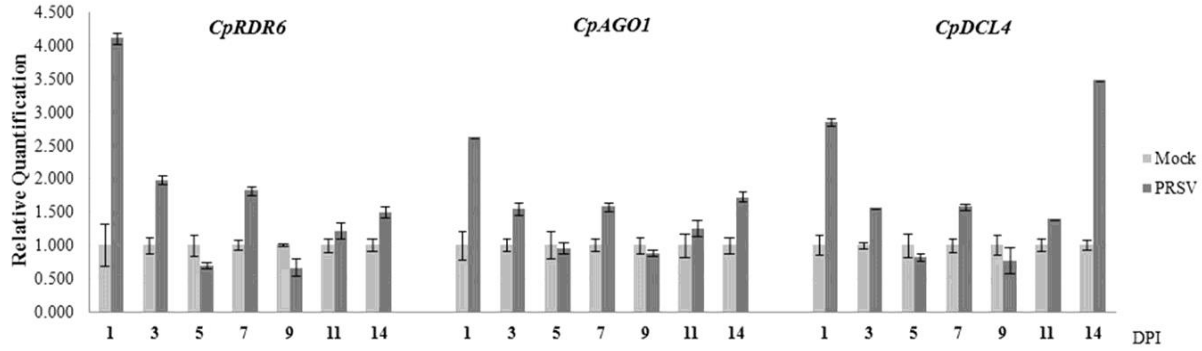

Figure 9 Expression pattern of *RDR6, AGO1,* and *DCL4* under PRSV at 1, 3, 5, 7, 9, 11 and 14 days post-inoculation (DPI). The expression patterns were analyzed using qPCR. The level of transcript in leaves in the mock treatment was set to 1 and level of PRSV treated was given relative to this. Error bars represent standard deviation of mean value from triplicates for each sample.

CONCLUSION

The complete nucleotide sequences of two CDS genes, *CpRDR6* and *CpAGO1*, and partial *CpDCL4* gene were directly cloned from papaya and analyzed with the aid of *in silico* tools. This is a useful, cost-effective tool and shortens the procedure for cloning and analysis of the target genes which were computationally predicted. The genes obtained were successfully deposited in GenBank with accession numbers: *CpRDR6* (KF668595. 1), *CpAGO1* (KY082908) and *CpDCL4* (KY082909). The 3D structures of full CDS CpRDR6 and CpAGO1 were generated and the ligand binding site residues were identified. The RNA silencing genes were expressed in healthy papaya tissues and changed in abundance during PRSV infection.

ACKNOWLEDGEMENTS

This research is supported by the Center of Excellence on Agricultural Biotechnology, Science and Technology Postgraduate Education and Research Development Office, Office of Higher Education Commission, Ministry of Education (AG-BIO/PERDO-CHE)

REFERENCES

Baghban B, Nayeri S (2016) Design, simplified cloning, and *in-silico* analysis of multisite small interfering RNA-targeting cassettes. Mol Biol Res Commun *5*: 31–43.

Blevins T, Rajeswaran R, Shivaprasad PV, Beknazariants D, Si-Ammour A, Park H, Vazquez F, Robertson D, Meins F, Hohn T, Pooggin MM (2006) Four plant dicers mediate viral small RNA biogenesis and DNA virus induced silencing. Nucleic Acids Res 34: 6233–6246.

Bozorov TA, Pandey SP, Dinh ST, Kim S, Heinrich M, Gase K, Baldwi IT (2012) DICER-like proteins and their role in plant-herbivore interactions in *Nicotiana attenuata*. J Integr Plant Biol 54: 189–206.

Burgyan J, Havelda Z (2011) Viral suppressors of RNA silencing. Trends Plant Sci 16: 265–272.

Carvalho FD, Renner SS (2012) A dated phylogeny of the papaya family (*Caricaceae*) reveals the crop's closest relatives and the family's biogeographic history. Mol Phylogenet Evol 65: 46–53.

Chumpookam J, Lin HL, Shiesh CC (2012) Effect of smoke-water on seed germination and seedling growth of papaya (*Carica papaya* cv. Tainung No. 2). HortScience 47: 741–744.

Dalmay T, Hamilton A, Rudd S, Angell S, Baulcombe DC (2000) An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. Cell 101: 543–553.

Eamens A, Wang M, Smith NA, Waterhouse PM (2008) RNA silencing in plants: yesterday, today, and tomorrow. Plant Physiol 147: 456–468.

Ekins S, Mestres J, Testa B (2007) *In silico* pharmacology for drug discovery: methods for virtual ligand screening and profiling. Brit J Pharmacol 152: 9–20.

Elkins KM (2011) An *in silico* dna cloning experiment for the biochemistry laboratory. Biochem Mol Biol Educ 39: 211–215.

Fire AZ (2007) Gene silencing by double-stranded RNA. Cell Death Differ. 14: 1998–2012.

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the expasy server. In: Walker JM (ed) the proteomics protocols handbook. Humana, Totowa, pp 571–607

Gonsalves D, Tripathi S, Carr JB, Suzuki JY (2010) Papaya ringspot virus. The Plant Health Instructor. DOI: 10.1094/PHI-I-2010-1004-01.

Gonsalves Dennis (1998) Control of papaya ringspot virus in papaya: A case study. Annu Rev Phytopathol46: 415–437.

Gronlund M, Olsen A, Johansen EI, Jakobsen I (2010) Protocol: using virus- induced gene silencing to study the arbuscular mycorrhizal symbiosis in *Pisum sativum*. Plant Methods 6: 28.

He L, Guoying Z, Huai-yun Z, Lin L, Jun-ang L (2010) *In silico* cloning and bioinformatic analysis of PEPCK gene in *Fusarium oxysporum*. Afr J Biotechnol 9: 1864–1870.

Janthasri R, Chaiayaboon W (2006) Present status of production and marketing of papaya for unripe consumption in Northeastern Thailand. IJAEB 9: 323–332.

Kalantidis K, Schumacher HT, Alexiadis T, Helm JM (2008) RNA silencing movement in plants. Biol Cell 100: 13–26.

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874.

Lai CJ, Yu Q, Hou S, Skelton RL, Jones MR, Lewis KT, Murray J, Eustice M, Guan P, Agbayani R, *et at*. (2006) Analysis of papaya BAC end sequence reveals first insights into the organization of a fruit tree genome. Mol Genet Genomics 276: 1–12.

Lele Rd (2009) Epigenetics- gene silencing. JAPI 57: 60–66.

Lilly ST, Drummond RM, Pearson MN, MacDiarmid RM (2011) Identification and validation of reference genes for normalization of transcripts from virus- infected *Arabidopsis thaliana*. Mol Plant Microbe Interact 24: 294–304.

Lines R, Persley D, Dale J, Drew R, Bateson M (2002) Genetically engineered immunity to papaya ringspot virus in Australian papaya cultivars. Mol Breed 10: 119–129.

Meng F, Jia H, Ling N, Xue Y, Liu H, Wang K, Yin J, Li Y (2013) Cloning and characterization of two argonaute genes in wheat (*Triticum aestivum* L.). BMC Plant Biol 13: 18.

Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KL, *et al*. (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). Nature 452: 991–996.

Mlotshwa S, Voinnet O, Mette MF, Matzke M, Vaucheret H, Ding SW, Pruss G, Vance VB (2002) RNA silencing and the mobile silencing signal. Plant Cell 14: S289–S301.

Molnar A, Melnyk C, Baulcombe DC (2011) Silencing signals in plants: a long journey for small RNAs. Genome Biol 12: 215.

National Research Council (US) Committee (1987) Computer- Assisted Modeling. Computer- Assisted Modeling: Contributions of Computational Approaches to Elucidating Macromolecular Structure and Function. Washington (DC): National Academies Press (US); Available from: https://www.ncbi.nlm.nih.gov/books/NBK218563/ doi: 10.17226/1136 (August 2017)

Passier R, Doevendans PA (2004) Methods in molecular cardiology: *in silico* cloning. Neth Heart J 12: 400–403.

Pedro CJ, Costa FR, Pereira T S, Neto MF, Pereira MG (2009) Karyotype determination in three *Caricaceae* species emphasizing the cultivated form (*C. papaya* L.). Caryologia 62: 10–15.

Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS (2004) *SGS3* and *SGS2/SDE1/RDR6* are required for juvenile development and the production of *trans*-acting siRNAs in *Arabidopsis*. Genes Dev 18: 2368–2379.

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

Plasterk RH, Ketting RF (2000) The silence of the genes. Curr Opin Genet Dev 10: 562–567.

Qu F, Ye X, Morris TJ. (2008) Arabidopsis DRB4, AGO1, AGO7, and RDR6 participate in a DCL4 initiated antiviral RNA silencing pathway negatively regulated by DCL1. Proc Natl Acad Sci USA. 105: 14732–14737.

Rao X, Lai D, Huang X (2013) A new method for quantitative real- time polymerase chain reaction data analysis. J Comput Biol 20: 703–711.

Ratcliff F, Harrison BD, Baulcombe DC (1997) A similarity between viral defense and gene silencing in plants. Science 276: 1558–1560.

Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5: 725–738.

Scofield SR, Nelson RS (2009) Resources for virus-induced gene silencing in the grasses. Plant Physiol 149: 152–157.

Shao F, Lu S (2014) Identification, molecular cloning and expression analysis of five RNA- dependent RNA polymerase genes in *Salvia miltiorrhiza*. PLoS One 9: e95117.

Sharma N, Singh SK, Lal S, Singh NK (2016) Genome sequence information in fruit crops: current status. Transcriptomics 4: 130.

Shimura H, Pantaleo V (2011) Viral induction and suppression of RNA silencing in plants. Biochim Biophys Acta, doi:10.1016/j.bbagrm.2011.04.005

Singh PK, Singh D, Vikas, Garg GK (2006) RNA interference and gene silencing. Indian J Crop Sci 1: 13–20.

Siriwan W, Takaya N, Roytrakul S, Chowpongpang S (2014) Study of interaction between papaya ringspot virus hc-pro and papaya (*Carica papaya*) proteins. J Gen Plant Pathol 80: 264–271.

Sliva K, Schnierle BS (2010) Selective gene silencing by viral delivery of short hairpin RNA. Virol J 7: 248.

Teixeira AJ, Teixeira S, Zinia R, Duong TN, Dharini S, Abed G, Manoel TS, Paula FT (2007) Papaya (*Carica papaya* L.) biology and biotechnology. Tree For Sci Biotech 48–73.

Voinnet O (2001) RNA silencing as a plant immune system against viruses. Trends Genet 17: 449–459.

Wang M, Li S, Yang H, Gao Z, Wu C, Guo X (2012) Characterization and functional analysis of *GhRDR6*, a novel *RDR6* gene from cotton (*Gossypium hirsutum* L.). Biosci Rep 32: 139.

Willmann MR, Endres MW, Cook RT, Gregory BD (2011) The Functions of RNA-dependent RNA polymerases in *Arabidopsis*. Arabidopsis Book 9: e0146.

Xie Z, Fan B, Chen C, Chen Z (2001) An important role of an inducible RNA-dependent RNA polymerase in plant antiviral defense. PNAS 98: 6516–6521.

Yang H, Wang M, Gao Z, Zhu C, Guo X (2011) Isolation of a novel RNA- dependent RNA polymerase 6 from *Nicotiana glutinosa*, NgRDR6, and analysis of its response to biotic and abiotic stresses. Mol Biol Rep 38: 929–937.

Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015) The I-TASSER Suite: protein structure and function prediction. Nat Methods 12: 7–8.

Zhang Z, Henzel W J (2004) Signal peptide prediction based on analysis of experimentally verified cleavage sites. Protein Sci 13: 2819–2824.

Zhu X, Li X, Chen W, Chen J, Lu W, Chen L, Fu D (2012) Evaluation of new reference genes in papaya for accurate transcript normalization under different experimental conditions. PLoS One 7(8): e44405.