Title page

Bioinformatic analysis based genome-wide identification, characterization, diversification and regulatory transcription components of RNA silencing machinery genes in wheat (*Triticum aestivum* L.)

Zobaer Akond^{1, 2, 5}, Hafizur Rahman^{3, 6}, Md. Asif Ahsan¹, Md. Parvez Mosharaf¹, Munirul Alam⁴, Md. Nurul Haque Mollah^{1*}

¹Bioinformatics Lab, Department of Statistics, University of Rajshahi, Rajshahi, Bangladesh

²Institute of Environmental Science, University of Rajshahi, Rajshahi, Bangladesh

³Rajshahi Institute of Biosciences, University of Rajshahi, Rajshahi, Bangladesh

⁴Molecular Ecology and Metagenomic Laboratory, Infectious Diseases Division, International Centre for Diarrheal Disease Research, Bangladesh (icddr,b)

⁵Agricultural Statistics and ICT Division, Bangladesh Agricultural Research Institute (BARI), Gazipur, Bangladesh

*Corresponding author

Email: mollah.stat.bio@ru.ac.bd

Abstract

Dicer-Like (DCL), Argonaute (AGO), and RNA-dependent RNA polymerase (RDR) gene families are known as RNA silencing machinery genes or RNAi genes. They have important activities at post-transcriptional and chromatin modification levels. They regulate gene expression relating to different stresses, growth, and development in eukaryotes. A complete cycle of gene silencing is occurred by the collaboration of these three families. However, these gene families are not yet rigorously studied in the economically important wheat genome. Our bioinformatic analysis based genome-wide identification, characterization, diversification and regulatory components of these gene families identified 7 TaDCL, 39 TaAGO and 16 TaRDR genes from wheat genome against RNAi genes of *Arabidopsis thaliana*. Phylogenetic analysis of wheat genome with *Arabidopsis* and rice RNAi genes showed that TaDCL, TaAGO and TaRDR proteins are clustered into four, eight and four subgroups respectively. Domain, motif and exon-intron structure analyses showed that the TaDCL, TaAGO and TaRDR proteins conserve identical characteristics within groups while retain diverse differences between groups. GO annotations implied that a number of biological and molecular pathways are linked to RNAi mechanism in wheat. Gene networking between transcription factors and RNAi proteins indicates that ERF is the leading family linked to maximum RNAi genes followed by MIKC-MADS, C2H2, BBR-BPC, MYB, and Dof. *Cis*-regulatory elements associated to RNAi genes are predicted to act as regulatory components against various environmental conditions. Expressed sequence tag analysis showed that larger numbers of RNAi genes are expressed in different tissues and organs predicted to play roles for healthy plants and grains. Expression analysis of 7 TaDCL genes using qRT-PCR showed that only TaDCL3a and TaDCL3b had root specific significant expression (p -value<0.05) with no expression in leaf validated EST results. Besides, TaDCL3b and TaDCL4 significantly prompted in drought condition indicating their potential role in drought stress tolerance. Overall results would however help researchers for in-depth biological investigation of these RNAi genes in wheat crop improvement.

Keywords: Wheat genome, RNAi genes, Transcription factor, *Cis*-regulatory elements, Expressed sequence tag, Gene ontology, *Arabidopsis* RNAi genes, Bioinformatic analysis

1. Introduction

Recently gene silencing technique has attracted many molecular biological scientists to develop transgene system in eukaryotic cell to solve complex biological problems and to make crop improvement by implementing the RNAi mechanisms. Plants and agricultural crops by nature establish some specific molecular mechanisms that help survive in diverse conditions in their life span. RNA silencing often known as RNA interference (RNAi) technique is one such mechanism, which is well-preserved in most multicellular eukaryotic groups and a unifying fact to define a sequence-specific RNA degradation process that controls sequence specific regulation of gene expression[1–3]. This technique was originally learned in plants and occurs broadly in eukaryotic organisms[4,5]. Eukaryotic organisms generate copious types of small RNAs[6]. It has been explored in previous studies that gene silencing or RNAi technique is triggered and maintained by double-stranded RNA(dsRNA) that provides increase to small RNA molecules[1,2,7–11]. There are two types of small RNA molecules nearly of size 20-30 nucleotides long ribonucleotides termed as microRNA (miRNA) and short interfering RNA (siRNA)[2,8,10,11]. microRNAs (miRNAs) are derived from self-complementary hairpin structures, whereas small-interfering RNAs (siRNAs) are resulting from double-stranded RNA (dsRNA) or hairpin precursors[12,13]. These minute size RNA particles play roles in controlling various biological and genetic processes in terms of growth and development, creation of heterochromatic, retaining genome stability by means of regulation of gene expression and suppression of transposon movement, anti-pathogen and fungal defense[1,3,10,11,14–16]. These small RNAs are mainly linked to both post-transcriptional gene silencing and triggering chromatic modifications[10,17]. In plants and crops, the production and function of these miRNAs and siRNAs are largely rely on three main protein families known as Dicer-like(DCL), Argonaute(AGO) and RNA-dependent RNA polymerase(RDR)[8,10]. A complete cycle of RNA silencing takes three steps: initiation, maintenance and signal amplification. The beginning of gene silencing requires the production of double-stranded RNA (dsRNA). dsRNAs that are created from plant-encoded RNA-dependent RNA polymerases(RDRs) on a variety of RNA templates[3]. The other processes of creation of dsRNA include bidirectional transcription of DNA, self-complementary RNA fold backs [7]. The complementary dsRNA are then processed by the act of RNaseIII-type enzymes called Dicer (DCR) in animals or Dicer-like (DCL) in plants which then alter these dsRNA into tiny RNAs of 19-31 nucleotides(nt) in size siRNA/miRNA[3,7,10]. Afterward, one strand of these siRNAs is attached to AGO protein containing complexes termed as RISC (RNA-induced silencing complex). RISC possesses endonuclease ability to operate cleavage activity of target mRNAs or DNAs

that are complementary/homologous to siRNA/miRNA by means of AGO's RNaseH-type enzyme property for RNA degradation, translational inhibition, or heterochromatin formation [1,3,7,8,10,18]. siRNAs are also predicted for transcriptional gene silencing by implementing RNA-directed DNA methylation or chromatic remodeling to regulatory sequences of target genes[7,10]. At the signal augmentation step, RDR enzymes are likely for synthesis of dsRNAs from ssRNA templates to start a new cycle of RNA silencing [8,19]

DCL, AGO and RDR protein families are however considered as the chief components of gene silencing or RNAi cycle. DCL proteins are known as the important protein families among the other RNA silencing machinery genes for small RNA biogenesis routes that transform the long dsRNAs into mature miRNAs and siRNAs[10,11]. DCL families are identified and characterized to possess six conserved domains, viz., dead box, helicase, dicer dimerization, PAZ (Piwi Argonaut and Zwillie), RNase III, and dsRNA binding domain (dsRBD)[7,8,10,11]. Although one or two domains are assumed to absent in the lower type eukaryotic organisms[11,20]. The helicase domain is termed to identify the single strand terminal loop structure of primary miRNA(pre-miRNA) and limits the proportion of substrate cleavage [11,21]. The dicer dimerization domain is engaged in target selection of small RNA handling [22]; the PAZ and RNaseIII domains contribute in dsRNA substrate cleavage[11,20]. Surprisingly, the distance between PAZ and the cleavage site in the RNaseIII domain decides the size of matured small RNAs[11,23]. DCL however contains a minor set of gene family in upper plants, insects, protozoa, and some fungi, while only one Dicer protein is present in vertebrates, nematodes, *Schizosaccharomyces pombe*, and green alga *Chlamydomonas reinhardtii*[8,20].

AGOs are also known as the special kind of protein families in RNA silencing mechanism located at multi-protein type RISC that direct to cleave the target mRNAs, which are complementary sequence of siRNAs. AGO proteins particularly in larger nearly 90-100kDa characterized by the presence of four functional domains or amino acid motifs, viz., ArgoN/Argo-L, C-terminal PIWI(P-element induced wimpy testis),PAZ and MID[6,24,25]. A chief PAZ domain possesses a particular binding pocket that binds siRNAs over the 3' end of the target mRNA[10,24] whereas a PIWI binds siRNAs through the 5' end of the target mRNA which holds the complete homology characteristics like RNaseH to cleave target mRNAs[1,10,26]. MID domains are placed between the PAZ and the PIWI domains. A very simple pocket characteristic of such MID domain particularly binds the 5' phosphate of the siRNAs and then anchors the siRNAs onto AGO proteins [10,26]. In eukaryotic species, minimum three subclasses of AGO proteins have been recognized[7]. These consists of AGO subfamily which is exist in plants, animals and

yeasts; the Piwi and worm-specific AGO(WAGO) subgroups are observed in animals and *C. elegans* respectively[7,27–29]. However, all members of the subgroup of AGO and Piwi conserve the particular DDH catalytic residues but absent in WAGO. The expression of Piwi subgroup proteins is limited to human germ cell and in rat and some mammals. Although this cluster of AGO proteins have not been identified in any plant species but in recent times a novel type of Argonaute, OsMEL1 has been identified in rice which is linked to male meiosis [7,30].

The third type of RNA silencing machinery gene is the RDR protein. This kind of protein family is exist and highly essential for RNAi mechanism in fungi, nematodes and plants, but have not been identified in insects or vertebrates[31]. RDR proteins were separated first from tomato[32]. RDR proteins are critically important for start and increase of silencing process and these protein families contain a well-preserved sequence motif like the catalytic β' subunit of DNA dependent RNA polymerases[7,33]. This gene family, however, possesses a RNA-dependent RNA polymerase (RdRP) domain, help form the dsRNA from single-stranded RNAs (ssRNAs) to start a new cycle of RNA silencing [7,34].

Recent study suggests that there are numerous copies of DCL, AGO and RDR genes are present in plants and animals[7,10]. Different study on a dozen of crops plant has been made to learn about the role of these RNAi gene families. All members of each of these gene families take part in diverse roles in RNA silencing pathway. A model plant in plant biology, for instance, *Arabidopsis thaliana* possesses 4 DCL (AtDCL1-AtDCL4) that definitely generate diverse kinds and sizes of small RNAs [18]. To activate various important biological functions in eukaryotic cell, different small RNAs produced in cell with diverse functions but they all are the associate members of DCL, AGO and RDR gene families. However, in *Arabidopsis thaliana* genome, 4 AtDCLs, 10 AtAGOs and 6 AtRDRs genes were identified [6,7,10]. 4 NbDCL, 9NbAGO and 3 NbrRDR RNAi genes were identified in tobacco(*Nicotiana benthamin*)[35]. Rice (*Oryza sativa*), a crop of monocot group, possesses 8 OsDCLs, 19 OsAGOs and 5 OsRDRs genes, in which OsAGO2 gene exhibited particular up-regulation in response to salt and drought [7,32]. Also, in tomato (*Solanum lycopersicum*), 7 SlDCLs, 15 SlAGOs and 6 SlRDRs genes were identified [36]. Five ZmDCLs, 18 ZmAGOs and 5 ZmRDRs genes were recognized in maize(*Zea mays*) genome [10]. There are 4 VvDCLs, 13 VvAGOs and 5 VvRDRs genes were detected in grapevine (*Vitis vinifera*)[37]. Likewise, 5, 7, and 8 CsDCLs, CsAGOs, and CsRDRs, respectively have been identified in cucumber (*Cucumis sativus*) [38]. On the other hand, the genome of allopolyploid species of *Brassica napus* contain 8 BnDCLs, 27 BnAGOs, and 16 BnRDRs [8,39]. Overall, 4 CaDCLs, 12 CaAGOs and 6 CaRDRs genes have been identified in pepper (*Capsicum*

Annuum) [1]. In coffee (*Coffea canephora*) 9CcDCLs, 11CcAGOs and 8 CcRDRs were identified for genome-wide study by means of RNA guided pathways in 2017[2]. Foxtail millet (*Setaria italica*) possesses 8 SiDCL, 19 SiAGO and 11 SiRDR genes. A very recent study suggests that sweet orange (*Citrus Sinensis*) genome possesses 5CsDCLs 13 CsAGOs, and 7 CsRDR RNAi genes [40]. However, successive investigation on RNAi genes in various important crop and fruit plants showed considerable divergence and have indispensable role in different genomic functions. In *A. thaliana*, AtDCL1 principally play role for miRNA biogenesis while AtDCL2, AtDCL3, and AtDCL4 mediate siRNA processing[10,11]. Besides, AtDCL3 and AtAGO4 are essential for RNA-directed DNA methylation of the FWA transgene, which is associated to histone H3 lysine 9 (H3K9) methylation[10,11]. Though AtDCL2 produces siRNAs responsible to create defensive mechanism against viral infection and by nature the creation of siRNAs are tend to happen from *cis*-acting antisense transcript but AtDCL4 regulate the vegetative stage transformation by the production of siRNAs from *trans*-acting transcript[10,41,42]. Surprisingly, AtDCL3 selects short dsRNAs but AtDCL4 cleaves long dsRNA substrate s[11]. Many other functions of DCL genes in plants such as AtDCL1 and AtDCL3 proteins contribute in promoting flowering[43].

On the other hand, AGO gene families act as the chief RNA-mediated gene silencing techniques and reported that they play noteworthy performance in growth and development in plants[24,44,45]. AtAGO1 is associated with the transgene silencing pathways[46] and AtAGO4 with the epigenetic silencing [47]. AtAGO7 and AtAGO10 however effect the growth [48] and meristem maintenance [49]. Other AtAGOs however have some important characteristics in gene silencing pathways. Earlier investigations also suggest that the RDR genes are biologically prompt in RNAi mechanism such as co-suppression, protect pathogen invasion, chromatin modification, and post-transcriptional gene silencing activities in plants, viz., *Arabidopsis*, maize [50–54].

Wheat (*Triticum aestivum*), a global common staple food as well as the second most-produced cereal crops after maize in the world (<http://www.fao.org/worldfoodsituation/csdb/en/>). It is also the high vital source of carbohydrates and is the leading source of vegetal protein in human food. There is little systemic thorough study has been made so far regarding RNA silencing machinery genes for this crop. The current study was carried out to obtain comprehensive bioinformatic understanding in terms of *in-silico* identification, characterization, diversity survey, regulatory components, and expression sequence tag(EST) analysis of all members of DCL, AGO and RDR gene families in wheat. Finally, wet-lab experimental expression analysis was investigated for the identified 7 wheat DCL

proteins in the two wheat plant organs (root and leaf) and in drought condition as well. Outcomes obtained from this work will offer fundamental source of genomic information for these RNAi genes and will certainly help plant molecular researchers for insights into the possible roles of these genes in plant growth, development, and performances against various biotic and abiotic stressors. The systemic schemes of this whole study in this article however have been demonstrated briefly in Fig.1.

Fig 1. The flow diagram of the complete work in this study

2. Materials and methods

2.1. Data sources of RNAi genes

Protein sequences of *Arabidopsis* and rice (*Oryza sativa*) DCL, AGO and RDR were obtained from The Arabidopsis Information Resource (TAIR) (<http://www.arabidopsis.org>) and Rice Genome Annotation Project (RGAP) (<http://rice.plantbiology.msu.edu/>) respectively following the profile Hidden Markov Model (profile HMM). Then these sequences were used as query sequence to search using Basic Local Alignment Search Tool (BLAST-P) program (p-value=0.001) against *T. aestivum* genome from the plant genome Phytozome database (<http://phytozome.jgi.doe.gov/pz/portal.html>). Paralogous protein sequences from the wheat (*Triticum aestivum*) genome were collected by considering the threshold value (≥ 50) and significant E-values. Only the primary transcripts were selected to avoid the redundancy of the sequences in the analysis. All retrieved sequences were searched for conserved functional domains using the Pfam (<http://pfam.sanger.ac.uk/>), web-based Simple Modular Architecture Research Tool (SMART) (<http://smart.embl-heidelberg.de/>) and the National Center for Biotechnology Information Conserved Domain Database (NCBI-CDD: <http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). Some basic information of these genes such as accession number, genomic location, gene length, and encoded protein length were downloaded from Phytozome wheat genome database. All newly identified genes from wheat genome were designated as TaDCL, TaAGO and TaRDR following taxonomy based on phylogenetic relationship of the same family members of the *Arabidopsis thaliana* genes as termed earlier investigations. Prediction of the molecular weight of each gene member of each DCL, AGO and RDR group was predicted from web-based tool ExPASyComputePI/Mwtool (<http://au.expasy.org/tools/pito ol.html>).

2.2 Bioinformatic analysis approaches

Several bioinformatic analysis approaches including sequence alignment and phylogenetic tree construction, prediction of functional domain and motif structure of the proteins, exon-intron structure of the RNAi genes, gene ontology (GO) analysis, prediction of subcellular location, regulatory network among the gene transcription factors and wheat RNAi genes, *cis*-regulatory element(CRE) analysis, express sequence tag (EST) analysis were carried out for comprehensive genome-wide identification, characterization, diversification and regulatory transcription components of wheat RNA silencing machinery genes. These approaches are described in the following sub-sections 2.2.1-2.2.8.

2.2.1 Sequence Alignment and Phylogenetic Tree Analysis

Protein sequences of DCL, AGO and RDR genes belong to respective *Arabidopsis thaliana*, rice and wheat were used for multiple sequence alignment using ClustalW method [55]. After that, to study the evolutionary relationship, three phylogenetic trees were constructed of the corresponding 7 TaDCLs, 39TaAGOs and 16 TaRDRs using MEGA7: Molecular Evolutionary Genetics Analysis software[56] with Neighbor-Joining (NJ) method. Phylogeny test was performed by following bootstrap method with 1000 replications to measure statistical support for nodes[57] and the evolutionary distances were computed using the p-distance substitution method with complete deletion of the gaps/missing data[56].

2.2.2 Conserved functional domain and motif structure analysis

Conserved functional domains for each of the TaDCL, TaAGO and TaRDR protein family were predicted from the three well-established genomic online databases, viz., Pfam, SMART and NCB-CDD by using their corresponding protein sequences to search similar conserved functional domains contained by RNAi genes of *Arabidopsis Thaliana* AtDCL, AtAGO and AtRDR. Higher numbers of functional domain possessing TaDCL, TaAGO and TaRDR genes corresponding to that of AtDCLs, AtAGOs and AtRDRs genes were selected. Six functional domains, viz., DEAD/ResIII, Helicase-C, Dicer-Dimer, PAZ, RNase III and DSRM were taken into consideration to select the potential DCL group genes(TaDCLs) in *Triticum aestivum*. The AGO gene families in wheat that contain Argo-N/Argo-L, PAZ, MID and PIWI important functional domains are termed as AGO class genes in wheat. TaRDR gene families were also listed for the presence of RdRP and RRM functional domains for protein structure.

In order to identify the most potential conserved metal-chelating catalytic triad regions or residues in the PIWI domain, i.e., aspartate, aspartate and histidine (DDH)[7] as well as histidine at 798 positions (H798), TaAGO protein families were considered. In addition, the two conserved motifs, viz, DLDGD and DFDGD in the RdRp domain of TaRDR protein groups were also detected. The alignment profiles for this identification were performed using ClustalW program in MEGA and GENEDOC software tool. All paralog AtAGO and AtRDR protein families were considered during alignment (Fig.2). Moreover, the conserved structural motif differences among DCL, AGO and RDR gene groups in *T. aestivum* were projected using a web-based tool called Multiple Expectation Maximization for Motif Elicitation (MEME-Suite)[58]. MEME analysis was carried out by specifying the parameters: (i) optimum motif width as ≥ 6 and ≤ 50 ; (ii) maximum 20 motifs.

2.2.3. Exon-intron structure analysis of the TaDCL, TaAGO and TaRDR gene families

The organization of the exon-intron structure of 7 TaDCLs, 39 TaAGOs and 16 TaRDRs genes were measured using the online-based Gene Structure Display Server (GSDS 2.0 : <http://gsds.cbi.pku.edu.cn/index.php>) by taking their respective complete coding sequences (CDS) and DNA/genomic sequences downloaded from Phytozome database. *Arabidopsis thaliana* 4 AtDCLs, 10AtAGOs, and 6AtRDRs genes were also included to observe the similarity study.

2.2.4. Gene ontology and sub-cellular localization analysis

Gene ontology (GO) enrichment analysis was performed using the web-based PlantTFDB v4.0[59] and QuickGO (<https://www.ebi.ac.uk/QuickGO>) tool in order to investigate the particular involvement of the RNAi genes of *T. aestivum* in terms of biological function (BP), molecular function(MF) and in certain cellular component(CC). Fisher's exact test was considered to determine respective p-values and Benjamin-Hochberg's adjustment correction. Enrichment outcomes with adjusted $p < 0.05$ were considered as statistically important. Furthermore, a web-based tool called PSI (Plant Subcellular localization Investigative Predictor) [60] for predicting the subcellular location of the 62 RNA silencing genes of *Triticum aestivum* and open source software R-3.5.2 [61] were used. A protein was considered locating in the certain cellular location if $p < 0.00001$ in the resulting output of the tool.

2.2.5. Identification and characterization of regulatory TFs and RNAi genes in *T.aestivum*

To identify and study the associated regulatory transcription factor (TF) families and their degree of relationship with the identified RNAi genes of *T. aestivum* were identified with the help of plantTFDB (<http://planttfdb.cbi.pku.edu.cn/>).

2.2.6. Construction of gene regulatory network between TFs and *T.aestivum* RNAi genes

To explore the interacting regulatory association between TFs and RNAi genes, all identified TF families and predicted *T.aestivum* RNAi genes were used to construct a whole regulatory gene network and sub-networks to find out the leading TF components and hub proteins by means of interaction network using Cytoscape 3.7.1[62].

2.2.7. Promoter *cis*-acting element prediction of the RNAi genes in *T.aestivum*

Cis-acting elements are highly conserved and commonly related to regulate the transcription of a particular gene or gene families in conjunction with the respective TFs. For in-depth analysis of the *T.aestivum* RNAi genes, the *cis*-acting regulatory components in the promoter region were collected and analyzed using web-based prediction bioinformatic tool PlantCARE(<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>). The extracted bundle of *cis*-acting components were classified into five different groups based on their involvement in various actions such as stress responsive, light responsive, hormone responsive and other which have unknown functions.

2.2.8. Expressed sequence tag (EST) analysis

In-silico EST investigation was performed for the identified wheat RNAi genes to predict their expression status in various important wheat plant tissues and organs using the web-based database PlantGDB (<http://www.plantgdb.org/cgi-bin/blast/PlantGDB/>). The DNA sequences of each TaDCL, TaAGO and TaRDR gene group were used to extract the EST information. The default parameter with e-value=1e-04 was set for blastn search. Finally, a heatmap was created to characterize the particular RNAi gene expression into various tissues and organs in *T.aestivum*.

2.3 RNA isolation and expression analysis of 7 TaDCL genes in *T.*

aestivum

Complete RNA was taken out using Trizol reagent (TAKARA, Japan) by following the manufacturer's guidelines. RNA was treated with DNaseI (TAKARA, Japan) and reverse-transcribed into cDNA using the PrimeScript RT reagent kit (TAKARA, Japan). The collected cDNAs were used for gene expression studies with quantitative real time PCR (qRT-PCR). The qRT-PCR was accomplished in StepOne Real-Time PCR System (Applied Biosystems, USA) using SYBER Premix Ex Taq reagents (TAKARA, Japan) following the program: 95°C for 30s, 95°C for 5s and 60°C for 45s for 40 cycles. To normalize the sample variance, *T.aestivum* 18s gene was used as internal control. Relative gene expression values were calculated using the $2^{-\Delta\Delta Ct}$ method[63]. The primers used for gene expression analyses are listed in S1 Table. For the statistical analysis of the gene expression data, ANOVA was performed with SPSS software (Version19.0, IBM, USA). An important difference between mean values was assessed following DMRT ($p < 0.05$) approach.

3. Results and discussion

3.1. Identification and characterization of TaDCL, TaAGO and TaRDR genes

All downloaded protein sequences were undergone several analysis procedures for selecting the best possible RNAi genes in *T.aestivum* (Fig.1). Then in total 62 RNAi gene set (7 TaDCL, 39 TaAGO and 16 TaRDR) were finally selected in wheat (*T. aestivum*) for subsequent bioinformatic analyses. By considering six functional domains, viz. DEAD/ResIII, Helicase-C, Dicer-dimer, PAZ, RNase III (Ribonuclease-3), and DSRM, 7 DCL loci were isolated and recognized in *T.aestivum* as TaDCLs using profile HMM with expect value threshold=0.01[64]. Conserved functional domains search of these 7 TaDCLs from the two databases(Pfam and NCBI-CDD) and SMART analysis reveals that almost all TaDCL protein families contain the said six domains (DEAD/ResIII, Helicase-C, Dicer-dimer, PAZ, RNase III and DSRM) which are termed as the full characteristics of plant DCL protein structure from the DCL family (class 3 RNase III family)[65]. Though all TaDCLs contain DEAD/ResIII but this helicase type DEXH-box domain of endoribonuclease dicer was absent in TaDCL3b and TaDCL4. This domain typically contains the ATP-binding region[66]. Domain identification using SMART showed that the Dicer-dimer domain is absent but Pfam and NCBI-CDD database analysis showed the presence of this domain. This domain is found in members of

the Dicer protein family, which functions in RNAi pathway and has weak dsRNA-binding action[64]. It facilitates heterodimerization of Dicer proteins with their corresponding protein partners[67]. Pfam analysis showed another domain of unknown function (DUF2838) in TaDCL1a. It is observed that all TaDCL protein families possess second DSRM domain shown by Pfam, SMART and NCBI-CDD databases that is entirely missed in non-plant DCLs, as described earlier [41]. The length of the TaDCL protein families however vary from 14743bp for TaDCL4 to 8347bp for TaDCL3d with their corresponding coding potential of 1392 and 1586 amino acids respectively (Table 1).

Through HMM analysis for both conserved functional domains, viz., PAZ and PIWI using the downloaded protein sequences, 39 AGO RNAi protein families were identified finally for wheat (*T. aestivum*) genome. Conserved functional domain analysis using Pfam, SMART and NCBI-CDD showed that all identified putative TaAGO protein families are tend to contain an N-terminus PAZ domain and a C-terminus PIWI super family domain that are known as the main characteristics of plant AGO RNAi genes. It is also shown that other than these two important domains, wheat TaAGO proteins retain some other functional domains like *A. thaliana*, viz., ArgoN, ArgoL1, DUF1785, ArgoL2, ArgoMid. The length of the TaAGO protein families however varied from 12443bp for TaAGO1g to 3234bp for TaAGO3 with their corresponding coding potential of 1085 and 849 amino acids respectively (Table 1).

Besides, earlier investigations suggest that the PIWI domain presents complete homology to RNase H binds the siRNA 5' end to the target RNA[26] and take part to cut target mRNAs that display sequence complementary to siRNA or miRNAs[68,69]. It is worth mentioning here that the three catalytic regions or residues in the PIWI domain such as aspartate(D), aspartate(D) and histidine(H) called DDH [7] are linked to the previous action and this triad was originally explored in *A.thaliana* AGO1, and a well-preserved histidine at position 798(H798) was also identified to be crucial for AGO1 for *in vitro* endonuclease function[69]. PIWI domains of all TaAGOs were aligned along with the paralogous 10 AtAGOs through multiple sequence alignment technique from their corresponding amino acid sequences using ClustalW program implemented in MEGA 7.0 (Fig.2). The PIWI domain sequence alignment showed that total 15 members namely 7TaAGO1s (TaAGO1a, TaAGO1b, TaAGO1c, TaAGO1d, TaAGO1i, TaAGO1j, TaAGO1k), 4TaAGO5s (TaAGO5c, TaAGO5d, TaAGO5f and TaAGO5g), 2TaAGO7s (TaAGO7a and TaAGO7b), 2TaAGO10s (TaAGO10c and TaAGO10d) possessed the four important active residues (DDH/H) like *Arabidopsis* AGO1(AtAGO1), implying that they might have endonuclease activity[68](Fig.

2). The rest of the 24 TaAGOs showed at least one dissimilarity among these catalytic traid residues. The D760 residue as AtAGO1 in TaAGO8 was changed by Tyrosine (Y) while for others remained same as AtAGO1 (Table 2

Table 1. Basic genomic information about TaDCL, TaAGO and TaRDR protein families in wheat (*T.aestivum*).

Gene No.	Gene Name	Accession Number	Genomic Location	Gene Length (bp)	No. of introns	Protein			Type
						Molecular Weight (D.a)	Protein length (a.a.)	PI	
DCL									
1	TaDCL1a	Traes_4BL_B3A1B8342.2	ta_iwgcsc_4bl_v1_6994694:2688..13103	10416	9	195505.38	1779	6.43	DCL1
2	TaDCL1b	Traes_5AL_72A7552B9.2	ta_iwgcsc_5al_v1_2752914:2855..12171	9317	9	187281.30	1674	6.19	DCL1
3	TaDCL3a	Traes_1AL_E7144546E.1	ta_iwgcsc_1al_v2_3918617:7862..20042	12181	18	184016.48	1627	6.39	DCL3
4	TaDCL3b	Traes_1DL_C646B6990.1	ta_iwgcsc_1dl_v1_2224660:33..11929	11897	15	169269.68	1494	6.46	DCL3
5	TaDCL3c	Traes_3AL_562D6614F.1	ta_iwgcsc_3al_v1_4401047:3091..12058	8968	16	177350.72	1586	6.04	DCL3
6	TaDCL3d	Traes_3DL_2DC78B18A.1	ta_iwgcsc_3dl_v1_6937552:192..8538	8347	16	177414.90	1586	6.06	DCL3
7	TaDCL4	Traes_2DL_E96DCDCB4.2	ta_iwgcsc_2dl_v1_9883052:5868..20610	14743	15	157968.69	1392	5.92	DCL4
AGO									
1	TaAGO1a	Traes_2AL_2512A7F91.1	ta_iwgcsc_2al_v1_6428016:16577..22920	6344	16	123940.53	1117	9.48	AGO1
2.	TaAGO1b	Traes_2BL_93099ACF4.1	ta_iwgcsc_6al_v1_5742333:938..12517	11580	11	72763.47	645	9.21	AGO1
3.	TaAGO1c	Traes_6AL_616161AAB.1	ta_iwgcsc_6al_v1_5742333:938..12517	11580	18	108532.11	978	9.48	AGO1
4.	TaAGO1d	Traes_6DL_58620B158.2	ta_iwgcsc_6dl_v1_3210843:3..6281	6279	18	101586.27	910	9.35	AGO1
5.	TaAGO1e	Traes_7DL_C255A109C.1	ta_iwgcsc_7dl_v1_3396214:4962..9930	4969	14	77484.04	721	9.22	AGO1
6.	TaAGO1f	Traes_6BL_9CFA54D4A.1	ta_iwgcsc_6bl_v1_4398549:8497..16669	8173	16	119364.74	1068	9.56	AGO1
7.	TaAGO1g	Traes_6AL_317133B3F.2	ta_iwgcsc_6al_v1_5770608:7589..20031	12443	17	120920.18	1085	9.60	AGO1
8.	TaAGO1h	Traes_6DL_804FB7F75.1	ta_iwgcsc_6dl_v1_3318900:4215..12555	8341	16	119065.35	1067	9.58	AGO1
9.	TaAGO1i	Traes_7AS_56569A5AC.2	ta_iwgcsc_7as_v1_4062808:1122..8394	7273	16	133941.63	1210	9.41	AGO1
10.	TaAGO1j	Traes_7DS_4D01B6175.1	ta_iwgcsc_7ds_v1_3925271:1281..7139	5859	16	97926.28	868	9.31	AGO1
11.	TaAGO1k	Traes_4AL_A118C6C84.2	ta_iwgcsc_4al_v2_7127490:1302..7644	6343	16	131032.87	1189	9.51	AGO1
12.	TaAGO2a	Traes_2AL_DFE4C65F6.2	ta_iwgcsc_2al_v1_6435863:9..4673	4665	3	104716.17	952	9.23	AGO2
13.	TaAGO2b	Traes_2BL_7713B3533.2	ta_iwgcsc_2bl_v1_8046272:2..4842	4841	2	94598.13	845	9.17	AGO2
14.	TaAGO3	Traes_2DL_A77212060.2	ta_iwgcsc_2dl_v1_9902460:2552..5785	3234	3	95781.21	849	8.54	AGO3
15.	TaAGO4a	Traes_3AS_8EE711E2C.2	ta_iwgcsc_3as_v1_3376626:7020..14138	7119	14	111230.93	999	8.88	AGO4
16.	TaAGO4b	Traes_3DS_57EA31670.1	ta_iwgcsc_3ds_v1_2601767:2..5053	5052	12	88926.36	799	9.22	AGO4
17.	TaAGO4c	Traes_3B_F4E466F78.1	ta_iwgcsc_3b_v1_10588096:6226..11781	5556	15	99622.27	918	9.21	AGO4
18.	TaAGO5a	Traes_2BS_8368F6B5D.1	ta_iwgcsc_2bs_v1_5155056:10..6355	6346	16	91375.57	815	9.02	AGO4
19.	TaAGO5b	Traes_2DS_4CC8FD7E3.1	ta_iwgcsc_2ds_v1_5380297:2..6639	6638	16	94300.29	838	9.10	AGO5
20.	TaAGO5c	Traes_5BL_F505BF164.1	ta_iwgcsc_5bl_v1_10819703:15714..21283	5570	14	82474.57	732	8.95	AGO5
21.	TaAGO5d	Traes_4AL_7CC35DF1D.2	ta_iwgcsc_4al_v2_7146652:9555..15647	6093	16	98771.25	875	9.18	AGO5
22.	TaAGO5e	Traes_4DS_88D2821C6.2	ta_iwgcsc_4ds_v1_2318247:117..6407	6291	16	103082.25	916	9.22	AGO5
23.	TaAGO5f	Traes_3AS_3F8424E4E.1	ta_iwgcsc_3as_v1_1779497:2..5110	5109	14	93220.12	832	9.18	AGO5
24.	TaAGO5g	Traes_3B_CA99AB66C.1	ta_iwgcsc_3b_v1_9317732:662..5865	5204	14	94133.21	841	9.25	AGO5
25.	TaAGO6a	Traes_1BL_05F7B7DFA.1	ta_iwgcsc_1bl_v1_2609948:748..7252	6505	13	100043.38	893	6.99	AGO5
26.	TaAGO6b	Traes_5AL_07EFD5712.1	ta_iwgcsc_5al_v1_2801172:3925..11424	7500	16	98228.56	883	9.31	AGO6
27.	TaAGO6c	Traes_5DL_672EE3605.1	ta_iwgcsc_5dl_v1_4504514:5254..15341	10088	16	98062.91	882	9.31	AGO6
28.	TaAGO6d	Traes_5BL_F611D65E0.1	ta_iwgcsc_5bl_v1_10861763:7394..17699	10306	16	98241.09	883	9.30	AGO6
29.	TaAGO6e	Traes_7AL_D88450A3C.2	ta_iwgcsc_7al_v1_944917:567..5236	4670	13	86268.90	768	9.09	AGO6
30.	TaAGO7a	Traes_2AL_3F3117458.1	ta_iwgcsc_2al_v1_6334464:10651..14850	4200	3	105787.59	934	9.23	AGO7
31.	TaAGO7b	Traes_2BL_24111235C.1	ta_iwgcsc_2bl_v1_7955795:2637..7172	4536	3	115033.19	1023	9.33	AGO7
32.	TaAGO8	Traes_7AL_1BAB53DCE.1	ta_iwgcsc_7al_v1_4478819:3359..8183	4825	16	101590.86	901	9.11	AGO8
33.	TaAGO9a	Traes_1AL_095416BC0.1	ta_iwgcsc_1al_v2_3876661:1149..9699	8551	16	102455.28	925	9.09	AGO9
34.	TaAGO9b	Traes_1BL_7C037D478.2	ta_iwgcsc_1bl_v1_3899789:1061..10207	9147	16	102872.66	927	9.06	AGO9
35.	TaAGO9c	Traes_1DL_64B330BBB.2	ta_iwgcsc_1dl_v1_2275968:1227..11215	9989	16	102361.16	922	9.06	AGO9
36.	TaAGO10a	Traes_6AS_FBB2AFAAB.1	ta_iwgcsc_6as_v1_4364847:5932..12979	7048	14	105777.22	948	9.37	AGO10
37.	TaAGO10b	Traes_6DS_9DD64BD48.1	ta_iwgcsc_6ds_v1_2082993:2072..10005	7934	13	99080.09	883	9.95	AGO10
38.	TaAGO10c	Traes_7AL_96766587F.2	ta_iwgcsc_7al_v1_4543530:3452..7823	4372	9	65968.92	583	8.88	AGO10
39.	TaAGO10d	Traes_7DL_C538856D4.1	ta_iwgcsc_7dl_v1_3364674:1..5438	5438	12	74927.28	664	8.95	AGO10
RDR									
1.	TaRDR1a	Traes_6DL_4B89E8742.2	ta_iwgcsc_6dl_v1_3272251:691..6878	6188	0	127488.70	1120	6.98	RDR1
2.	TaRDR1b	Traes_6BL_78BEF51DD.1	ta_iwgcsc_6bl_v1_4353480:4946..8011	3066	0	82713.43	727	6.82	RDR1
3.	TaRDR1c	Traes_6AL_393C6B853.1	ta_iwgcsc_6al_v1_5823227:4526..9827	5302	0	127483.82	1119	7.15	RDR1
4.	TaRDR1d	Traes_6BL_0A9D15EDC.2	ta_iwgcsc_6bl_v1_4254864:3..1709	1707	0	54258.16	484	8.42	RDR1
5.	TaRDR1e	Traes_6BL_0BB5C493D.1	ta_iwgcsc_6bl_v1_4369818:3771..6377	2607	0	38289.35	333	5.40	RDR1
6.	TaRDR1f	Traes_6AL_13BC97E04.1	ta_iwgcsc_6al_v1_5769298:833..7192	6360	0	127303.48	1116	8.53	RDR1
7.	TaRDR1g	Traes_6BL_DF680C2AF.1	ta_iwgcsc_6bl_v1_4369819:1195..5786	4592	2	108562.11	947	7.34	RDR1
8.	TaRDR2a	Traes_2DL_6DB81005E.1	ta_iwgcsc_2dl_v1_9891666:1..7150	7150	2	82516.60	732	6.38	RDR2
9.	TaRDR2b	Traes_4AS_8D6311711.1	ta_iwgcsc_4as_v2_5940771:10843..26735	15893	2	126947.12	1127	6.60	RDR2
10.	TaRDR2c	Traes_4DL_2E9CE89D9.2	ta_iwgcsc_4dl_v3_1439157:7408..18499	11092	4	130417.41	1156	8.15	RDR2
11.	TaRDR2d	Traes_4DL_A54C80661.1	ta_iwgcsc_4dl_v3_14383226:1730..7149	5420	2	97124.16	869	6.82	RDR2
12.	TaRDR3	Traes_7BL_8CEC8F99B.2	ta_iwgcsc_7bl_v1_6687438:7..10355	10349	10	70361.58	616	6.20	RDR3
13.	TaRDR4	Traes_3AS_F27BB108C.2	ta_iwgcsc_3as_v1_3345716:2930..10167	7238	5	50505.60	436	5.75	RDR4
14.	TaRDR5	Traes_3B_2C6DB84FB.2	ta_iwgcsc_3b_v1_10500163:2098..12737	10640	0	63011.89	560	6.84	RDR5
15.	TaRDR6a	Traes_3B_DC77B5E89.1	ta_iwgcsc_3b_v1_10414255:3..4337	4335	0	104445.56	923	7.29	RDR6
16.	TaRDR6b	Traes_3DL_F32B49981.1	ta_iwgcsc_3dl_v1_6945798:1..3009	3009	9	60535.30	543	6.71	RDR6

[Genomic info such as gene names, accession number, genomic/chromosomal location, genome length, protein length were collected from Phytozome database (<https://phytozome.jgi.doe.gov/pz/portal.html>) and the molecular weight and PI values were predicted by the ExPASy ComputePI/Mwtool (http://au.expasy.org/tools/pi_tool.html). Molecular weights are in Daltons and “aa” means amino acid.]

and Fig.2). The third residue histidine (H) at 986th position (H986) is missed (“-”) in all TaAGO1e, TaAGO4b, TaAGO6a and TaAGO6e as well as the fourth residue histidine (H) at 798th position (H798) was replaced by the proline (P) in TaAGO4b, TaAGO6a and TaAGO6e (Table 2 and Fig.2). The fourth residue in the 798th position (H798) in AtAGO1 of the corresponding position in TaAGO1f, TaAGO1g, and TaAGO1h is replaced by arginine (R) (Table 2 and Fig.2). Unlike AtAGO1, TaAGO2a, and TaAGO2b, TaAGO3 however conserved aspartate (D) as a third catalytic residue instead of histidine (H) in the 986th position (H986) (Table 2 and Fig.2). On the other hand, the fourth triad residue histidine (H) in the 798th position (H798) is replaced by the amino acid proline (P) in all 11 TaAGO RNAi genes such as 2 TaAGO4s (TaAGO4a and TaAGO4c), 3 TaAGO5s (TaAGO5a, TaAGO5b and TaAGO5e), 3 TaAGO6s (TaAGO6b, TaAGO6c and TaAGO6d) and 3 TaAGO9s (TaAGO9a, TaAGO9b and TaAGO9c) (Table 2 and Fig.2). Furthermore, tyrosine (Y) was conserved instead of histidine in the 798th position (H798) by TaAGO10a and TaAGO10b proteins (Table 2 and Fig.2). The replacement/mutational changes of the D760, D845, H986 and H798 (DDH/H) catalytic residues in the reported wheat AGO genes are clearly indicating the functional distinction compared to the AGO proteins in Rice and *Arabidopsis*. The further investigation of these genes is demanded for deeper understanding about the PIWI domain nuclease activities as well as the biological functionality in Wheat.

Finally, total 16 TaRDR proteins however possessed an RdRp domain (Table 3). The gene length of the TaRDR proteins varied from 15893bp for TaRDR2b to 1707 bp for TaRDR1d with their corresponding potentially encoding 1127 and 484 amino acids respectively (Table 1). Additionally, TaRDR proteins were studied to identify the presence of the catalytic motif in the RdRp domain. The sequence alignment revealed that all members of TaRDR1, TaRDR2, and TaRDR6 groups shared a common DLDGD motif in the catalytic domain, whereas the proteins TaRDR3, TaRDR4, and TaRDR5 retained DFDGD motif (Fig. 3).

Fig.2 Alignment profile of PIWI domain amino acids of wheat (*T.aestivum*) AGO proteins. The protein sequences were aligned using MEGA 7.0. The conserved Asp, Asp and His (DDH) triad residues corresponding to D760, D845, and H986 of *Arabidopsis* AGO1 are indicated with downward arrows, whereas the conserved H residue corresponding to H798 of *Arabidopsis* AGO1 is boxed with blue color. Amino acid positions corresponding to each protein are indicated at the end of each line.

Fig.3 Alignment of catalytic regions in RdRp domains of RDR proteins of *T. aestivum*, rice, and *A. thaliana*.

The conserved motif DLDGD and DFDGD were colored. The alignments were performed using MEGA 7.0. A ‘-’ indicates of an amino acid missing of the corresponding protein sequence. The locations analogous to each protein are mentioned at the end of line.

Table 2: Comparison of argonaute proteins with missing catalytic residue(s) present in PIWI domain of wheat (*T.aestivum*), rice^b and *Arabidopsis*^b

	Wheat		Rice		<i>Arabidopsis thaliana</i>	
	Argonaute	Motif ^a	Argonaute	Motif ^b	Argonaute	Motif ^b
1	TaAGO1a	DDH/H	OsAGO1	DDH/P	AGO2	DDD/H
2	TaAGO1b	DDH/H	OsAGO2	DDD/H	AGO3	DDD/H
3	TaAGO1c	DDH/H	OsAGO3	DDD/H	AGO4	DDH/S
4	TaAGO1d	DDH/H	OsAGO4a	DDH/P	AGO6	DDH/P
5	TaAGO1e	DD-/H	OsAGO4b	DDH/P	AGO9	DDH/R
6	TaAGO1f	DDH/R	OsAGO11	GDH/H		
7	TaAGO1g	DDH/R	OsAGO13	-DH/H		
8	TaAGO1h	DDH/R	OsAGO15	DDH/P		
9	TaAGO1i	DDH/H	OsAGO16	DDH/P		
10	TaAGO1j	DDH/H	OsAGO17	HDR/C		
11	TaAGO1k	DDH/H	OsAGO18	DDH/S		
12	TaAGO2a	DDD/H				
13	TaAGO2b	DDD/H				
14	TaAGO3	DDD/H				
15	TaAGO4a	DDH/P				
16	TaAGO4b	DD-/P				
17	TaAGO4c	DDH/P				
18	TaAGO5a	DDH/P				
19	TaAGO5b	DDH/P				
20	TaAGO5c	DDH/H				
21	TaAGO5d	DDH/H				
22	TaAGO5e	DDH/P				
23	TaAGO5f	DDH/H				
24	TaAGO5g	DDH/H				
25	TaAGO6a	DD-/P				
26	TaAGO6b	DDH/P				
27	TaAGO6c	DDH/P				
28	TaAGO6d	DDH/P				
29	TaAGO6e	DD-/P				
30	TaAGO7a	DDH/H				
31	TaAGO7b	DDH/H				
32	TaAGO8	YDH/P				
33	TaAGO9a	DDH/P				
34	TaAGO9b	DDH/P				
35	TaAGO9c	DDH/P				
36	TaAGO10a	DDH/Y				
37	TaAGO10b	DDH/Y				
38	TaAGO10c	DDH/H				
39	TaAGO10d	DDH/H				

^aMotifs are correspond to conserved D760, D845, H986/H798 of Arabidopsis AGO1; D(aspartate), H(Histidine), K(Lysine), A(Alanine), P(Proline), Q(Glutamine), G(Glycine), R(Arginine), C(Cysteine), S(Serine); ‘-’ represents the missing catalytic residue

^b Reviewed in Kapoor *et al* (2008)

3.2 Phylogenetic tree analysis of the putative RNAi genes in *T.aestivum*, *A.thaliana* and rice

In order to infer the evolutionary history of the three groups of RNAi genes in wheat (*T. aestivum*) with those of the *Arabidopsis* and rice homologs, three independent phylogenetic trees were created from the multiple sequence alignment. Encoded protein sequences (S1, S2 and S3 data) of all RNAi genes were used to construct the trees using

neighbor-joining(NJ) method with 1000 bootstrap replications (Fig. 4).The TaDCL proteins showed high sequence conservation compared with their *Arabidopsis* and rice counterpart. Fig. 4A shows that the 7TaDCLs were obviously classified into four distinct clades. All the TaDCL members of the corresponding group are clustered with the similar ortholog AtDCLs and OsDCLs at a high similarity context. In relation to the phylogenetic association and sequence homology with AtDCLs, the TaDCLs were named as TaDCL1a, TaDCL1b, TaDCL3a, TaDCL3b, TaDCL3c, TaDCL3d, and TaDCL4. However, there were no TaDCL2 gene(s) corresponding to AtDCL2 and OsDCL2.

Fig.4 Phylogenetic trees of all *A. thaliana*, rice (*Oryza sativa*) wheat (*T. aestivum*) RNAi gene families. (A) DCL (B) AGO and (C) RDR RNAi group. The trees were produced by MEGA 7.0 software using Neighbor-Joining (NJ) method with bootstrap of 1000.*T. aestivum* proteins are indicated with a red colour circle before the corresponding taxon names.

Based on NJ phylogenetic analysis and the protein sequence homology with AtAGOs, the 39 candidates of TaAGO family consisted of 11 AGO1s (TaAGO1a-TaAGO1k), 2AGO2s (TaAGO2a and TaAGO2b), one AGO3 (TaAGO3), 3AGO4s (TaAGO4a, TaAGO4b and TaAGO4c), 7 AGO5s (TaAGO5a-TaAGO5g), 5AGO6s (TaAGO6a-TaAGO6e), 2AGO7s (TaAGO7a and TaAGO7b), one AGO8 (TaAGO8), 3AGO9s (TaAGO9a-TaAGO9c) and 4AGO10s (TaAGO10a-TaAGO10d) for 8 clusters (Fig. 2B). It is however observed from the phylogenetic tree that TaAGO8 and 3TaAGO9s (TaAGO9a-TaAGO9c) form a group with the nearest taxa OsAGO4b not with the corresponding AtAGO8 and AtAGO9. On the other hand, it is unusually obvious that TaAGO5a and TaAGO5b group with OsAGO14. TaAGO5c, TaAGO5d, TaAGO5e form a group with OsAGO11 and OsAGO12 as well as TaAGO5f and TaAGO5g specifically with taxa OsAGO18 but not with any AtAGOs. The characteristic of these genes may however indicates that TaAGO8, 3TaAGO9s, and 7TaAGO5s are the monocot specific that have originated during evolutionary process[7,10].

Finally, like DCL and AGO, RDR genes in *T. aestivum* were named after the *Arabidopsis* homologs, which were the evolutionary neighboring in the phylogenetic tree, and exhibit the top protein sequence homologies. Fig. 2C illustrates that the phylogenetic tree however produced from the RDR protein sequences is separated into four major clusters. RDR1 clade has seven members (TaRDR1a, TaRDR1b, TaRDR1c, TaRDR1d, TaRDR1e, TaRDR1f, TaRDR1g,) form a cluster with their AtRDR1 and OsRDR1 counterparts. Clade RDR2 has four members

(TaRDR2a, TaRDR2b, TaRDR2c, TaRDR2d,) and clearly generates a distinct cluster with their AtRDR2 and OsRDR2 homologs. TaRDR3, TaRDR4, and TaRDR5 together form a cluster with their similar genes to AtRDR3, AtRDR4, AtRDR5, OsRDR3, and OsRDR4. Clade RDR6 has two members TaRDR6a and TaRDR6b produced a distinct group with AtRDR6 and OsSHL2 genes in the phylogenetic tree imply that these TaRDR protein families convey the similar genetic information of their AtRDR and OsRDR counterparts.

3.3 Discovery of conserved domains and motifs of the putative DCL, AGO and RDR proteins in *T.aestivum*

Protein domains have functional importance in protein structure. They are made up of a chain of amino acids of independent molecular folding. Domains have unique function in respective protein. Functional conserved domain analysis for protein structure in *T.aestivum* of the predicted TaDCLs, TaAGOs and TaRDRs were performed using Pfam, SMART and NCBI-CDD databases. The obtained search results are listed in Table 3.

It is observed from the search results from all three databases that all TaDCL protein families possessed all most all assumed six common domains viz., DEAD box/ResIII, Helicase-C, Dicer-dimer, PAZ, RNase III and DSRM (Table 3 and Fig.5). These were also conserved by the 4 DCL protein families in model plant *Arabidopsis*[7,10,70]. Previous study suggests that these functional domains are known as the highly important plant DCL domains for protein structure[7,10,11,70]. Ribonuclease-3(RNaseIII) domain characteristics of TaDCL proteins might enable them to cleave dsRNA to generate small interfering RNAs(siRNAs) that play crucial functions to regulate gene expression in plants[10,11,64]. On the other hand, earlier investigations revealed that Argonaute proteins usually retain a Piwi-Argonaute- Zwillie (PAZ) domain and a PIWI domain[71,72]. Functional domain search results (Table 3) and domain organization by Pfam (Fig.5) showed that all 39 TaAGO protein families also hold the two most common domains, viz., N-terminus PAZ and C-terminus PIWI domain similar to that of *Arabidopsis* and rice known as potential characteristic of plant AGO families[7,13]. Additionally, a DUF1785 domain also can be recognized before PAZ domain in all TaAGO proteins except for TaAGO1b, TaAGO10c and TaAGO10d encodes the shortest amino acids chain, which may however be for the possible of loss of its N-terminal sequence during evolution

Table 3. List of predicted conserved functional domains of the DCL, AGO and RDR proteins in wheat (*T.aestivum*) using Pfam, SMART and NCBI-CDD.

Gene No.	Gene Name	Accession Number	Pfam	Domains SMART	NCBI-CDD
DCL					
1.	TaDCL1a	Traes_4BL_B3A1B8342.2	Ribonuclease_3,Ribonuclease_3, PAZ,Dicer_dimer,Helicase_C,DND1_DSRM,ResIII,dsrm	DEXDc,HELICc,PAZ,RIBOc,RIBOc,DSR M,DSRM	DEXHc_dicer,SF2_C_dicer(C-terminal domain/HelicC), PAZ super family, RIBOc, RIBOc, Dicer_dimer, DSRM_SF super family, DSRM_SF super family
2.	TaDCL1b	Traes_5AL_72A7552B9.2	Ribonuclease_3,Ribonuclease_3, PAZ,Dicer_dimer,Helicase_C,DND1_DSRM, dsrm, ResII,, DUF2838	DEXDc,HELICc,PAZ,RIBOc,RIBOc,DSR M,DSRM	DEXHc_dicer, PAZ_CAF_like, SF2_C_dicer(C-terminal domain/HelicC), Rnc, DSRM_DCL_plant, RIBOc, Dicer_dimer
3.	TaDCL3a	Traes_1AL_E7144546E.1	Ribonuclease_3,Ribonuclease_3, Dicer_dimer,Helicase_C,DEAD,PAZ,dsrm, DND1_DSRM	DEXDc,HELICc,PAZ,RIBOc,RIBOc,DSR M,DSRM	DEXHc_dicer, MPH1 super family(ERCC4-related helicase), RIBOc, Dicer_dimer, RIBOc, DSRM_SF super family
4.	TaDCL3b	Traes_1DL_C646B6990.1	Ribonuclease_3,Ribonuclease_3, Dicer_dimer,Helicase_C,PAZ,dsrm, DND1_DSRM	HELICc,PAZ,RIBOc,RIBOc,DSRM,DSRM	RIBOc, DEAD-like_helicase_N super family, DEAD-like_helicase_C super family(also called Helicase-C), Dicer_dimer, RIBOc, PAZ super family, DSRM_SF super family
5.	TaDCL3c	Traes_3AL_562D6614F.1	Ribonuclease_3,Ribonuclease_3, Dicer_dimer,Helicase_C,PAZ,ResIII, dsrm, DND1_DSRM	DEXDc, HELICc, DSRM, PAZ, RIBOc, RIBOc, DSRM	DEXHc_dicer, SF2_C_dicer (C-terminal domain/HelicC), Dicer_dimer, PAZ super family , RIBOc RIBOc, , DSRM_SF super family, DSRM_SF super family
6.	TaDCL3d	Traes_3DL_2DC78B18A.1	Ribonuclease_3,Ribonuclease_3, Dicer_dimer,Helicase_C,PAZ,ResIII,dsrm	DEXDc,HELICc,DSRM,PAZ,RIBOc,RIBOc,DSRM, DSRM	DEXHc_dicer, SF2_C_dicer (C-terminal domain/HelicC), PAZ super family ,RIBOc , RIBOc, Dicer_dimer, DSRM_SF super family, DSRM_SF super family
7.	TaDCL4	Traes_2DL_E96DCDCB4.2	Ribonuclease_3,Ribonuclease_3, Dicer_dimer,Helicase_C,DND1_DSRM,PAZ,dsrm	HELICc,PAZ,RIBOc,RIBOc,DSRM	DEAD-like_helicase_C super family(also called Helicase-C), Rnc, PAZ super family , Dicer_dimer, RIBOc, DSRM_SF super family, DSRM_SF super family
AGO					
1	TaAGO1a	Traes_2AL_2512A7F91.1	Piwi,Gly-rich_Ago1,ArgoN,PAZ,ArgoL1,ArgoL2,ArgoMid	DUF1785,PAZ,Piwi	Gly-rich-Ago, ArgoN, ArgoL1, PAZ-argonaute, PIWI
2.	TaAGO1b	Traes_2BL_93099ACF4.1	Piwi,PAZ,ArgoL2	PAZ,Piwi	PAZ, PIWI
3.	TaAGO1c	Traes_6AL_616161AAB.1	Piwi,ArgoN,PAZ,ArgoL1,ArgoL2,ArgoMid	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PAZ, PIWI
4.	TaAGO1d	Traes_6DL_58620B158.2	Piwi,ArgoN,PAZ,ArgoL1,ArgoL2,ArgoMid	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PAZ, PIWI
5.	TaAGO1e	Traes_7DL_C255A109C.1	Piwi,ArgoN,PAZ,ArgoL1,ArgoMid	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PAZ, PIWI
6.	TaAGO1f	Traes_6BL_9CFA54D4A.1	Piwi,Gly-rich_Ago1,ArgoN,PAZ,ArgoL1,ArgoL2,ArgoMid	DUF1785,PAZ,Piwi	Gly-rich-Ago, ArgoN, ArgoL1, PAZ ,PIWI
7.	TaAGO1g	Traes_6AL_317133B3F.2	Piwi,Gly-rich_Ago1,ArgoN,PAZ,ArgoL1,ArgoL2,ArgoMid	DUF1785,PAZ,Piwi	Gly-rich-Ago, ArgoN, ArgoL1, PAZ, PIWI
8.	TaAGO1h	Traes_6DL_804FB7F75.1	Piwi,Gly-rich_Ago1,ArgoN,PAZ,ArgoL1,ArgoL2,ArgoMid	DUF1785,PAZ,Piwi	Gly-rich-Ago, ArgoN, ArgoL1, PAZ, PIWI
9.	TaAGO1i	Traes_7AS_56569A5AC.2	Piwi,Gly-rich_Ago1,ArgoN,PAZ,ArgoL1,ArgoL2,ArgoMid	DUF1785,PAZ,Piwi	Gly-rich-Ago, ArgoN, ArgoL1, PAZ, PIWI
10.	TaAGO1j	Traes_7DS_4D01B6175.1	Piwi,ArgoN,PAZ,ArgoL1,ArgoL2,ArgoMid	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PIWI
11.	TaAGO1k	Traes_4AL_A118C6C84.2	Piwi,ArgoN,PAZ,ArgoL1,ArgoL2,ArgoMid	DUF1785,PAZ,Piwi	Gly-rich-Ago
12.	TaAGO2a	Traes_2AL_DFE4C65F6.2	Piwi,PAZ,ArgoN,ArgoL1,ArgoL2,ArgoMid	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PAZ, PIWI
13.	TaAGO2b	Traes_2BL_7713B3533.2	Piwi,PAZ,ArgoN,ArgoL2,ArgoL1,ArgoMid	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PAZ, PIWI
14.	TaAGO3	Traes_2DL_A77212060.2	Piwi,PAZ	DUF1785,PAZ,Piwi	ArgoL1, PAZ, ArgoN, PIWI
15.	TaAGO4a	Traes_3AS_8E711E2C.2	Piwi,ArgoN,PAZ,ArgoL1,ArgoL2	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PAZ, PIWI
16.	TaAGO4b	Traes_3DS_57EA31670.1	Piwi,ArgoN,PAZ,ArgoL1,ArgoL2	DUF1785,PAZ,Piwi	ArgoN, ArgoL1,PAZ, PIWI
17.	TaAGO4c	Traes_3B_F4E4667F8.1	Piwi,ArgoN,PAZ,ArgoL1,ArgoL2	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PAZ, PIWI
18.	TaAGO5a	Traes_2BS_8368F6B5D.1	Piwi,ArgoN,PAZ,ArgoL1,ArgoL2,ArgoMid	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PAZ, PIWI
19.	TaAGO5b	Traes_2DS_4CC8F07E3.1	Piwi,ArgoN,PAZ,ArgoL1,ArgoL2,ArgoMid	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PAZ, PIWI
20.	TaAGO5c	Traes_5BL_F505BF164.1	Piwi,PAZ,ArgoL1,ArgoL2,ArgoMid	DUF1785,PAZ,Piwi	ArgoL1, PAZ, PIWI
21.	TaAGO5d	Traes_4AL_7CC35DF1D.2	Piwi,ArgoN,PAZ,ArgoL1,ArgoMid,ArgoL2	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PAZ, PIWI
22.	TaAGO5e	Traes_4DS_88D2821C6.2	Piwi,ArgoN,PAZ,ArgoL1,ArgoMid,ArgoL2	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PAZ, PIWI
23.	TaAGO5f	Traes_3AS_3F8424E4E.1	Piwi,ArgoN,PAZ,ArgoL1,ArgoMid,ArgoL2	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PAZ, PIWI
24.	TaAGO5g	Traes_3B_CA99AB66C.1	Piwi,ArgoN,PAZ,ArgoL1,ArgoL2,ArgoMid	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PAZ, PIWI
25.	TaAGO6a	Traes_1BL_05F7B7DFA.1	Piwi,ArgoN,PAZ,ArgoL1,ArgoL2	DUF1785,PAZ,Piwi	ArgoL1, PAZ, PIWI
26.	TaAGO6b	Traes_5AL_07EFD5712.1	Piwi,ArgoN,PAZ,ArgoL2,ArgoL1	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PAZ, PIWI
27.	TaAGO6c	Traes_5DL_672EE3605.1	Piwi,ArgoN,PAZ,ArgoL2,ArgoL1	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PAZ, PIWI
28.	TaAGO6d	Traes_5BL_F611D65E0.1	Piwi,ArgoN,PAZ,ArgoL2,ArgoL1	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PAZ, PIWI
29.	TaAGO6e	Traes_7AL_D88450A3C.2	Piwi,PAZ,ArgoN,ArgoL2,ArgoL1	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PAZ, PIWI
30.	TaAGO7a	Traes_2AL_3F3117458.1	Piwi,ArgoN,PAZ,ArgoL1,ArgoL2	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PAZ, PIWI
31.	TaAGO7b	Traes_2BL_24111235C.1	Piwi,ArgoN,PAZ,ArgoL1,ArgoL2	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PAZ
32.	TaAGO8	Traes_7AL_1B4B53DCE.1	Piwi,ArgoN,PAZ,ArgoL1,ArgoL2,ArgoMid	DUF1785,PAZ,Piwi	PIWI
33.	TaAGO9a	Traes_1AL_095416BC0.1	Piwi,ArgoN,PAZ,ArgoL1,ArgoL2	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PAZ, PIWI
34.	TaAGO9b	Traes_1BL_7C037D478.2	Piwi,ArgoN,PAZ,ArgoL1,ArgoL2	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PAZ, PIWI
35.	TaAGO9c	Traes_1DL_64B330BBB.2	Piwi,ArgoN,PAZ,ArgoL1,ArgoL2	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PAZ, PIWI
36.	TaAGO10a	Traes_6AS_FBB2AFAAB.1	Piwi,ArgoN,PAZ,ArgoL1,ArgoL2,ArgoMid	DUF1785,PAZ,Piwi	ArgoN, ArgoL1, PAZ, PIWI
37.	TaAGO10b	Traes_6DS_9DD64BD48.1	Piwi,PAZ,ArgoL1,ArgoL2,ArgoMid	DUF1785,PAZ,Piwi	ArgoL1, PAZ, PIWI
38.	TaAGO10c	Traes_7AL_96766587F.2	Piwi,ArgoL2,ArgoMid,PAZ	Piwi	PAZ, PIWI
39.	TaAGO10d	Traes_7DL_C538856D4.1	Piwi,PAZ,ArgoL2,ArgoMid	PAZ,Piwi	PAZ, PIWI
RDR					
1.	TaRDR1a	Traes_6DL_4B89E8742.2	RdRP	RRM	RRM_SF,RdRP
2.	TaRDR1b	Traes_6BL_78BEF51DD.1	RdRP	transmembrane region	RdRP
3.	TaRDR1c	Traes_6AL_393C6B853.1	RdRP	RRM	RRM_SF,RdRP
4.	TaRDR1d	Traes_6BL_0A9D15EDC.2	RdRP	no domains	RdRP
5.	TaRDR1e	Traes_6BL_0BB5C493D.1	RdRP	low complexity region	RdRP
6.	TaRDR1f	Traes_6AL_13BC97E04.1	RdRP	RRM	RRM_SF,RdRP
7.	TaRDR1g	Traes_6BL_DF680C2AF.1	RdRP	low complexity region	RdRP
8.	TaRDR2a	Traes_2DL_6DB81005E.1	RdRP	no domains	RdRP
9.	TaRDR2b	Traes_4AS_8D6311711.1	RdRP	low complexity region	RdRP
10.	TaRDR2c	Traes_4DL_2E9CE89D9.2	RdRP	low complexity region	RdRP
11.	TaRDR2d	Traes_4DL_A54C80661.1	RdRP	low complexity region	RdRP
12.	TaRDR3	Traes_7BL_8CEC8F99B.2	RdRP	no domains	RdRP
13.	TaRDR4	Traes_3AS_F27BB108C.2	RdRP	no domains	RdRP
14.	TaRDR5	Traes_3B_2C6DB84FB.2	RdRP	no domains	RdRP
15.	TaRDR6a	Traes_3B_DC77B5E89.1	RdRP	low complexity region	RdRP
16.	TaRDR6b	Traes_3DL_F32B49981.1	RdRP	low complexity region	RdRP

Fig.5 Domain organization of DCL, AGO and RDR protein families in wheat (*T.aestivum*). Conserved

functional domains were predicted from Pfam database. Domains are indicated in different color boxes with their corresponding name inside the box.

Besides, the *T.aestivum* genome encrypted 16 TaRDR protein families sharing a mutual motif analogous to the catalytic β' subunit of RNA-dependent RNA-polymerases(RdRp)(Table 3 and Fig.5), which is also supported by the earlier study[33].

Fig.6 Discovery of 20 conserved motif structures. Motifs with consensus amino acid sequence in wheat DCL,

AGO and RDR protein families predicted by the web-based MEME suite. Each motif in the protein domain is represented by the color box and the order of the motifs corresponds to position of motifs in individual protein sequence.

Motifs in a DNA, RNA or a protein sequence encode different biological functions[73]. Their identification and characterization is significant to study molecular interactions in cell such as the regulation of gene expression. These motifs are also known as the small active site of an enzyme or a structural component essential for accurate folding of the protein and molecular evolution[73]. In this work, Multiple EM for Motif Elicitation (MEME) was used to discover motifs for each group of TaDCL, TaAGO and TaRDR protein sequences along with that of RNAi protein families of *Arabidopsis*. The search was done for each member for each of the family and 20 motifs were predicted for each of the proteins. It is observed that most of the motifs were conserved in each of the TaDCL protein members and distributed in order in almost all four subfamilies of the DCL protein family like *Arabidopsis* (Fig.6). Wheat DCL4 protein subfamily, TaDCL4, however was identified to possess a few C-terminal motifs reorganized in comparison of the paralogs AtDCL4, which was also supported by the previous demonstration[10]. All TaDCL protein subfamilies retain 20 conserved motifs with the exception of that TaDCL3b and TaDCL4 contain 19 and 16 motifs respectively (Fig.6).

On the other hand, out of 39 TaAGO protein subfamilies, the maximum 28 protein members of wheat DCL, AGO and RDR retained 16-20 conserved motifs in their protein domain structures with some variations in other 11 subfamilies. High conservation was found in TaAGO1, TaAGO9, and TaAGO10 family members with some exceptions that TaAGO1b, TaAGO1e, and TaAGO10c contained 16, 14, and 15 motifs respectively (Fig.6).

Motif search using MEME for the RDR protein family in wheat demonstrates that 10 TaRDR subfamilies possess 17-20 functional motifs in the RdRp domain similar to that of the AtRDR protein members (Fig.6) but with little high distinction that TaRDR1e, TaRDR3, TaRDR4, and TaRDR5 contain only 7-8 motifs. The study of sequence motifs of the protein subfamilies of each group of RNAi genes, it suggests that there is a tendency of high conservation of functional motifs in respective domains within the members of same subfamilies in TaDCL, TaAGO and TaRDR groups that are assumed to hold and perform same structural functions. Therefore, discovery and broad study of these motifs is important to structure models of cellular processes at the molecular scale and to identify the mechanisms of diseases in eukaytic groups[58,73].

3.4 Assembly of exon-intron structure of *T. aestivum* DCL, AGO and RDR Genes

The exon-intron organization of TaDCL, TaAGO and TaRDR genes in wheat (*T.aestivum*) was studied using GSDS2.0 web-based tool for more broad understanding regarding their probable structural progression and level of similarity with the respective RNAi genes of *Arabidopsis thaliana*. Our analysis for all three gene families illustrated that intron and exon numbers were usually conserved among the members of the same group though varied considerably in different groups of the same family that is similar to those of the AtDCL, AtAGO, and AtRDR gene families. The intron numbers varied from 9 to 18 among TaDCL protein members while 4 AtDCLs possessed 11-17 introns (Table 1 and Fig.7). The two members of TaDCL1 sub-group contained 9 introns, which is the lowest in the group whereas only TaDCL3a contained the double introns compared to this sub-group. Other four members retained 15/16 introns (Table 1 and Fig.7). On the other hand, almost all TaAGO protein families also follow the same range of introns of their AtAGOs counterpart commonly measuring 13-18 although TaAGO2a/2b, TaAGO3 and TaAGO7a/7b displayed only lowest intron numbers among all argonaute gene members in wheat(*T.aestivum*) like AtAGO2 , AtAGO3 and AtAGO7(Table 1 and Fig.7). The members of the TaAGO1 sub-family maintained high similarity conservation in containing introns16-18 as in AtAGO1 with exception 11 introns in TaAGO1b (Table 1 and Fig.7).Furthermore all members of the rest of the sub-families follow almost the same pattern in containing introns and exons of their *Arabidopsis* counterpart. Among the TaRDR protein elements, TaRDR3 contained the highest number of 10 introns followed by 9 and 5 introns in TaRDR5 and TaRDR4

respectively (Table 1 and Fig.7). Interestingly, all gene members of TaRDR1 and TaRDR6 sub-group had no introns though AtRDR1 possessed 2 introns and AtRDR6 had 0 introns that might be of some significant implication that these gene sub-families may be responsible for any organ or tissue specific expression characteristics and bear any particular disease resistance potentiality (Table 1 and Fig.7). Earlier investigation suggested CaRDR1 possessed important role in pepper resistance against TMV[74].It was however observed that the exon-intron structure of TaDCL, TaAGO and TaRDR genes was very similar within members of the same groups though substantially different through various groups of the same family indicate that these gene families particularly TaAGO and TaRDR families have experienced recurrent gene replication and recombination during evolution.

Fig.7 Exon-intron structure of *T. aestivum* RNAi genes corresponding to *A. thaliana* DCL, AGO and RDR genes. Exons (green box), intron (black lines), and intron phases (0, 1, and 2) are mentioned. The gene structure was estimated using the online GSDS1.0 server by considering their full-length genomic/DNA and coding sequences (CDS).

3.5 Gene set enrichment analysis of the RNAi genes in wheat (*T.aestivum*)

Gene set enrichment analysis(GESA) such as gene ontology (GO) analysis was conducted to demonstrate that how the predicted RNAi genes in *T.aestivum* are involved in different biological processes (BP), what molecular functions(MF) they perform with their interrelations and what are the cellular components where these molecular functions occur(S4 data , S1 Fig. and Fig.8). An interconnected relationship using QuickGO web-based tool from less specific concepts to more specific concepts among the GO terms and the corresponding gene ontologies are demonstrated using the three hierarchical GO slim trees for BP (S1A Fig.), MF (S1B Fig.), and CC (S1C Fig.). It is shown from the GO analysis that 11 gens are involved in post-transcriptional gene silencing (PTGS) pathway (GO:0016441;p-value: 5.00e-24), 9 genes are responsible for RNAi interference mechanism(GO:0016246;p-value:3.60e-22) and 11 genes are linked to gene silencing pathway(GO:0016458;p-value:1.30e-19)(S4 data). Post-transcriptional gene silencing (PTGS) in plants is an RNA-degradation technique that displays connections to RNA interference (RNAi) [75,76]. GO slim hierarchy tree illustrates that RNA interference is a sub-type of PTGS ultimately connected to gene expression (GO: 0010467; p-value: 7.20e-05) (S1A Fig.). There are also a number of biological processes are related to be occurred by the predicted RNAi genes in wheat (*T.aestivum*) some of which

important for wheat plant are response to stimulus (GO: 0050896; 0.00017), immune system process (GO: 0002376; 3.70e-14) and developmental process (GO: 0032502, 5.40e-06) (S4 data and S1A Fig). GO tree displayed that response to stimulus is termed into three important sub-types GO such as response to stress (GO:0006950; 2.80e-05) linked to 10 RNAi genes, response to external stimulus (GO:0009605; 7.30e-11) and response to biotic stimulus (GO:0009607; 1.20E-11). These three terms related to defense response to virus (GO: 0051607; 4.80e-22) called ultimately virus induced gene silencing (GO: 0009616; 6.10e-20). It is also obvious from the GO tree that response to stimulus and immune system process is associated to the defense response to other organism (GO: 0098542; 3.90E-13) (S1A Fig.). GO analysis for BP also suggests that some very important functions are maintained by the RNAi genes in wheat (*T.aestivum*), viz., post-embryonic development (GO: 0009791; 6.90E-05), anatomical structure development (GO: 0048856; 0.00014) which are eventually connected to wheat plant developmental process (GO: 0032502; 2.87e-04) (S1A Fig.) Previous investigation also reported that *Arabidopsis* DCL and RDR RNAi genes are actively associated to defense response to various biotic, abiotic stressors and play role for developmental progression [41,42,50–54].

Fig.8 Heatmap and Venn diagram for the predicted GO terms corresponding to the wheat RNAi genes. (A)

biological process (B) molecular function (C) cellular components whether the genes are related (Present) or not (Absent). Histograms corresponding to each GO terms represent the p-value (-log (p-value)) showed in the opposite site of GO terms in the heatmap. The Venn diagrams display the common GO terms shared by three gene families in terms of (D) biological process (E) molecular functions (F) cell components.

GO enrichment analysis showed that 8 putative RNAi genes out of 62 are associated to perform endonuclease type molecular function (GO:0004519;p-value: 1.20e-10)(S4 Data) that have positive link to the RNA-induced silencing complex(RISC-mediated) mediated cleavage actions into cell [7,10,76,77]. This multi-protein molecular compartment play significant role in mRNA degradation through that regulate gene expression. Argonaute proteins also act as attaching themselves with siRNAs in RISC for cleavage defined as endonucleolytic function that results in completion of PTGS for target mRNAs substrate [78]. GO analysis also demonstrates that three essential molecular functions(MF), viz., nucleic acid binding (GO:0003676; 5.00e-26), protein binding (GO:0005515; 2.80e-17), RNA binding (GO:0003723;p-value:0.00257) are associated to 44, 42 and 7 predicted RNAi genes in wheat (*T.aestivum*)(S4 Data) which implied that RNAi proteins are actively engaged in RISC as well as gene silencing. Nucleic acid binding is also of two types: heterocyclic compound binding (GO: 1901363; 1.80e-10) and

organic cyclic compound binding (GO: 0097159; 1.80e-10) (S4 Data). GO ancestry tree showed that all these MFs are the child terms of the GO term binding (GO: 0005488; 0.00217) (S1B Fig.). There is another MF termed as ribonuclease III activity (GO: 0004525; p-value: 7.00E-15) is related to 7 RNAi genes that is the child GO term of double-stranded RNA-specific ribonuclease activity (GO: 0032296; p-value: 7.00e-15) and nuclease activity (GO: 0004518; p-value: 6.70e-09) (S1B Fig.). The identified TaAGO proteins possess the two common domains like other plant AGO proteins such as PAZ and PIWI that perform essential functions to build protein complex with RNA or DNA template. The PAZ domain has a particular nucleic acid-binding fold which help domain to bind to the certain site of the nucleic acids[79,80]. Different molecular functions are usually tend to occur in various cellular components (CCs). GO analysis also suggests that the different MFs are predicted to happen significantly in nucleus (GO: 0005634; 8.00e-05), intracellular organelle lumen (GO: 0070013; 2.20e-05), organelle lumen (GO: 0043233; 2.20e-05), and membrane-enclosed lumen (GO: 0031974; 2.50e-05) (S1C Fig.).

The Venn diagram also illustrates that TaDCLs, TaAGOs and TaRDRs shared significant number of GO pathways. For BP, there were 91 GO enrichment pathways in general (Fig.7D) shared by the identified RNAi genes in wheat (*T.aestivum*), that suggests that RNAi genes are significantly associated to many important BPs. Furthermore, it were shown that the identified putative RNAi genes displayed a pool of mutual pathways for performing MFs and these MFs area predicted to occur in 6 common CCs(Fig.8E and Fig.8F). GESA however provided insights about the identification of the involvement of 62 wheat (*T.aestivum*) RNAi genes in different biological processes; engagement in performing a number of molecular actions in different cellular components for the survival of wheat plant in diverse environments.

3.6 Prediction of sub-cellular location of the TaDCL, TaAGO and TaRDR gene families in wheat (*T.aestivum*)

Prediction of sub-cellular localization (SCL) helps to know the presence or absent status of the identified RNAi genes in certain cellular organs in plants. A number of RNA silencing machinery genes in wheat (*T.aestivum*) were found locating in more than one cellular locations (Fig. 9). It is however observed that most of the RNA silencing genes were found to be located in cytosol (DCL 71.4 %, AGO 87.2 % and RDR 87.5 %) followed by plastid (DCL 14.3 %, AGO 33.3 % and RDR 31.2 %). A few of the AGO (20.5 %) and RDR (12.5 %) genes were located in

mitochondria but no DCL protein was found in mitochondria. On the other hand, ER, golgi and peroxisome contain no RNA silencing genes (**Fig. 9D and S2 Table**). Cytosol is the place of maximum metabolism in plants and most of the proteins in the cell are located in cytosol [81,82]. Since the PTGS takes place in the cytoplasmic region [77], this suggests that the identified putative RNAi genes are actively engaged in PTGS process.

3.7 Exploration of gene regulatory relationship between TFs and RNAi genes in wheat (*T.aestivum*)

A transcription factor (TF) acts as a key for a particular gene expression in lives. It is also known as a DNA-binding amino acid chain, which particularly works attaching with a particular *cis*-acting component in the promoter site of a gene for gene expression in plants and other eukaryotic groups [83]. Recent research in plant molecular biology has attracted genomics researchers to study the important transcription factor families to study the promoting and controlling power of gene expression against different biotic and abiotic stresses such as fungi, bacteria, viruses, drought, salinity, cold, hormone, pathogenesis, cell growth, and development in higher plants. Therefore, identification and characterization of TF families regulating the RNAi genes is however essential to study the gene silencing process in wheat (*T.aestivum*).

Fig.9 Prediction of subcellular locations of TaDCL, TaAGO and TaRDR gene families. Prediction were made for cytosol (cytos), endoplasmic reticulum (ER), extracellular (extra), golgi apparatus (golgi), membrane (membr), mitochondria (mito), nuclear (nucl), peroxisome (pero), plastid (plast) and vacuole (vacu) for each of the (A) DCL, (B) AGO and (C) RDR gene families in *T. aestivum* with the help of PSI and R-3.6.3.

In this analysis, 375 TFs were identified those regulate the predicted RNAi genes in wheat (*T.aestivum*) (S5 Data and S2 Fig.) for particular gene expression. The identified TFs were classified into 27 groups based on the TF families (S2 Fig.). Among the putative TF families, ERF, MIKC-MADS, C2H2, BBR-BPC, MYB, Dof, LBD, CPP and AP2 families are assumed highly contribute in regulating RNAi genes for specific gene expression in wheat (*T. aestivum*) in diverse conditions as these families bind with higher number of RNAi genes in wheat. Out of these top 7 TF families, first three one, ERF possesses 157(41%), MIKC-MADS possesses 38(10%) and C2H2 retains 34(9%) TFs, which are in leading position accounted for nearly 61% of the total identified TFs (S3 Table). This, implied that

those set of TF families could be high potential in regulating gene expression in terms of predicted RNAi genes in wheat (*T.aestivum*).

Furthermore, a regulatory gene network was created using Cytoscape to analyze in-depth relationship among the TFs and RNAi genes in wheat(*T.aestivum*).This suggests that there is some diverse relationships among the predicted groups of TFs and predicted RNAi genes (Fig.10) in wheat. The ERF, top most TF family, largely binds to the genes TaDCL3a (18 TFs), TaAGO7a (17 TFs), TaDCL1a (15 TFs), TaAGO9b (15 TFs), TaRDR2b (16 TFs), TaAGO2a (13 TFs), and TaAGO10b with 13 TFs (Fig.10 and S2 Fig.). MIKC-MADS, second top most TF family, also binds with the higher number of TaRNAi genes among those TaAGO1f(6 TFs) ,TaAGO6c(3 TFs), TaAGO6d(3 TFs), TaRDR1e(4 TFs) and TaRDR3(3 TFs) followed by C2H2, BBR-BPC and MYC. It is also observable that other than these leading TF families in wheat, the two families, viz., AP2 (one TF with each TaDCL3a, TaAGO1f, TaAGO6b, TaAGO6d, TaAGO6e TaAGO9c, TaAGO10c, TaRDR1f, TaRDR2d and TaRDR3) and bZIP (4 TFs with TaAGO2b) are predicted to bind with some specific *cis*-acting elements in the RNAi genes to response some physiological and biochemical stimuli in wheat for regulating gene expression (S2 Fig.). Some previous study suggested that a few TFs regulate the expression of a number of important genes for some particular physiological and biochemical reasons[83,84]. Earlier investigation also reported that both AP2 and bZIP TF families contribute in gene regulation for development and resistance response to various environmental stressors[83]. From the analysis results, it is however observed that AP2 TF family only interacts with TaAGO2b (S2 Fig.) in wheat that might be predicted that the DNA-binding domain of this TF family is responsible to play for regulating gene in a specific situation.

Fig.10. Regulatory gene network among the TF families and the predicted RNAi genes in wheat (*T.aestivum*).

The nodes of the network were colored based on RNAi genes and TFs. TaDCL, TaAGO and TaRDR genes were represented by pink, green, and yellow color nodes respectively and the TFs were represented by paste color nodes.

Different node symbols were used for different TF families.

Additionally, four hub TF families were selected based on the node degree criterion that possessed three or more interaction with the predicted wheat (*T.aestivum*) RNAi genes. Among them 16 TFs belong to ERF family, one TF from MIKC-MADS and C2H2, two TFs from LBD family (Fig.11E). The ERF family with accession no. Traes_2AL_E5A9615E2 regulates all 10 and a TF of ERF family with accession no Traes_5BL_F5D379AFC

regulates minimum 5 hub wheat (*T. aestivum*) RNAi genes in the network followed by the two TF of family LBD with accession no.TRAES3BF084400010CFD_g and Traes_1DS_BB8508CC6 that regulate 8 and 5 RNAi genes respectively (Fig. 11E). The TF family C2H2 and MIKC-MADS each regulates 4 and 3 RNAi genes respectively in wheat (*T.aestivum*) in the network (Fig. 11E). It is however observed from the Fig. 11E that TFs from ERF family are largely bind to the hub wheat RNAi genes that suggests that TFs from this family contain high tendency to bind with RNAi genes in particular gene expression in wheat. From earlier investigation, it is also broadly well-known that transcription factors of ERF family play a significant role in regulating the growth, evolution, and response of plants to various stresses as a signal transmission pathway in plants [85]. This TF family contains DRE-binding proteins (DREBs) that stimulate the expression of abiotic stress-responsive genes through specific binding to the dehydration-responsive element/C-repeat (DRE/CRT) *cis*-acting component in their promoters[86]. Transcription factors of ethylene-responsive factor (ERF) family also possess evolving role as the plant TF for coordinating wound defense responses and repair[87].

Furthermore, from this analysis (Fig.11E), it is noticeable that, all hub TF families are associated with 10 RNA silencing machinery gens in wheat (*T. aestivum*) viz., two RNAi genes from TaDCL family (TaDCL3a and TaDCL1a), 6 RNAi genes from TaAGO family (TaAGO2a/2b, TaAGO6d, TaAGO7a, TaAGO8, and TaAGO9), and two from TaRDR family (TaRDR2b and TaRDR2c). Interestingly, only TaDCL3a member from DCL family is regulated by all four TF families (Fig. 11E) that is assumed to possess high biological importance in regulating gene expression in wheat (*T.aestivum*) plants.

Fig.11. RNAi gene mediated sub-networks with different TF families. (A) ERF, (B) MIKC-MADS, (C) C2H2, (D) BBR-BPC (E) MYB (F) Dof TF family. (G) Sub-network among the hub TF families those regulate three or more RNAi genes in wheat (*T.aestivum*).

3.8 *Cis*-acting regulatory element analysis in wheat (*T.aestivum*)

RNAi gene families

Cis-regulatory elements (CREs) act as enhancers and promoters for particular gene expression in living organisms. CREs take part in indispensable activities for development and physiology through the regulation of gene expression in plants[88]. Generally these *cis*-acting components are consists of non-coding DNA possessing binding regions

for TFs and/or other regulatory molecules for triggering gene transcription[88,89]. In this work, CREs analyses were also carried out to study their functional divergence in wheat (*T.aestivum*) plant for the identified RNAi genes using PlantCARE database. It is obvious from the Fig.12 that most of the *cis*-regulatory sequences/motifs are involved in light responses (LR) which are diversely possessed by most of the RNAi gene families' promoter sites in wheat (*T.aestivum*). Plants photosynthesis phenomenon is greatly connected to light response that mostly takes place in leaves. Photosynthesis is also the key physiological parameter in wheat plants that relates ultimately in many aspects to increase grain quality and crop productivity[90]. Increase photosynthesis rate can utilize the solar radiation properly which leads to create early flowering time because flowering signals are produced in leaves [91,92]. Therefore, the identified CREs related to LR is assumed to retain direct connections for high photosynthesis level in wheat plant leaves. Amongst the LR CREs the ATCT-motif, ATC-motif, Box-4, AE-box, G-box, I-box, GAT-motif, GT1-motif were shared by the most of the RNAi genes in wheat (*T.aestivum*) (Fig 12). The TC-rich repeats (CRE involved in defense and stress responsiveness), MBS (MYB binding site involved in drought inducibility), and LTR elements (CRE involved in low-temperature responsiveness) were commonly found as stress responsive CREs among the RNAi gene families in wheat (*T.aestivum*).

Fig 12. The CREs of the promoter sites in the predicted RNAi gene families of DCL, AGO and RDR in wheat (*T.aestivum*). The dark color represents the existence of a *cis*-acting element corresponding to each RNAi gene member.

It is well-studied that various plant hormones or phytohormones are essential for plant growth and developmental timing[93]. More than a dozen common CREs is related to wheat plant hormonal activities was also identified from the PlantCARE database analysis. The ABRE (*cis*-acting element involved in the abscisic acid responsiveness), AuxRR-core (*cis*-acting regulatory element involved in auxin responsiveness), GC-motif (enhancer-like element involved in anoxic specific inducibility), GARE-motif (gibberellin-responsive element), O2-site (*cis*-acting regulatory element involved in zein metabolism regulation), P-box (gibberellin-responsive element), TATC-box (*cis*-acting element involved in gibberellin-responsiveness), TCA-element (*cis*-acting element involved in salicylic acid responsiveness) and TGA-element (auxin-responsive element), RY-element (*cis*-acting regulatory element involved in seed-specific regulation), MBSI(MYB binding site involved in flavonoid biosynthetic genes regulation), GCN4_motif(*cis*-regulatory element involved in endosperm expression), CGTCA-motif and TGACG-motif (*cis*-acting regulatory element involved in the MeJA-responsiveness), CAT-box and NON-box(*cis*-acting regulatory

element related to meristem specific activation) were predicted as the hormone responsive CREs in wheat (*T.aestivum*) (Fig.12). These however are retained by the three TaDCL, TaAGO and TaRDR protein families which are termed as phytohormones CREs. Previous studies also suggested that five most important plant hormones: auxin, gibberellin, cytokinin, ethylene, and abscisic acid act collectively or individually to effect plant growth and development [93–97].

Besides, there are some other CREs were also found and identified in the promoter and enhancer sites of the RNAi genes in wheat (*T.aestivum*) assumed to be responsible for a few momentous activities for controlling gene expression. The AT-rich element (binding site of AT-rich DNA binding protein (ATBP-1), AT-rich sequence (element for maximal elicitor-mediated activation), CAAT-box (common *cis*-acting element in promoter and enhancer regions), CCAAT-box (MYBHv1 binding site), TATA-box (core promoter element around-30 of transcription start), circadian (*cis*-acting regulatory element involved in circadian control) etc. were recognized as other *cis*-acting regulatory element shared by RNAi genes in wheat (*T.aestivum*) (Fig 12). In addition, some unknown CREs was discovered together with the reported *cis*-elements (S6, S7, and S8 Data). However, the identification and characterization of the predicted CREs associated with the suggested RNAi genes in wheat (*T.aestivum*) would however provide strong basis to plant molecular researchers for in-depth molecular investigation about their role in plant growth and development in diverse environmental conditions as whole.

3.9 *In silico* expressed sequence tag (EST) analysis for RNAi genes in wheat (*T.aestivum*)

In order to obtain further insights about the genomic information in terms of gene expression at various tissues or organs with varied conditions, computational expressed sequence tag (EST) data analysis was carried out for all 62 RNAi gene families in wheat (*T.aestivum*) using web-based PlantGDB database. This analysis results discovered that numerous RNA silencing machinery genes of DCL, AGO and RDR family showed their expression in several different important tissues and organs in wheat (*T.aestivum*). Also the expression study of the RNAi gene families of DCL, AGO and RDR groups in several plant species was investigated and thus the analyses results also suggested that RNAi gene families showed considerable and significant expression eminence in root, leaf, flower, seed, endosperm, spike and other important organs or tissues [1,6,7,10,35–38,98]. It is however observed from the Fig.13

that all most all members of TaDCL, TaAGO and TaRDR gene families exhibited their expression at least in one tissue or organ while TaAGO5c, TaAGO6e and TaAOG7b did not show any expression in any tissue or organ in wheat (*T.aestivum*). Nearly 50% RNAi genes of all 62, expressed in root, spike, anther, heads, shoots and endosperm followed by floret, leaf and seed (Fig.13) that certainly implied that these tissues and organs provide major contribution for improved wheat grain formation resulting in increased wheat yield. Among the DCL gene members, TaDCL4 exhibited on in root, wheat head, and seedling shoot specific expression. All TaDCL3 gene sub members displayed expression in root, spike, anther, and endosperm only except TaDCL3d that was in endosperm specific (Fig.13). On the other hand, TaDCL1a and TaDCL1b had no root, leaf, wheat head/shoot and floret/flower specific expression (Fig.13) but interestingly they had only endosperm-embryo, seed, spike specific expression characteristics which indicates that these two gene members of TaDCL family might play role in formation of healthy and sufficient number of grains in wheat plant (*T.aestivum*).

Among the all TaAGO and TaRDR gene families, TaAGO1e, TaAGO3, TaAGO2a/2b, TaAGO10d/10c, all 7-sub members of TaAGO5 family, TaRDR2d, TaRDR4 and TaRDR6a/6b had no root specific expression (Fig.13). Though the TaAGO2a/2b genes expressed in wheat endosperm, head, and spike (Fig.13) implied that these two genes contribute in grain development. But surprisingly three sub members of TaAGO1 (TaAGO1i/j/k) as well as all three sub families of TaAGO4 and TaAGO9 showed expression in all mentioned tissues and organs in the expression map except pistil, spikelet and sheath (Fig.13). This indicates that these 6 members of TaAGO4 and TaAGO9 might have roles in wheat plant growth and development. The gene sub-family TaAGO3 had only endosperm-embryo specific expression. The gene sub families TaAGO6a/6b/6c had shoot anther, spike, and root specific expression whereas the two members TaAGO6c/6d showed leaf and only TaAGO6d showed endosperm specific expression. TaAGO7a also showed root, wheat head/shoot and endosperm specific expression (Fig.13). Among the four sub families of TaAGO10 gene, the TaAGO10a/10b retained expression in all tissues and organs of wheat except leaf, sheath, and floret (Fig.13). Earlier wet lab study reported that RNAi genes of AGO family are tend to show diverse expression intensity in leaf, stem, flower in *Brassica napus*, pepper (*Capsicum Annuum*) [1,39].

Furthermore, EST analysis for all sub members of TaRDR1 showed their expression in flower but had no expression in any tissues and organs in wheat (*T.aestivum*) (Fig.12). It that suggests that these sub gene families are tend to responsible for expression in flower like two sub families TaRDR6a/6b (Fig.13). Previous study on pepper (*Capsicum*

Annuum) suggested that large number of RDR gene families showed flower specific expression[1]. In contrast, four sub members of TaRDR2 family did not provide any expression in flower. However, all members of TaRDR2 showed expression all most in all mentioned important tissues and organs with additional expression of TaRDR2b in pistil in wheat (*T.aestivum*) but TaRDR2d expressed in only spike (Fig.12). It can be said from this EST analysis of all 62 RNAi genes in wheat (*T.aestivum*) that genes those expressed in root and leaf might play role in resistance to abiotic stresses such as drought and water logging etc. Genes expressed in leaves may contribute in promoting photosynthesis rate to produce sufficient energy for the wheat plants. Increasing photosynthesis level stimulate the flowering time of the crop as the respective signals are produced in leaves[91,92].Seed and endosperm specific expression of the RNAi genes demonstrates that they might have significant role in maintaining quality seed formation.

Fig.13. The *in silico* expressed sequence tag (EST) analysis of the identified RNAi genes in *T.aestivum*. The green color represents the existence of expression of the corresponding RNAi genes in wheat (*T. aestivum*).

3.10 Expression analysis of TaDCL Genes in leaves, roots and against drought Stress

To obtain evidence for the likely functions of the TaDCL genes, their expression profile in leaves and roots of two-week aged plants were analyzed by qRT-PCR. The results showed that two out of seven TaDCL genes (TaDCL3a and TaDCL3b) expressed significantly in root and had no expression in leaves indicating root specific function of TaDCL3a and TaDCL3b gene families that validates the respective EST analysis too (Fig. 14A and Fig.13). To explore the role of TaDCL genes in response to drought stress tolerance, the expression patterns of TaDCLs were measured at 3-days post treatment against drought. The results showed that drought treatment positively induced the expression of all the seven TaDCL candidate genes. Among them, TaDCL3b and TaDCL4 showed the significant change in expression in response to drought signifying these genes might play a vital role against drought stress in wheat(*T.aestivum*) (Fig. 14B). The results from this analysis however can enrich our understanding of DCL genes in wheat and constitute a base for advance study.

Fig.14 qRT-PCR expression of TaDCL genes in (A) leaves and roots and (B) under drought stress. *T.aestivum*

18s was considered as the reference gene, and three biological replicates were accomplished for the experiments.

Error bars indicate the standard error. Asterisks indicate the significance difference between control and treatment at
p-value < 0.05

4. Conclusion

Wheat as a cereal crop is consumed by a large community of the people across the globe for the key source of carbohydrates after rice. Due to of its high protein content, this crop is also used for producing many important human foods for each day. Genome-wide identification, characterization, and diverse functional analyses of the RNA silencing machinery gene families were top most priority because these set of gene families carry great significance for regulating gene expression by adopting gene-silencing mechanism in plants by restricting the transcript buildup in cells. In this work, a number of bioinformatic techniques were applied for comprehensive identification, characterization, as well as *in-silico* biological and molecular functional activities were predicted in wheat (*T.aestivum*). Analyses results suggest that the wheat genome possessed 7 DCL, 39 AGO and 16 RDR RNAi genes. Initial analysis provided some standard genomic and physicochemical info for the identified genes and corresponding proteins. The phylogenetic analysis provided that all subfamilies of these three gene sets maintain their evolutionary relationships similar to their rice and *Arabidopsis* homologs. First subfamily of DCL, AGO and RDR possessed the multiple copies of genes higher than that of corresponding rice and *Arabidopsis*. Conserved functional domain and motif structure analyses suggested that these genes also contained consistent domain and motif structure similar to those of rice and *Arabidopsis* RNAi genes. Besides, GESA and SCL analyses established the insights into the functional routes and their involvement scenario in RNAi process in wheat (*T.aestivum*). We discovered the most important regulatory relationship network amongst the popular TF families and the putative RNAi gene families. Promoter and enhancer as CREs was analyzed those also attached with TFs to control the gene expression for some specific reasons during the wheat plant's growth and development. CRE and EST analyses hence suggests that the identified TaDCL, TaAGO and TaRDR gene families have diversifying participation for plant growth, development, maintenance the grain quality and increased crop production in wheat(*T.aestivum*). Wet lab expression study demonstrated that only TaDCL3a and TaDCL3b had root specific expression that validates EST analysis as well. Finally, all TaDCL transcripts induced up-regulation in drought stress condition with high expression in TaDCL3b and TaDCL4. The results of this study thus provide the important indication of evolutionary resemblances of DCL, AGO and RDR genes with their rice and *Arabidopsis* counterpart. Obtained results would

however provide valuable sources for further biological and molecular justification and implementation to draw more specific conclusion regarding any particular gene(s) and its domain of activity against different biotic and abiotic stresses as well as growth and improvement in wheat plants for more improvement of this potential cereal crop across the world.

Acknowledgments

We are very grateful to Bangladesh Agricultural Research Institute and Bioinformatics Lab, Department of Statistics, University of Rajshahi to provide us opportunity to conduct this research using lab facilities. MA of icddr,b thanks the Governments of Bangladesh, Canada, Sweden, and the United Kingdom for core/unrestricted support.

Abbreviations

RNA, Ribonucleic Acid; RNAi, RNA interference; siRNA, Short interfering RNA; miRNA, micro RNA; ssRNA, Single-stranded RNA; dsRNA, Double-stranded RNA; DCL, Dicer-like; AGO, Argonaute; RDR, RNA-dependent RNA polymerase; RISC, RNA-induced silencing complexes; DRM, Double-stranded RNA-binding motif; BLAST, Basic Local Alignment Search Tool; MEGA, Molecular Evolutionary Genetics Analysis; GO, Gene Ontology; SCL, Sub Cellular Location; NJ, neighbor-joining; PAZ, Piwi Argonaut and Zwillig; DUF, Domain of unknown function; GO, Gene Ontology; GSDS, Gene Structure Display Server; HMM, Hidden Markov Model; MEME, Multiple EM for Motif Elicitation; PTGS, post-transcriptional gene silencing; TF, transcription factor, CRE, *cis*-regulatory elements, qRT-PCR, quantitative reverse transcription polymerase chain reaction

References

1. Qin L, Mo N, Muhammad T, Liang Y. Genome-wide analysis of DCL, AGO, and RDR gene families in pepper (*Capsicum Annuum* L.). *International Journal of Molecular Sciences*. 2018. doi:10.3390/ijms19041038
2. Fernandes-Brum CN, Rezende PM, Ribeiro THC, De Oliveira RR, De Sousa Cardoso TC, Do Amaral LR, et al. A genome-wide analysis of the RNA-guided silencing pathway in coffee reveals insights into its regulatory mechanisms. *PLoS ONE*. 2017. doi:10.1371/journal.pone.0176333

870 3. Vaistij FE, Jones L. Compromised virus-induced gene silencing in rdr6-deficient plants. *Plant Physiology*.
871 2009. doi:10.1104/pp.108.132688

872 4. Tijsterman M, Ketting RF, Plasterk RHA. The Genetics of RNA Silencing. *Annual Review of Genetics*.
873 2002. doi:10.1146/annurev.genet.36.043002.091619

874 5. Ullu E, Tschudi C, Chakraborty T. RNA interference in protozoan parasites. *Cellular Microbiology*. 2004.
875 doi:10.1111/j.1462-5822.2004.00399.x

876 6. Vaucheret H. Plant ARGONAUTES. *Trends in Plant Science*. 2008. doi:10.1016/j.tplants.2008.04.007

877 7. Kapoor M, Arora R, Lama T, Nijhawan A, Khurana JP, Tyagi AK, et al. Genome-wide identification,
878 organization and phylogenetic analysis of Dicer-like, Argonaute and RNA-dependent RNA Polymerase gene
879 families and their expression analysis during reproductive development and stress in rice. *BMC Genomics*.
880 2008;9: 1–17. doi:10.1186/1471-2164-9-451

881 8. Cao J-Y, Xu Y-P, Li W, Li S-S, Rahman H, Cai X-Z. Genome-Wide Identification of Dicer-Like,
882 Argonaute, and RNA-Dependent RNA Polymerase Gene Families in Brassica Species and Functional
883 Analyses of Their Arabidopsis Homologs in Resistance to *Sclerotinia sclerotiorum*. *Frontiers in Plant*
884 *Science*. 2016;7: 1–17. doi:10.3389/fpls.2016.01614

885 9. Fahlgren N, Montgomery TA, Howell MD, Allen E, Dvorak SK, Alexander AL, et al. Regulation of
886 AUXIN RESPONSE FACTOR3 by TAS3 ta-siRNA Affects Developmental Timing and Patterning in
887 Arabidopsis. *Current Biology*. 2006. doi:10.1016/j.cub.2006.03.065

888 10. Qian Y, Cheng Y, Cheng X, Jiang H, Zhu S, Cheng B. Identification and characterization of Dicer-like,
889 Argonaute and RNA-dependent RNA polymerase gene families in maize. *Plant Cell Reports*. 2011;30:
890 1347–1363. doi:10.1007/s00299-011-1046-6

891 11. Wang T, You L, Li R, Fu DQ, Zhu BZ, Luo YB, et al. Cloning, identification, and expression analysis of a
892 Dicer-Like gene family from *Solanum lycopersicum*. *Biologia Plantarum*. 2016. doi:10.1007/s10535-016-
893 0620-8

894 12. Chen X. Small RNAs and Their Roles in Plant Development. *Annual Review of Cell and Developmental*
895 *Biology*. 2009. doi:10.1146/annurev.cellbio.042308.113417

896 13. Borges F, Martienssen RA. The expanding world of small RNAs in plants. *Nature Reviews Molecular Cell*
897 *Biology*. 2015. doi:10.1038/nrm4085

- 898 14. Carrington JC, Ambros V. Role of MicroRNAs in Plant and Animal Development. *Science*. 2003;301: 336
899 LP – 338. doi:10.1126/science.1085242
- 900 15. Lai EC. microRNAs: Runt of the Genome Assert Themselves. *Current Biology*. 2003.
901 doi:10.1016/j.cub.2003.11.017
- 902 16. Finnegan EJ, Matzke MA. The small RNA world. *Journal of Cell Science*. 2003. doi:10.1242/jcs.00838
- 903 17. Finnegan EJ, Matzke MA. The small RNA world. *Journal of Cell Science*. 2003;116: 4689 LP – 4693.
904 doi:10.1242/jcs.00838
- 905 18. Bologna NG, Voinnet O. The Diversity, Biogenesis, and Activities of Endogenous Silencing Small RNAs in
906 *Arabidopsis*. *Annual Review of Plant Biology*. 2014. doi:10.1146/annurev-arplant-050213-035728
- 907 19. Sijen T, Fleenor J, Simmer F, Thijssen KL, Parrish S, Timmons L, et al. On the role of RNA amplification
908 in dsRNA-triggered gene silencing. *Cell*. 2001. doi:10.1016/S0092-8674(01)00576-1
- 909 20. Liu Q, Feng Y, Zhu Z. Dicer-like (DCL) proteins in plants. *Functional and Integrative Genomics*. 2009.
910 doi:10.1007/s10142-009-0111-5
- 911 21. Zhu H, Zhou Y, Castillo-González C, Lu A, Ge C, Zhao YT, et al. Bidirectional processing of pri-miRNAs
912 with branched terminal loops by Arabidopsis Dicer-like1. *Nature Structural and Molecular Biology*. 2013.
913 doi:10.1038/nsmb.2646
- 914 22. Qin H, Chen F, Huan X, Machida S, Song J, Yuan YA. Structure of the Arabidopsis thaliana DCL4
915 DUF283 domain reveals a noncanonical double-stranded RNA-binding fold for protein - Protein interaction.
916 *RNA*. 2010. doi:10.1261/rna.1965310
- 917 23. Hammond SM. Dicing and slicing: The core machinery of the RNA interference pathway. *FEBS Letters*.
918 2005. doi:10.1016/j.febslet.2005.08.079
- 919 24. Meng F, Jia H, Ling N, Xue Y, Liu H, Wang K, et al. Cloning and characterization of two Argonaute genes
920 in wheat (*Triticum aestivum* L.). *BMC Plant Biology*. 2013;13: 1–10. doi:10.1186/1471-2229-13-18
- 921 25. Hutvagner G, Simard MJ. Argonaute proteins: Key players in RNA silencing. *Nature Reviews Molecular*
922 *Cell Biology*. 2008. doi:10.1038/nrm2321
- 923 26. Höck J, Meister G. The Argonaute protein family. *Genome Biology*. 2008;9. doi:10.1186/gb-2008-9-2-210
- 924 27. Faehnle CR, Joshua-Tor L. Argonautes confront new small RNAs. *Current Opinion in Chemical Biology*.
925 2007. doi:10.1016/j.cbpa.2007.08.032

- 926 28. Girard A, Sachidanandam R, Hannon GJ, Carmell MA. A germline-specific class of small RNAs binds
927 mammalian Piwi proteins. *Nature*. 2006. doi:10.1038/nature04917
- 928 29. Yigit E, Batista PJ, Bei Y, Pang KM, Chen CCG, Tolia NH, et al. Analysis of the *C. elegans* Argonaute
929 Family Reveals that Distinct Argonautes Act Sequentially during RNAi. *Cell*. 2006.
930 doi:10.1016/j.cell.2006.09.033
- 931 30. Nonomura KI, Morohoshi A, Nakano M, Eiguchi M, Miyao A, Hirochika H, et al. A germ cell-specific gene
932 of the ARGONAUTE family is essential for the progression of premeiotic mitosis and meiosis during
933 sporogenesis in rice. *Plant Cell*. 2007. doi:10.1105/tpc.107.053199
- 934 31. Djupedal I, Ekwall K. Epigenetics: Heterochromatin meets RNAi. *Cell Research*. 2009.
935 doi:10.1038/cr.2009.13
- 936 32. Schiebel W, Pélissier T, Riedel L, Thalmeir S, Schiebel R, Kempe D, et al. Isolation of an RNA-directed
937 RNA polymerasespecific cDNA clone from tomato. *Plant Cell*. 1998. doi:10.1105/tpc.10.12.2087
- 938 33. Iyer LM, Koonin E V., Aravind L. Evolutionary connection between the catalytic subunits of DNA-
939 dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA
940 polymerases. *BMC Structural Biology*. 2003. doi:10.1186/1472-6807-3-1
- 941 34. Wassenegger M, Krczal G. Nomenclature and functions of RNA-directed RNA polymerases. *Trends in*
942 *Plant Science*. 2006. doi:10.1016/j.tplants.2006.01.003
- 943 35. Nakasugi K, Crowhurst RN, Bally J, Wood CC, Hellens RP, Waterhouse PM. De Novo Transcriptome
944 Sequence Assembly and Analysis of RNA Silencing Genes of *Nicotiana benthamiana*. *PLoS ONE*. 2013;8.
945 doi:10.1371/journal.pone.0059534
- 946 36. Bai M, Yang GS, Chen WT, Mao ZC, Kang HX, Chen GH, et al. Genome-wide identification of Dicer-like,
947 Argonaute and RNA-dependent RNA polymerase gene families and their expression analyses in response to
948 viral infection and abiotic stresses in *Solanum lycopersicum*. *Gene*. 2012;501: 52–62.
949 doi:10.1016/j.gene.2012.02.009
- 950 37. Zhao H, Zhao K, Wang J, Chen X, Chen Z, Cai R, et al. Comprehensive Analysis of Dicer-Like, Argonaute,
951 and RNA-dependent RNA Polymerase Gene Families in Grapevine (*Vitis Vinifera*). *Journal of Plant*
952 *Growth Regulation*. 2015. doi:10.1007/s00344-014-9448-7
- 953 38. Gan D, Liang D, Wu J, Zhan M, Yang F, Xu W, et al. Genome-Wide Identification of the Dicer-Like,

954 Argonaute, and RNA-Dependent RNA Polymerase Gene Families in Cucumber (*Cucumis sativus* L.).
955 *Journal of Plant Growth Regulation*. 2016. doi:10.1007/s00344-015-9514-9

956 39. Zhao X, Zheng W, Zhong Z, Chen X, Wang A, Wang Z. Genome-wide analysis of RNA-interference
957 pathway in *Brassica napus*, and the expression profile of BnAGOs in response to *Sclerotinia sclerotiorum*
958 infection. *European Journal of Plant Pathology*. 2016;146: 565–579. doi:10.1007/s10658-016-0942-6

959 40. Sabbione A, Daurelio L, Vegetti A, Talón M, Tadeo F, Dotto M. Genome-wide analysis of AGO, DCL and
960 RDR gene families reveals RNA-directed DNA methylation is involved in fruit abscission in *Citrus sinensis*.
961 *BMC Plant Biology*. 2019. doi:10.1186/s12870-019-1998-1

962 41. Margis R, Fusaro AF, Smith NA, Curtin SJ, Watson JM, Finnegan EJ, et al. The evolution and
963 diversification of Dicers in plants. *FEBS Letters*. 2006. doi:10.1016/j.febslet.2006.03.072

964 42. Deleris A, Gallago-Bartolome J, Bao J, Kasschau KD, Carrington JC, Voinnet O. Hierarchical action and
965 inhibition of plant dicer-like proteins in antiviral defense. *Science*. 2006. doi:10.1126/science.1128214

966 43. Schmitz RJ, Hong L, Fitzpatrick KE, Amasino RM. DICER-LIKE 1 and DICER-LIKE 3 redundantly act to
967 promote flowering via repression of FLOWERING LOCUS C in *Arabidopsis thaliana*. *Genetics*. 2007.
968 doi:10.1534/genetics.107.070649

969 44. Ji L, Liu X, Yan J, Wang W, Yumul RE, Kim YJ, et al. ARGONAUTE10 and ARGONAUTE1 regulate the
970 termination of floral stem cells through two microRNAs in *Arabidopsis*. *PLoS Genetics*. 2011.
971 doi:10.1371/journal.pgen.1001358

972 45. Carbonell A, Fahlgren N, Garcia-Ruiz H, Gilbert KB, Montgomery TA, Nguyen T, et al. Functional analysis
973 of three *Arabidopsis* argonautes using slicer-defective mutants. *Plant Cell*. 2012.
974 doi:10.1105/tpc.112.099945

975 46. Fagard M, Boutet S, Morel JB, Bellini C, Vaucheret H. AGO1, QDE-2, and RDE-1 are related proteins
976 required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals.
977 *Proceedings of the National Academy of Sciences of the United States of America*. 2000.
978 doi:10.1073/pnas.200217597

979 47. Zilberman D, Cao X, Jacobsen SE. ARGONAUTE4 control of locus-specific siRNA accumulation and
980 DNA and histone methylation. *Science*. 2003. doi:10.1126/science.1079695

981 48. Hunter C, Sun H, Poethig RS. The *Arabidopsis* Heterochronic Gene ZIPPY Is an ARGONAUTE Family

982 Member. *Current Biology*. 2003. doi:10.1016/j.cub.2003.09.004

983 49. Moussian B, Schoof H, Haecker A, Jürgens G, Laux T. Role of the ZWILLE gene in the regulation of
984 central shoot meristem cell fate during Arabidopsis embryogenesis. *EMBO Journal*. 1998.
985 doi:10.1093/emboj/17.6.1799

986 50. Dalmay T, Hamilton A, Rudd S, Angell S, Baulcombe DC. An RNA-dependent RNA polymerase gene in
987 arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus.
988 *Cell*. 2000. doi:10.1016/S0092-8674(00)80864-8

989 51. Mourrain P, Béclin C, Elmayan T, Feuerbach F, Godon C, Morel JB, et al. Arabidopsis SGS2 and SGS3
990 genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell*. 2000.
991 doi:10.1016/S0092-8674(00)80863-6

992 52. Jovel J, Walker M, Sanfaçon H. Recovery of *Nicotiana benthamiana* Plants from a Necrotic Response
993 Induced by a Nepovirus Is Associated with RNA Silencing but Not with Reduced Virus Titer. *Journal of*
994 *Virology*. 2007. doi:10.1128/jvi.01192-07

995 53. Matzke M, Kanno T, Huettel B, Daxinger L, Matzke AJM. Targets of RNA-directed DNA methylation.
996 *Current Opinion in Plant Biology*. 2007. doi:10.1016/j.pbi.2007.06.007

997 54. Dorweiler JE, Carey CC, Kubo KM, Hollick JB, Kermicle JL, Chandler VL. Mediator of paramutation 1 is
998 required for establishment and maintenance of paramutation at multiple maize loci. *Plant Cell*. 2000.
999 doi:10.1105/tpc.12.11.2101

1000 55. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the sensitivity of progressive multiple
1001 sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.
1002 *Nucleic Acids Research*. 1994. doi:10.1093/nar/22.22.4673

1003 56. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger
1004 Datasets. *Molecular biology and evolution*. 2016;33: 1870–1874. doi:10.1093/molbev/msw054

1005 57. Felsenstein J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution*. 1985.
1006 doi:10.2307/2408678

1007 58. Bailey TL. Discovering Novel Sequence Motifs with MEME. *Current Protocols in Bioinformatics*. 2003.
1008 doi:10.1002/0471250953.bi0204s00

1009 59. Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, et al. PlantTFDB 4.0: Toward a central hub for

transcription factors and regulatory interactions in plants. *Nucleic Acids Research*. 2017.
doi:10.1093/nar/gkw982

60. Liu L, Zhang Z, Mei Q, Chen M. PSI: A Comprehensive and Integrative Approach for Accurate Plant Subcellular Localization Prediction. *PLoS ONE*. 2013. doi:10.1371/journal.pone.0075826

61. R Core Team. R: A language and environment for statistical computing. <http://www.R-project.org/>. In: R Foundation for Statistical Computing, Vienna, Austria. 2017.

62. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*. 2003.
doi:10.1101/gr.1239303

63. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods*. 2001. doi:10.1006/meth.2001.1262

64. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research*. 2017. doi:10.1093/nar/gkw1129

65. MacRae IJ, Doudna JA. Ribonuclease revisited: structural insights into ribonuclease III family enzymes. *Current Opinion in Structural Biology*. 2007. doi:10.1016/j.sbi.2006.12.002

66. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. *Nucleic Acids Research*. 2015. doi:10.1093/nar/gku1221

67. Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, Deweese-Scott C, Fong JH, et al. CDD: Specific functional annotation with the Conserved Domain Database. *Nucleic Acids Research*. 2009.
doi:10.1093/nar/gkn845

68. Rivas F V., Tolia NH, Song JJ, Aragon JP, Liu J, Hannon GJ, et al. Purified Argonaute2 and an siRNA form recombinant human RISC. *Nature Structural and Molecular Biology*. 2005. doi:10.1038/nsmb918

69. Baumberger N, Baulcombe DC. Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proceedings of the National Academy of Sciences*. 2005.
doi:10.1073/pnas.0505461102

70. Henderson IR, Zhang X, Lu C, Johnson L, Meyers BC, Green PJ, et al. Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nature Genetics*. 2006. doi:10.1038/ng1804

1038 71. Jinek M, Doudna JA. A three-dimensional view of the molecular machinery of RNA interference. *Nature*.
1039 2009. doi:10.1038/nature07755

1040 72. Simon B, Kirkpatrick JP, Eckhardt S, Reuter M, Rocha EA, Andrade-Navarro MA, et al. Recognition of 2'-
1041 o-methylated 3'-end of piRNA by the PAZ domain of a Piwi protein. *Structure*. 2011.
1042 doi:10.1016/j.str.2010.11.015

1043 73. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. *Nucleic Acids Research*. 2015.
1044 doi:10.1093/nar/gkv416

1045 74. Qin L, Mo N, Zhang Y, Muhammad T, Zhao G, Zhang Y, et al. CaRDR1, an RNA-dependent RNA
1046 polymerase plays a positive role in pepper resistance against TMV. *Frontiers in Plant Science*. 2017.
1047 doi:10.3389/fpls.2017.01068

1048 75. Vaucheret H, Béclin C, Fagard M. Post-transcriptional gene silencing in plants. *Journal of Cell Science*.
1049 2001.

1050 76. Kusaba M. RNA interference in crop plants. *Current Opinion in Biotechnology*. 2004.
1051 doi:10.1016/j.copbio.2004.02.004

1052 77. Agrawal N, Dasaradhi PVN, Mohammed A, Malhotra P, Bhatnagar RK, Mukherjee SK. RNA Interference:
1053 Biology, Mechanism, and Applications. *Microbiology and Molecular Biology Reviews*. 2003.
1054 doi:10.1128/MMBR.67.4.657-685.2003

1055 78. Lingel A, Izaurralde E. RNAi: Finding the elusive endonuclease. *RNA*. 2004. doi:10.1261/rna.7175704

1056 79. Lingel A, Simon B, Izaurralde E, Sattler M. Structure and nucleic-acid binding of the Drosophila Argonaute
1057 2 PAZ domain. *Nature*. 2003. doi:10.1038/nature02123

1058 80. Ma JB, Ye K, Patel DJ. Structural basis for overhang-specific small interfering RNA recognition by the PAZ
1059 domain. *Nature*. 2004. doi:10.1038/nature02519

1060 81. Kholodenko BN. Four-dimensional organization of protein kinase signaling cascades: the roles of diffusion,
1061 endocytosis and molecular motors. *Journal of Experimental Biology*. 2003. doi:10.1242/jeb.00298

1062 82. Ohlrogge JB, Kuhn DN, Stumpf PK. Subcellular localization of acyl carrier protein in leaf protoplasts of
1063 *Spinacia oleracea*. *Proceedings of the National Academy of Sciences of the United States of America*. 1979.
1064 doi:10.1073/pnas.76.3.1194

1065 83. Liu Q, Zhang G, Chen S. Structure and regulatory function of plant transcription factors. *Chinese Science*

1066 Bulletin. 2001. doi:10.1007/bf03187184

1067 84. Kasuga M, Liu Q, Miura S, Yamaguchi-Shinozaki K, Shinozaki K. Improving plant drought, salt, and
1068 freezing tolerance by gene transfer of a single stress-inducible transcription factor. *Nature Biotechnology*.
1069 1999. doi:10.1038/7036

1070 85. Najafi S, Sorkheh K, Nasernakhaei F. Characterization of the APETALA2/Ethylene-responsive factor
1071 (AP2/ERF) transcription factor family in sunflower. *Scientific Reports*. 2018. doi:10.1038/s41598-018-
1072 29526-z

1073 86. Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. AP2/ERF family transcription factors in plant abiotic stress
1074 responses. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. 2012.
1075 doi:10.1016/j.bbagrm.2011.08.004

1076 87. Heyman J, Canher B, Bisht A, Christiaens F, De Veylder L. Emerging role of the plant ERF transcription
1077 factors in coordinating wound defense responses and repair. *Journal of Cell Science*. 2018.
1078 doi:10.1242/jcs.208215

1079 88. Wittkopp PJ, Kalay G. Cis-regulatory elements: Molecular mechanisms and evolutionary processes
1080 underlying divergence. *Nature Reviews Genetics*. 2012. doi:10.1038/nrg3095

1081 89. Ong CT, Corces VG. Enhancer function: New insights into the regulation of tissue-specific gene expression.
1082 *Nature Reviews Genetics*. 2011. doi:10.1038/nrg2957

1083 90. Hidayati N, Triadiati, Anas I. Photosynthesis and Transpiration Rates of Rice Cultivated Under the System
1084 of Rice Intensification and the Effects on Growth and Yield. *HAYATI Journal of Biosciences*. 2016.
1085 doi:10.1016/j.hjb.2016.06.002

1086 91. Karki S, Rizal G, Quick WP. Improvement of photosynthesis in rice (*Oryza sativa* L.) by inserting the C4
1087 pathway. *Rice*. 2013. doi:10.1186/1939-8433-6-28

1088 92. Lee YS, An G. Regulation of flowering time in rice. *Journal of Plant Biology*. 2015. doi:10.1007/s12374-
1089 015-0425-x

1090 93. Gray WM. Hormonal regulation of plant growth and development. *PLoS Biology*. 2004.
1091 doi:10.1371/journal.pbio.0020311

1092 94. Goda H, Sawa S, Asami T, Fujioka S, Shimada Y, Yoshida S. Comprehensive comparison of auxin-
1093 regulated and brassinosteroid-regulated genes in arabidopsis. *Plant Physiology*. 2004.

- 1094 doi:10.1104/pp.103.034736
- 1095 95. Peng ZY, Zhou X, Li L, Yu X, Li H, Jiang Z, et al. Arabidopsis hormone database: A comprehensive
- 1096 genetic and phenotypic information database for plant hormone research in Arabidopsis. Nucleic Acids
- 1097 Research. 2009. doi:10.1093/nar/gkn873
- 1098 96. Liu Y, Guo J, Zhu H. Gene prediction in metagenomic fragments based on the SVM algorithm. Proceedings
- 1099 - 2011 4th International Conference on Biomedical Engineering and Informatics, BMEI 2011. 2011.
- 1100 doi:10.1109/BMEI.2011.6098588
- 1101 97. Ogawa M, Hanada A, Yamauchi Y, Kuwahara A, Kamiya Y, Yamaguchi S. Gibberellin biosynthesis and
- 1102 response during Arabidopsis seed germination. Plant Cell. 2003. doi:10.1105/tpc.011650
- 1103 98. Mirzaei K, Bahramnejad B, Shamsifard MH, Zamani W. In silico identification, phylogenetic and
- 1104 bioinformatic analysis of argonaute genes in plants. International Journal of Genomics. 2014.
- 1105 doi:10.1155/2014/967461

1106 Author's contributions

1107 ZA and MM designed the research. ZA analyzed the data using bioinformatics and statistical tools as well as drafted

1108 the manuscript. ZA and HR carried out the gene expression study. MAA performed GO and SCL analysis. MA,

1109 MJA, and MPM attended at the meeting regarding this manuscript. All authors read and approved the final version

1110 of the manuscript.

1111

1112 Supporting information

1113 **S1 Table. Primer sequences of seven TaDCL genes for qRT-PCR analysis**

1114 **S2 Table. Percentage of three groups of RNA silencing genes involved in different cellular location in wheat**

1115 **(*T.aestivum*)**

1116 **S3 Table. Distribution of TF families those regulating RNAi genes in wheat (*T.aestivum*)**

1117 **S1 Data.** Protein sequences of the identified DCL genes in wheat(*Triticum aestivum*).

1118 **S2 Data.** Protein sequences of the identified AGO genes in wheat(*Triticum aestivum*).

1119 **S3 Data.** Protein sequences of the identified RDR genes in wheat(*Triticum aestivum*).

1120 **S4 Data.** GO enrichment analysis result for reported RNAi genes.

1121 **S5 Data.** List of transcript factors and their families regulating the predicted RNAi genes.

1122 **S6 Data.** List of *cis*-regulatory element associated with the TaDCL protein families

1123 **S7 Data.** List of *cis*-regulatory element associated with the TaAGO protein families

1124 **S8 Data.** List of *cis*-regulatory element associated with the TaRDR protein families

1125 **S1 Fig.** GO slim hierarchy tree analysis of the predicted RNAi genes in wheat (*T.aestivum*) was projected using

1126 QuickGO web-based bioinformatic tool.(A) Biological process(BP) (B) Molecular Function(MF) and (C) Cellular

1127 Component(CC). GO IDs are highlighted in upper part of the boxes shaded blue color and their corresponding

1128 functional terms are mentioned in the middle of each box. Boxes with light yellow color are statistically significant

1129 GO functions whereas the boxes with white color are insignificant GO.

1130 **S2 Fig.** Distribution of TF families corresponding to each RNAi gene member. Rows of the figure represent the

1131 predicted RNAi genes and the columns represent the TF families. The number inside the cell indicates the number

1132 of TFs of the corresponding family involved in regulating the RNAi genes in wheat(*T.aestivum*).

Part-I: Genomic data collection

BLAST search results for *T.aestivum*
DCL, AGO, RDR genes against protein
sequences of *Arabidopsis thaliana*
AtDCL, AtAGO and AtRDR genes

First, all protein sequences of *T.aestivum* were downloaded from Phytozome database based on similarity score(>50.0) using BLAST against the protein sequences of *Arabidopsis thaliana*

BLAST Results

- BLAST Inputs

Query: your.seq (992 letters)

Target *Triticum aestivum* v2.2 proteome (293053 sequences, 89057418 total letters)

Program BLASTP 2.2.26+

Hits Found 84

Download results 12

Views	Define	Score	E	Query View
	Traces_30_AAACDCDF6F.1 (1 of 11) PTHG23079 SF13 - RNA...	60.1	1.4E-9	
	Traces_48L_746C0C0EA1.1 (1 of 49) 2.7.7.48 - RNA-directed	61.2	2.6E-9	
	Traces_58L_08B0CA49D0.1 (1 of 9) PTHG23079 SF9 - RNA-d...	62.0	4.7E-9	
	Traces_58L_30AFA1543B.8	61.2	7.7E-9	
	Traces_58L_30AFA1543B.7	61.2	7.7E-9	
	Traces_GAL_13BC93FE04.9	57.0	1.7E-7	
	Traces_GAL_13BC93FE04.8	56.2	5.5E-7	
	Traces_GAL_13BC93FE04.6	55.8	8.1E-7	
	Traces_28L_439F77709.1 (1 of 49) 2.7.7.48 - RNA-directed	51.2	2.1E-6	
	Traces_AAL_1CDBA1961.1 (1 of 3) PTHG23079 SF1 - PROT...	49.7	5.5E-6	
	Traces_30S_581EEFEP.1 (1 of 11) PTHG23079 SF13 - RNA...	48.1	1.1E-5	

bioRxiv preprint doi: <https://doi.org/10.1101/2020.05.21.108100>; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

for pre-bioinformatic analyses

Part-III

Comprehensive bioinformatic analyses on the finally identified 62 RNAi genes in wheat (*T.aestivum*)

- Prediction of DDH/H catalytic regions in the PIWI domain of TaAGO protein families using MEGA and ClustalW method
- Exon-Intron structure analysis of the wheat RNAi genes using GSDS 2.0
- Gene Set Enrichment Analysis (GESA)/GO analysis using QuickGO
- Sub Cellular Localization(SCL) analysis using PSI
- Prediction of transcription factor (TF) families from plantTFDB and construction of gene-RNAi network using Cytoscape
- Prediction of Cis-regulatory elements (CREs) using PlantCARE database
- Expression sequence tag (EST) analysis using PlantGDB web-based tool

Part-IV

Expression Analysis

- Expression analysis of the 7 TaDCL genes in roots and leaves using qRT-PCR
- Expression analysis of the 7 TaDCLs against drought stress using qRT-PCR

Figure

	1080	1100	1120	1140	1180	1200
AtAGO1	: TIIIFGADVTHHPGEDS-SFSIAAVVASQDWPEITKYAGLVCAQFHRCELIQDLFKWKDPQKGV-VTGGMIKELLIAFRSTG-HKPLRIIFRDGVSEGQFYQVLLYELDAIRKACASLEA-----GYQPP	880				
AtAGO2	: VMFIGADVNHFAARDKM-SFSIVAVVGTNLWPEANRYAARVIAQIRKEEIQGFC-----DACLELVKAHVQATG-KRPNKIVIFRDGVSDAQFDMVLNVELLDVK-LTFERN-----GYNDK	849				
AtAGO3	: VMFIGADVNHFAAHDNM-SFSIVAVVGTNLWPEANRYAARVIAQIRKEEIQGFG-----ETCNELIEAHSQAPE-KRPNKIVIFRDGVSDGQFDMVLNVELCNVK-DVFAKV-----GYNPQ	1028				
AtAGO4	: TIIILGNDVSHGSPGQSD-VFSIAAVVSSREWPLISKYRASVRTQSKAEMIESLVK-KNGTE-----DDGIKELLVDFTYSSNKRKPEHIIIFRDGVSESOQFNOVLNIELDOIIEACKLL-----DANWNP	774				
AtAGO5	: TIIIMGADVTHHPGEDS-SFSIAAVVASNDWPEINKYRGLVSAQFHRCEIIQDLKYLVQDPQKGL-VHSGLIREFHIAFRRTG-QIPCRIIFRDGVSEGOFSQVLLHENTAIRKACNSLOE-----NYVPR	839				
AtAGO6	: TIIILGNDVSHGPPGRAD-VFSVAAVVGSKCWPLISRYRAAVRTQSRLEMIIDSLFQPIENTEKG---DNGIMNELFVEFYRTSRARKPKQIIIFRDGVSESOFEQVLKIEVDQIIKAYORLG-----SEVVPK	740				
AtAGO7	: VIFMGADVTHHPFDCC-SFSVAAVVGSINWPEANRYVSRMRSQHRCEIIQDLD-----LMVKELLDDFYKAVK-KLPNRIIFRDGVSETQFKKVLQEELQSIK-TACSKFQ-----DYNDS	838				
AtAGO8	: TIIIGNDVSHGSPGQSDHIPISIAAVVSSREWPLISKYRACVRTQSEKEMIDSLFKFVSDKD-----DQGMIRELLDDFHSSS-GKKPNHIIIFRDGVSESOQFNOVLNIELDQMMQ-----707					
AtAGO9	: TIIIVGNDVSHGSPGQSD-IFISIAAVVSSROWPLISKYKACVRTQSEKMEMIDNLFKFPVNGKD-----E-GMFRELLDDFYYSSENKRKPEHIIIFRDGVSESOQFNOVLNIELDOMMOACKFL-----DDTWHPK	746				
AtAGO10	: TIIIFGADVTHHPGEES-SFSIAAVVASQDWPEVTKYAGLVCAQFHRCELIQDLKYTWQDPVRGT-VSGGMIRDLLISFRKATG-QKPLRIIFRDGVSEGQFYQVLLYELDAIRKACASLEP-----NYQPP	827				
TaAGO1a	: TIIIFGADVTHHPGEDS-SFSIAAVVASQDWPEVTKYAGLVSAQFHRCELIQDLFKVWQDPQKGT-VTGGMIKELLISFRKATG-QKPCRIIFRDGVSEGQFYQVLLYELDAIRKACASLEP-----NYQPP	950				
TaAGO1b	: TIIIFGADVTHHPGEDS-SFSIAAVVASQDWPEVTKYAGLVSAQFHRCELIQDLFKVWQDPQKGT-VTGGMIKELLISFRKATG-QKPCRIIFRDGVSEGQFYQVLLYELDAIRKACASLEP-----NYQPP	478				
TaAGO1c	: TIIIFGADVTHHPGEDS-SFSIAAVVASQDWPEVTRYAGLVSAQFHRCELIEDLYKVRQDPQKGP-VSSGMIRELLISFRKSTG-EKPCRIIFRDGVSEGQFYQVLLFELNNAIRKACASLEA-----NYQPP	806				
TaAGO1d	: TIIIFGADVTHHPGEDS-SFSIAAVVASQDWPEVTRYAGLVSAQFHRCELIEDLYKVRQDPQKGP-VSSGMIRELLISFRKSTG-EKPCRIIFRDGVSEGQFYQVLLFELNNAIRKACASLEA-----NYQPP	738				
TaAGO1e	: TIIIFGADVTHHPGEDS-SFSIAAVVASQDWPEVTKYAGLVSAQFHRCELIEDLYNVTHDPQKGT-IHCGMIRELLISFRKSTG-EKPCRIIFRDGVSEGQFYQVLLHEDDAIRKACASLEA-----D-----721					
TaAGO1f	: TIIIFGADVTHHPGEDS-SFSIAAVVASQDWPEVTKYAGLVSAQFHRCELIQDLFKVWQDPQKGT-VNGGMIRELLISFRKSTG-EKPCRIIFRDGVSEGQFYQVLLYELDAIRKACASLES-----NYQPP	897				
TaAGO1g	: TIIIFGADVTHHPGEDS-SFSIAAVVASQDWPEVTKYAGLVSAQFHRCELIQDLFKVWQDPQKGT-VNGGMIRELLISFRKSTG-EKPCRIIFRDGVSEGQFYQVLLYELDAIRKACASLES-----NYQPP	914				
TaAGO1h	: TIIIFGADVTHHPGEDS-SFSIAAVVASQDWPEVTKYAGLVSAQFHRCELIQDLFKVWQDPQKGT-VNGGMIRELLISFRKSTG-EKPCRIIFRDGVSEGQFYQVLLYELDAIRKACASLES-----NYQPP	897				
TaAGO1i	: TIIIFGADVTHHPGEDS-SFSIAAVVASQDWPEITKYAGLVSAQFHRCELIQDLFKVWQDPQKGT-VTGGMIKELLISFRKATG-QKPCRIIFRDGVSEGQFYQVLLFELDAIRKACASLEP-----NYQPP	1041				
TaAGO1j	: TIIIFGADVTHHPGEDS-SFSIAAVVASQDWPEITKYAGLVSAQFHRCELIQDLFKVWQDPQKGT-VTGGMIKELLISFRKATG-QKPCRIIFRDGVSEGQFYQVLLFELDAIRKACASLEP-----NYQPP	699				
TaAGO1k	: TIIIFGADVTHHPGEDS-SFSIAAVVASQDWPEITKYAGLVSAQFHRCELIQDLFKVWQDPQKGT-VTGGMIKELLISFRKATG-QKPCRIIFRDGVSEGQFYQVLLFELDAIRKACASLEP-----NYQPP	1023				
TaAGO2a	: YMFIGADVNHPSGNVE-SFSIAAVVASMNR-GATKYVPRIRAOHRCEVKNLG-----EIVQELIGVFEKKAG-VKPCRIIFRDGVSDGQFEMVNEELADME-KVIKVK-----GYSPT	791				
TaAGO2b	: YMFIGADVNHPSGNVE-SFSIAAVVASMNR-GATKYVPRIRAOHRCEVKNLG-----EIVQELIGVFEKKAG-VKPCRIIFRDGVSDGQFEMVNEELADME-KVIKVK-----GYSPT	684				
TaAGO3	: YMFIGADVNHPPGDTV-SLSIAAVVASMDPCSAQYVPRIRAOHRTEIDNLG-----TMCKELIQVYMERANGGFKPKDIIIFRDGVSDGQFEMVNEELADME-KVICED-----DYSPT	700				
TaAGO4a	: TIIILGNDVSHGSPGQSD-IFISIAAVVGSREWPLVSKYRASVRSQSEKLEMIDSLFK-PQGT-----DDGLVRECLIDFYTSSGKRKPDQIIIFRDGVSESOQFNOVLNIELDOIIEACKFL-----DENWNP	848				
TaAGO4b	: TIIILGNDVSHGSPGQSD-IFISIAAVVGSREWPLVSKYRASVRSQSEKLEMIDSLFK-PQGT-----DDGLVRECLIDFYTSSGKRKPDQIIIFRDGVSESOQFNOVLNIELDOIIEACKFL-----799					
TaAGO4c	: TIIILGNDVSHGSPGQSD-IFISIAAVVGSREWPLVSKYRASVRSQSEKLEMIDSLFK-PQGT-----DDGLVRECLIDFYTSSGKRKPDQIIIFRDGVSESOQFNOVLNIELDOIIEACKFL-----767					
TaAGO5a	: TIIIFGADVTHPHAGDSD-SFSIAAVVASNDWPEITKYKAVVSAQFHRCEIIQDLKYCRGTPEKGTVEHGMIRELLISFYRATG-HKPSRIIFRDGVSEGQFAQVLMYENDAIRKACASLOD-----DYQPP	672				
TaAGO5b	: TIIIFGADVTHPHAGDSD-SFSIAAVVASNDWPEITKYKAVVSAQFHRCEIIQDLKYCRGTPEKGTVEHGMIRELLISFYRATG-HKPSRIIFRDGVSEGQFAQVLMYENDAIRKACASLOE-----DYQPP	672				
TaAGO5c	: TIVFGADVTHPHAGEDA-CFSIAAVVASNDWPEVTKYKCLVSSQHRCEIIADLYAETKDPQKGL-VGGGMIRELLISFYRATG-CKPHRIIFRDGVSEGQFSQVLLYEND-----ACATLOA-----GYLPP	578				
TaAGO5d	: TIVFGADVTHPHAGEDA-SFSIAAVVASNDWPEVTKYKCLVSSQHRCEIIADLYAETKDPQKGL-VGGGMIRELLISFYRATG-CKPHRIIFRDGVSDGQFSQVLLYENDAIRKACATLOA-----GYLPP	721				
TaAGO5e	: TIVFGADVTHPHAGEDA-SFSIAAVVASNDWPEVTKYKCLVSSQHRCEIIADLYAETKDPQKGL-VGGGMIRELLISFYRATG-CKPHRIIFRDGVSEGQFSQVLLYENDAIRKACATLOA-----GYLPP	762				
TaAGO5f	: TIIIFGADVTHSALDIT-APSIASVVASQDWPEVTKYHGDVHAQHRVLEIEGL-----EGIVKALLLSFEKQSK-QKPCQLIFRDGVSEGQFRKVLDEIIFLIEKAWKALYN-----EKDP	682				
TaAGO5g	: TIIIFGADVTHSALDIT-APSIASVVASQDWPEVTKYHGDVHAQHRVLEIEGL-----EGIVKALLLSFEKQSK-QKPCQLIFRDGVSEGQFRKVLDEIIFLIEKAWKALYS-----CKPA	691				
TaAGO6a	: TIIILGNDVSHGPPGRSD-KFSIAAVVGSRYWPLISRYRASVRTQSEKEMIDSLFKPLKGEK-----DDGIIRELLDDFYTSQQRKPEHIIIFRDGVSESOFSQVLLDEVDQIIKAYKNMD-----OGALPK	871				
TaAGO6b	: TIIILGNDVSHGPPGRSD-IFISIAAVVGSRCWPLISRYRASVRTQSEKEMIDSLFKPLDNK-----DDGIIXLLDDFYTSQQRKPEHIIIFRDGVSESOFSQVLLNLEVDQIIKAYQNMG-----QGPPK	733				
TaAGO6c	: TIIILGNDVSHGSPGRSD-IFISIAAVVGSRCWPLISRYRASVRTQSEKEMIDSLFKPLDNK-----DDGIIRELLDDFYTSQQRKPEHIIIFRDGVSESOFSQVLLNLEVDQIIKAYQNMG-----QGPPK	733				
TaAGO6d	: TIIILGNDVSHGSPGRSD-IFISIAAVVGSRCWPLISRYRASVRTQSEKEMIDSLFKPLDNK-----DDGIIRELLDDFYTSQQRKPEHIIIFRDGVSESOFSQVLLNLEVDQIIKAYQNMG-----QGPPK	733				
TaAGO6e	: TIIIFGNDVSHGSPGRSD-VFSVAAVSSLEWPIISKYRASVCTLEKCEMIKTLFK-QDGD-----DHGLIKDSLVAFLGKNKPKPEQIIIFRDGVSESOQFNOVLNIELDOIIEACKFL-----DETWEK	687				
TaAGO7a	: VMFIGADVTHHPPLDSD-SFSVAVVASMNPAAKNYISMRMSTHRKEIEHLD-----VMTGELLEFLKEVG-KLPGRIFRDGVSETQFQVLLYELDAIRKACASLEP-----GYKPL	778				
TaAGO7b	: VMFIGADVTHHPPLDSD-SFSVAVVASMNPAAKNYISMRMSTHRKEIEHLD-----VMTGELLEFLKEVG-KLPGRIFRDGVSETQFQVLLYELDAIRKACASLEP-----GYKPL	867				
TaAGO8	: TIIILGNYVSHGQPGQSD-RPSTAAVNSREWPLISKYRATVHTQSEKLEMSSSLFK-PWKG-----DDGLIRESLIDFYTSSGNRKPEQIIIFRDGVSESOQFNOVLNIELDOIIEACKFL-----DETWEK	762				
TaAGO9a	: TIIILGNDVSHGQPGQSD-RPSTAAVNSREWPLISKYRATVHTQSEKEMMASLFLK-PRGTE-----DDGLIRESLIDFYTSSGNRKPDQVIFRDGVSESOQFNOVLNIELDOIIEACKCL-----DDKWEK	774				
TaAGO9b	: TIIILGNDVSHGQPGQSD-RPSTAAVNSREWPLISKYRATVHTQSEKEMMASLFLK-PRGTE-----DDGLIRESLIDFYTSSGNRKPDQVIFRDGVSESOQFNOVLNIELDOIIEACKCL-----DDKWEK	776				
TaAGO9c	: TIIILGNDVSHGQPGQSD-RPSTAAVNSREWPLISKYRATVHTQSEKEMMASLFLK-PRGTE-----DDGLIRESLIDFYTSSGNRKPDQVIFRDGVSESOQFNOVLNIELDOIIEACKCL-----DDKWEK	771				
TaAGO10a	: TIIIFGADVTHPETREDT-SFSIAAVVASQDWPEVTKYAGLVCAQFHRCELIQDLKYTWQDPQKGT-VTGGMIKELLISFRKATG-QKPLRIIFRDGVSEGQFYQVLLYELDAIRKACASLEP-----NYQPP	795				
TaAGO10b	: TIIIFGADVTHPETREDT-SFSIAAVVASQDWPEVTKYAGLVCAQFHRCELIQDLKYTWQDPQKGT-VTGGMIKELLISFRKATG-QKPLRIIFRDGVSEGQFYQVLLYELDAIRKACASLEP-----NYQPP	730				
TaAGO10c	: TIIIFGADVTHPETGEDS-SFSIAAVVASQDWPEVTKYAGLVCAQFHRCELIQDLKYTWQDPQKGT-VTGGMIKELLISFRKATG-QKPLRIIFRDGVSEGQFYQVLLYELDAIRKACASLEP-----NYQPP	422				
TaAGO10d	: TIIIFGADVTHPETGEDS-SFSIAAVVASQDWPEVTKYAGLVCAQFHRCELIQDLKYTWQDPQKGT-VTGGMIKELLISFRKATG-QKPCRIIFRDGVSEGQFYQVLLYELDAIRKACASLEP-----NYQPP	508				

	1220	1240	1260	1280	1300	1320	
AtAGO1	: VTFVVVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	1005		
AtAGO2	: ITVIVAQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	973		
AtAGO3	: ITVIVAQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	1153		
AtAGO4	: FLLLVAKNNHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	893		
AtAGO5	: VTFVIVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	964		
AtAGO6	: FTVIVAQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	850		
AtAGO7	: ITFAVVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	958		
AtAGO8	: -----INHHTKFFQTESFN-----NVLPGTIIDSNICHCHNNDYFYLCAHAGIKGT	RETHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	819		
AtAGO9	: FTVIVAQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	865		
AtAGO10	: VTFVIVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	952		
TaAGO1a	: VTFVVVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	1075		
TaAGO1b	: VTFVVVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	603		
TaAGO1c	: VTFVIVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	931		
TaAGO1d	: VTFVIVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	863		
TaAGO1e	: -----				-		
TaAGO1f	: VTFVVVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	1023		
TaAGO1g	: VTFVVVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	1040		
TaAGO1h	: VTFVVVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	1023		
TaAGO1i	: VTFVVVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	1166		
TaAGO1j	: VTFVVVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	824		
TaAGO1k	: VTFVVVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	1148		
TaAGO2a	: ITVIVAQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	917		
TaAGO2b	: ITVIVAQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	810		
TaAGO3	: ITVIVAQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	821		
TaAGO4a	: FTLIVAQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	967		
TaAGO4b	: -----				-		
TaAGO4c	: FTLIVAQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	886		
TaAGO5a	: VTFVVVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	796		
TaAGO5b	: VTFVVVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	796		
TaAGO5c	: VTFVVVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	703		
TaAGO5d	: VTFVVVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	846		
TaAGO5e	: VTFVVVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	887		
TaAGO5f	: ITFIVVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	806		
TaAGO5g	: ITFIVVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	815		
TaAGO6a	: ITVIVAQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	893		
TaAGO6b	: ITVIVAQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	852		
TaAGO6c	: ITVIVAQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	851		
TaAGO6d	: ITVIVAQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	852		
TaAGO6e	: FTVIVAQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	768		
TaAGO7a	: TTFVVVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	902		
TaAGO7b	: TTFVVVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	991		
TaAGO8	: FTVIVAQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	878		
TaAGO9a	: FTVIVAQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	893		
TaAGO9b	: FTVIVAQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	895		
TaAGO9c	: FTVIVAQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	890		
TaAGO10a	: VTFVIVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	917		
TaAGO10b	: VTFVIVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	852		
TaAGO10c	: VTFVVVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	552		
TaAGO10d	: VTFVVVQKRHHHTLRFACNHNDRHSVD-RSGNILPGTVVDSKICHETEFDFYLC	SHAGIQGTSREAHYHVLWDENNFTADGLQSLTNNLCYTYARCTRSVSI-----	VFPAYYAHAAFRARFYMEP--	ETSDS	633		

Figure

	840	860	880	900	920	940	
AtRDR1	: EYGQVVVQYSDPMRPGR-----RFIITGPVVVAKNPCLHPGDVRVLQA---VNVP--ALNHMVDCCVFPQKGLRPHPNECSSGS	DLDGD	IYFVCWDQELVP-----	813			
AtRDR2	: EYGQVYVRVTLTKAELKSRDQSYFR--KIDEETSVVIGKVVVTKNPCLHPGDIRVLDA---IYEVHFEEKGYLDCIIFPQKGERPHPNECSSGG	DLDGD	QFFVSWDEKIIP-----	846			
AtRDR3	: KENEICVILHSGQ-----ISGDVLVYRNPGLHFGDIHVLKATYVKALED-YVGNAKFAVFFPQKGRSLGDEIAGG	CFDGI	MYFISRNPKLLEHFKPSEPW	717			
AtRDR4	: KEDEVSGLAKS-----QDVLVYRNPGLHFGDIHILKATYVKSLEQ-YVGNASKYGVFFPQKGRSLGDEIAGG	CFDGI	MYFISRNPKLLEHYKPSSEP	706			
AtRDR5	: KEDEVVCVILESGQ-----ISGEVLVYRNPGLHFGDIHILKATYVKALEE-YVGNASKFAVFFPQKGRSLGDEIAGG	CFDGI	MYFISRNPELLENFKPSEPW	706			
AtRDR6	: EHGQCFIQVSKPSIENCFSKHGSRFK-ETKTDLEVVKGYVAIAKNPCLHPGDVRILEA---VDVP--QLHMYDCLIFPQKGRPHPTNEASGS	DLDGD	LYFVAVWDQKLIPP	881			
CsSHL2	: EQGQCFIRATVPSLNSYFVKHGSRFS-STDKNTEVILGTVVIKPNPCLHPGDVRILEA---VDVP--ELHHLVDCLVFPQKGERPHANEASGS	DLDGD	LYFVTWDEKLIPPG-----	905			
CsRDR1	: KYGQVFIRATSGVNDND-----RFTVTGKVVIKPNPCLHPGDIRILHA---VDVP--VLHMFNCVFPQGGPRPHPNECSSGS	DLDGD	IYFVSWDPSLIP-----	445			
CsRDR2	: EYGQVYIRITKNSKEQKDSNQSYFYNDG--KTATVVGKVAITKNPCLHPGDIRVLEA---IYDPDLVG--MVDCLVFPQGRGERPHPNECSSGG	DLDGD	LYFITWDDKLIP-----	840			
CsRDR3	: RPNEVCVILENGQ-----FSGDVLVYKHPGLHFGDIHVLKATYIRDLEKEYVGYAKYAILFPIISGPRSLADEMANS	CFDGI	IYWVSKNPKLLEHFKPSEPW	155			
CsRDR4	: GPNEVCVILDYQG-----VSGDVLVYKYPGLHPGDIHVLKATYSSDIEK-VVGNASKHAILFPTTGQRSLADEMANS	CFDGI	IYWVSLNPKLLEHFKPSKPW	300			
TaRDR2a	: EYGQVYIRITKNSKEQKDNCPYFALDNGKEKTAVVVGKVAVSKNPCLHPGDIRVLEA---VYDHGLYAKNLVDCVFPQGRGERPHPNECSSGG	DLDGD	LYFITWDEKLIP-----	660			
TaRDR5	: KANEVCVILDNGP-----YCGNVLVYKHPGLHFGDIHVLTSRYIEDIHD-VVGYSRYAILFPTSGPRSLADEMANS	CFDGI	MYWVSINEQLLKQFKPSKPW	422			
TaRDR2b	: EYGQVYIRISKNSKEQKDNCPYFSEDNGTEKTAVVVGKVAVSKNPCLHPGDIRVLEA---VYDHGLYAKNLVDCVFPQGRGERPHPNECSSGG	DLDGD	LYFITWDEKLIP-----	844			
TaRDR1d	: KYGQVFIQASNSANDRG-----KSIVTGKVVVAKNPCLHPGDIRILQA---VHSP--PLGHMVNCVFPQLGPRPHPNECSSGS	DLDGD	IYFVSWDPSLIP-----	452			
TaRDR1a	: KYGQVFIQASNSANDSD-----KFIVTGKVVVAKNPCLHPGDIRILEA---VYTP--VLDHMVNCVFPQGGPRPHPNECSSGS	DLDGD	IYFVSWDPSLIP-----	819			
TaRDR6a	: EQGQCFIRASSPVLNNSLLKHAPRSS--SENNAETVIGTVVMAKNPCLHPGDVRILEA---IDVP--ALHHLVDCLVFPKNGERPHANEASGS	DLDGD	LYFVTWDEKLIPPG-----	611			
TaRDR2d	: EYGQVYIRISKNSKEQKDNCPYFSKDNGKEKTAVVVGKVAVSKNPCLHPGDIRVLEA---VYDHGLYAKNLVDCVFPQGRGERPHPNECSSGG	DLDGD	LYFITWDEKLIP-----	845			
TaRDR1e	: -----MVNCVFPQLGPRPHPNECSSGS	DLDGD	IYFVSWDPSLIP-----	39			
TaRDR6b	: EQGQCFIRASSPVLNNSLLKHGPRSS--SANNAETVKGTVVMAKNPCLHPGDVRILEA---IDVP--ALHHLVDCLVFPKNGERPHANEASGS	DLDGD	LYFVTWDEKLIPPG-----	434			
TaRDR1f	: EYGQVFIQVNSADNRG-----KSIVTGKVVVAKNPCLHPGDIRILQA---VQSP--LLGHMVNCVFPQLGPRPHPNECSSGS	DLDGD	IYFVSWDPSLIP-----	818			
TaRDR1b	: KYGEVFIQASNSAIDSD-----KFVVTGKVIVAKNPCLHPGDIRILEA---VYTP--VLDHMVNCVFPQGGPRPHPNECSSGS	DLDGD	IYFVSWDPSLIP-----	436			
TaRDR3	: KPHEVCVILDHGP-----ISGEVLVYRHPGLHFGDIHVLATYSEAIQD-FVGDSKYAILFPVSGPRSLADEMAGG	CFDGI	MYWVSRNPQLLKYPFKPSEPW	377			
TaRDR4	: KPHEVCVILDHGP-----ISGEVLVYRHPGLHFGDIHVLATYSEAIQD-FVGDSKYAILFPVSGPRSLADEMAGG	CFDGI	MYWVSRNPQLLKYPFKPSEPW	155			
TaRDR2c	: KYGQVYIRITKNSKEQKDNQPYFSKDNRKDKTAVVVGKVAISKNPCLHPGDIRVLEA---VYDHGLYANNLVDCVFPQGRGERPHPNECSSGG	DLDGD	LYFITWDEKLIP-----	840			
TaRDR1c	: KYGEVFIQASNSANDSD-----KFVVTGKVVVAKNPCLHPGDIRILEA---VYTP--VLDHMVNCVFPQGGPRPHPNDECSSGS	DLDGD	IYFVSWDPSLIP-----	818			
TaRDR1g	: KYGQVFIQASNSANDRG-----KSIVTGKVVVAKNPCLHPGDIRILQA---VHSP--PLGHMVNCVFPQLGPRPHPNECSSGS	DLDGD	IYFVSWDPSLIP-----	649			

Figure

Figure

Figure

Figure

Figure

Figure

Figure

Figure

Figure

Figure