Sel'skokhozyaistvennaya Biologiya [Agricultural Biology], 2011, № 5, p. 51-55

UDC 635.263:578.85/.86:578.2:577.217

THE DICER-LIKE PROTEINS CODING NUCLEOTIDE SEQUENCES IN SHALLOT *Allium cepa* L. var. *aggregatum* G. Don

A.V. Arkhipov, V.K. Vishnichenko
All-Russia Research and Development Institute of Agricultural Biotechnology, RAAS, Moscow 127550, Russia
e-mail: vish@iab.ac.ru
Received June 6, 2011

S u m m a r y

For the first time in the world literature the authors presented the evidence of expression in shallot plants(*Allium cepa* L. var. *aggregatum* G. Don) of nucleotide sequences coding the polypeptides homologous to known plant Dicer-like protein (DCL), obtained by PCR, sequencing and computer analysis. The investigated polypeptides contain specific for DCL-protein the dsRNA bind, PAZ and RIBOc domens. The solution of assigned task is available only with use of diagnostic oligonucleotide primers, the design of which was accomplished on the basis of information on EST-sequences (expressed sequence tags) in *Allium sativum* (GarlicESTdb integrated database). The results permit to draw a conclusion about high extent of divergence of DCL-proteins in *Arabidopsis*, poplar, rice, on the one hand, and the plants of *Allium* genus, on the other hand, that is revealed on level both nucleotide sequences of mRNA, and amino acid sequences of studied polypeptides.

Keywords: RNA-interference, Dicer&Dicer-like (DCL) proteins, shallot (*Allium cepa* L. var. *aggregatum*), Shallot virus X.

In plants, the degree of expression of many genes can be modulated at epigenetic level trough RNA interference, which mechanisms involve the key elements – endoribonucleases (RNase III) Dicer or Dicer-like (DCL), endonucleases Argonautes (AGOs) and cellular RNA-dependent RNA polymerase (RDR) (1). The action of these enzymes results in a wide range of small RNAs (microRNAs and small interfering RNA - siRNA) participating different biological processes, particularly those involved in development of plants (2). Phytoviruses are potent exogenous inducers of RNA interference in plants; viral infection leads to a so-called virus-induced gene silencing, VIGS (3), that develops as the result of specific cleavage of replication viral RNA (double-stranded) by plant DCL-proteins. Degradation of viral RNA by these endoribonucleases is accompanied by formation of virus-specific siRNA, and this is the first step of antiviral defense system of plants based on the phenomenon of RNA interference (4).

Plant genomes contain a number of homologous sequences encoding different DCL-proteins (5). So, in *Arabidopsis thaliana*, basic set of four paralogs encodes AtDCL 1-4 – the functionally specialized proteins highly specific to a substrate. (6). AtDCL1 generates small RNAs using as a substrate the single-stranded RNA whose certain areas contain a stem-loop - the imperfect two-stranded structure. AtDCL2, AtDCL3 and AtDCL4 show great affinity to RNA molecules with a perfect double-stranded structure. At viral infection, all four types of *Arabidopsis* DCL-proteins are involved in disintegration of the replicative form of viral RNA, while AtDCL4 plays the main anti-virus function: this protein contributes to formation of virus-specific siRNA of 21 bp length. These siRNAs are, firstly, actively participate the specific degradation of viral RNA by RISC (the effector RNA-induced silencing complex), and secondly, they operate as a mobile signal of RNA interference (7). Small interfering RNAs generated by AtDCL2 play a subordinate role in these processes, while AtDCL1 and AtDCL3 have no antiviral activity though being capable to cut the double-stranded virus RNA into short fragments (24 bp). However, it is possible that the presented model is not universal, and plants having a set of DCL-proteins distinct from that of *Arabidopsis* can develop a distinct type of functional relationships between them.

The purpose of this study was detecting the expression and analysis of nucleotide sequences encoding DCL-proteins in *Allium cepa* var. aggregatum - the only susceptible host for X-virus of shallot (HVSH, prototype of the genera Allexivirus) (8). Prior to this study, in the world scientific literature there were found no information about DCL-proteins of *Allium cepa*.

Technique. Vegetative reproduction of HVSH-infected shallot plants (*Allium cepa* L. var. *aggregatum* G. Don) was performed under the conditions of monovirus infection. Control plants were non-infected shallot the variety Red Sun ("Crocus Co.", UK).

Total RNA was extracted from leaves 2 weeks after planting the bulbs with trizol (TRIzol Reagent, "Life Technologies", USA) according to the manufacturer's protocol. Synthesis of cDNA was done under the scheme: at the annealing stage, the incubation mixture containing 0,5-2,0 ug total RNA, 2 ul oligo$(dT)_{18}$ (40 uM) or "random" primers ("random" hexamers, 40 uM), 4 ul of dNTP mixture (each at a concentration of 2,5 mM; pH 7,0; Tris-HCl) and nuclease-free H_2O (total volume of the mixture 16 ul) was heated at 70 °C for 5 min, centrifuged and immediately placed on ice; the mixture was added with 2 ul 10× RT buffer, 1 ul (10 units) ribonuclease inhibitor (RNasin) and 1 ul (50 units) revertase (M-MLV, Moloney) ("Synthol", Russia). Reverse transcription (in volume 20 ml) was performed at 42 °C for 1 h, and then the enzyme was inactivated by heating at 90 °C for 10 min. The design of PCR-primers was performed using two groups of them: diagnostic primers (correspond to conservative fragments of genomic sequences of *Arabidopsis* DCL-proteins; allow PCR identification of 1-4 DCL-proteins) (9), and primers designed via the alternative approach based on using the database GarlicESTdb (10). Polymerase chain reaction was performed under the following regime: denaturation at 94 °C for 2 min, then 30 cycles (94 °C, 30 s; 60 °C, 30 s; 72 ° C, 30 s) and final elongation at 72 °C for 2 min. Amplicons were fractionated in 0,8% agarose gels, extracted and purified using DNA extraction kit № K0513 ("Fermentas", Lithuania), nucleotide sequence of the product was determined on the sequencer 3130xl ("Applied Biosystems", USA). Computer analysis of the obtained sequences was performed using databases and software package of NCBI (The National Center for Biotechnology Information, USA).

Results. The target was supposed to be reached by using the diagnostic primers corresponding to conservative sites in genomic sequences of arabidopsis DCL-proteins (9). Such experimental approach was based on the concept that all angiosperms and possibly all multi-cellular plants have a basic set of four highly homologous types of DCL-proteins. The existing experimental data suggest that these primers allow PCR-identification of the genes for 1-4 DCL-proteins in many crops along with arabidopsis – barley, maize, cotton and lupine (9). In authors' experiments, these primers provided reliable detection of 1-4DCL-proteins in arabidopsis plants, but applying them to nuclear DNA or cDNA of shallots was unsuccessful: amplicons were not synthesized at all, or the reaction products had no any pronounced homology with sequences encoding 1-4 DCL-proteins (data not shown). This fact was interpreted as the evidence of divergence between the genes for DCL-protein in *A. thaliana* and *Al. cepa*. A direct comparison of DNA sequences of arabidopsis and table onion can't be performed owing to the absence of necessary data about *Al. cepa*: sequencing

of its genome is a difficult and still unresolved problem; the recently published results of multiyear work on sequencing the genome of onion (11) have revealed in it only the extended regions of degenerated retroviral elements and transposons, but no information about any genes for DCL-proteins. In this regard, this study was a new attempt based on using the database GarlicESTdb (http://garlicdb.kribb.re.kr) keeping the data about 21,595 ESTs (expressed sequence tags) of *Al. sativum* (10). Though, using a similar database on *Al. cepa* (12) containing a much smaller number of EST-sequences, was unsuccessful.

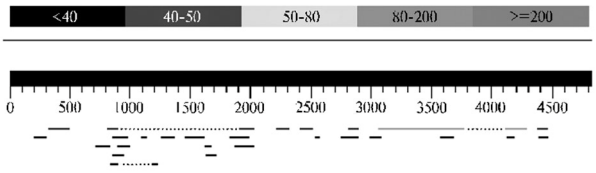

Fig. 1. Results of screening of EST-sequences complementary to mRNA for AtDCL2 upon the database GarlicESTdb.

The identified ESTs (expressed sequence tags) correspond to nucleotide sequences of mRNA at positions 3100-3750 and 4100-4250. The similar data were obtained using mRNA encoding all known DCL-proteins of plants.

Design of primers based on the database GarlicESTdb was carried out as follows. The program BLAST/tblastx, which translates the proposed nucleotide sequence and then compares the resulting amino acid sequences to each other, was used to find in GarlicESTdb the EST-sequences encoding polypeptides with a high degree of homology to various DCL-proteins. Figure 1 shows the result of such screening for mRNA of AtDCL2; the similar data were obtained for mRNAs of all known plant DCL-proteins. Similarly, three EST-sequences were selected from GarlicESTdb - EPP005KGAA12S003959 (783 bp), EPP004KGAA12S004834 (679 bp) and EPP004KGAA12S005240 (675 bp). All of them contain one open reading frame, and the level of homology of translated polypeptide sequences to different DCL-protein of arabidopsis, rice and poplar ranging from 60 to 70%. Along with it, the computer analysis has revealed in these polypeptide chains the domains characteristic to the molecule of DCL-protein dsRNA bind, PAZ and RIBOc (RNase IIIa and RNase IIIb) (data not shown). This information was a basis for creation of primers corresponding to various parts of the three selected EST-sequences (Table).

During a reverse transcription and polymerase chain reaction using total RNA of healthy and infected shallot plant under standard conditions, the pairs of primers L-106-R-644 (EPP005KGAA12S003959), L-133-R-632 (EPP004KGAA12S004834) and L-38-R-552 (EPP004KGAA12S005240) were found to be the most effective and providing stable results (Fig. 2, A, B, C).

Diagnostic PCR-primers used for amplification of nucleotide sequences encoding DCL-proteins in shallot.

Primer designation	Nucleotide sequence of a primer
	E S T - sequence EPP004KGAA12S005240
L-8	CGCTCAGCCCTTTTAAACAA
L-38[a]	TATGAGCTGGCCATTGCTAA
L-102	ACTTCTTTGCATTGGGCAGA
L-378	ACGAGGAAGCCAAAAAGGAT
R-552[b]	GCATCCCCAAGATACTCCAA
R-397	ATCCTTTTTGGCTTCCTCGT
R-120	CTGCCCAATGCAAAGAAGTT
R-628	AAATCTGAAATTTGACCTGGTTTC
	E S T - sequence EPP004KGAA12S004834
L-133[a]	GTCCCCGTTTTGTGATTGAC
L-90	TTCTACTGCTCCCGACTTCC
L-38	TGAATCAACTGGAGCAAAACC
L-323	AGCAGAGCTTGCTTCTGAGG
L-395	GGTTGATTCTTGGTCCTCGT
R-567	CGCCCTGGTCTTCTTCTATG
R-632[b]	AGCCATCAAAACCTGCTCAC
R-415	ACGAGGACCAAGAATCAACC
R-4834	ATATTCTGACGAATCAGTAAATTGA
	E S T - sequence EPP005KGAA12S003959
L-285	CCGGAACTTTGCAAAATCAT
L-106[a]	TGAAGAGCAGGGAAACTCTCA
L-3	CATTGTGCCTTTGAAAATTCTG
R-432	TGACATTTTGCATGCAGTGATCC
R-644[b]	TCCTATCCGGCATAAGTTGG

Note. a, b – most effective pairs of primers

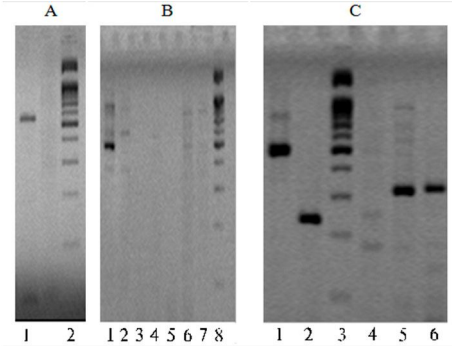

Fig. 2. Electrophoregrams showing amplicons during the detection of expressed nucleotide sequences encoding DCL-proteins in shallot using PCR with diagnostic primers designed upon the database GarlicESTdb: A — primer EST-sequences EPP005KGAA12S003959 (1 and 2 — respectively, the pair L-106—R-644 and molecular weight marker); B — primer EST-sequence EPP004KGAA12S005240 (1-8 — respectively, the pairs L38—R552, L38-F—R628, L102—R397, L102—R552, L102—R628, L378—R552, L378—R628 and molecular weight marker); C — primer EST-sequences EPP004KGAA12S004834

(1, 2, 4, 5 and 6 — respectively, L133—R632, L395—R632, L90—R4834, L90—R408, L90—R415, 3 — molecular weight marker). EST — expressed sequence tags; the molecular weight marker M25 100-3000 bp DNA Ladder ("SibEnzyme", Russia).

After extraction from the gel, purification and sequencing of amplicons, the computer analysis revealed in corresponding databases no sequences homologous to the studied ones. At the same time, the polypeptides translated from these sequences manifested high homology to all plant DCL-proteins; particularly, the highest degree of homology (81%) was observed with DCL-proteins of *Populus trichocarpa*. In this case, amino acid sequences in all three polypeptides were not identical while carrying in the molecule different specific domains (dsRNA bind, PAZ or RIBOc, fig. A, B, C).

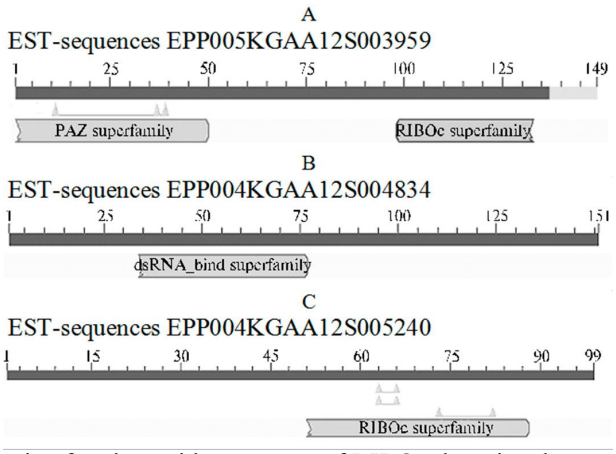

Fig. 3. Results of computer domain analysis of polypeptides translated from nucleotide sequences amplified at the use of the most effective pairs of primers with total cDNA of shallot: A — L-106—R-644, B — L-133—R-632, C — L-38—R-552; EST — expressed sequence tags.

During the analysis of amino acid sequences of RIBOc domains detected in two polypeptides, there were revealed that each of the three characteristic sites of domains (active site, metal binding site and dimerization interface) contains two amino acid substitutions, whereas the sequences of these sites in all other known plant DCL-proteins are identical.

Thus, for the first time in world scientific literature, this work presents the evidence of expression in normal and XVSH-infected shallot plants (*Allium cepa* L. var. *aggregatum* G. Don) of nucleotide sequences encoding three domains for different DCL-proteins - dsRNA bind, PAZ and RIBOc (RNase IIIa and RNase IIIb). In addition, the obtained data show a diversity of DCL-proteins in plants of different species, on the one hand, and within the genus *Allium* – on the other. This fact was observed at both levels of mRNA nucleotide sequences and amino acid sequences of studied polypeptides. At the same time, evolutionary relationships between nucleotide sequences encoding these domains still remain unknown.

REFERENCES

1. Meister G. and Tuschl T., Mechanisms of Gene Silencing by Double-Stranded RNA, *Nature,* 2004, vol. 431, pp. 343-349.
2. Ketting R.F., The Many Faces of RNAi, *Developmental Cell*, 2011, vol. 20, pp. 148-161.
3. Baulcombe D., RNA Silencing in Plants, *Nature*, 2004, vol. 431, pp. 356-363.
4. Garcia-Ruiz H., Takeda A., Chapman E.J. et al., Arabidopsis RNA-Dependent RNA Polymerases and Dicer-Like Proteins in Antiviral Defence and Small Interfering RNA Biogenesis during Turnip Mosaic Virus Infection, *The Plant Cell*, 2010, vol. 22, pp. 481-496.
5. Liu Q., Ruiz-Ferrer V. and Voinnet O., Roles of Plant Small RNAs in Biotic Stress Responses, *Annu. Rev. Plant Biol.,* 2009, vol. 60, pp. 485-510.
6. Dunoyer P., Schott G., Himber C. et al., Small RNA Duplexes Function as Mobile Silencing Signals between Plant Cells. *Science*, 2010, vol. 328, pp. 912-916.
7. Vishnichenko V.K., Konareva T.N. and Zavriev S.K., A New Filamentous Virus in Shallot, *Plant Pathol*., 1993, vol. 42, pp. 121-125.
8. Margis R., Fusaro A.F., Smith N.A., Curtin S.J., Watson J.M., Finnegan E.J. and Waterhouse P.M., The Evolution and Diversification of Dicers in Plants, *FEBS Let.,* 2006, vol. 580, pp. 2442-2450.
9. Kim D.-W., Jung T.-S., Nam S.-H. et al, GarlicESTdb: an Online Database and Mining Tool for Garlic EST Sequences, *BMC Plant Biol.,* 2009, no. 9, p. 61.
10. Jakne J., Meyer J.D.F., Suzuki G. et al., Pilot Sequencing of Onion Genomic DNA Reveals Fragments of Transposable Elements, Low Gene Densities, and Significant Gene Enrichment after Methyl Filtration, *Mol. Genet. Genomics*, 2008, vol. 280, pp. 287-292.
11. Joseph C., Kuhl J.C., Cheung F. et al., A Unique Set of 11,008 Onion Expressed Sequence Tags Reveals Expressed Sequence and Genomic Differences between the Monocot Orders Asparagales and Poales, *The Plant Cell*, 2004, vol. 16, pp. 114-125.