IUFS Journal of Biology *Short communication*
IUFS J Biol 2013, 72(1):53-63

In silico analysis of dicer-like protein (DCLs) sequences from higher plant species

Ertugrul Filiz[1*], Ibrahim Koc[2]

[1]*Department of Crop and Animal Production, Cilimli Vocational School, Duzce University, 81750 Duzce, Turkey*
[2]*Department of Molecular Biology and Genetics, Gebze Institute of Technology, 41400 Kocaeli, Turkey*

Abstract

Dicer and Dicer like (DCLs) proteins are essential part of small RNA biogenesis pathway, is a type of RNase III digesting long dsRNA (pre-miRNA) to small RNA segments (miRNA). A total of 20 full length of Dicer like proteins (DCL1, DCL2, DCL3 and DCL4) from different organisms available in NCBI were evaluated by bioinformatics tools to investigate properties, structure of DCLs, domain analysis, multiple sequence alignment and phylogenetics tree construction. All DCLs protein sequences have *Ribonuclease III* protein family that contains *RNaseIII domain* including Helicase ATP-binding type-1, Helicase C-terminal, Dicer double-stranded RNA-binding fold, PAZ, Ribonuclease III, Double stranded RNA-binding domain (dsRB). Physicochemical analysis offers data such as pI, EC, Al, GRAVY and instability index about these enzymes. Putative phosphorylation sites were also identified which are found to be conserved in plant species and the results showed that the most abundant phosphorylation site is Serine residues in DCLs proteins. Patterns and profile analysis were performed using Prosite and conserved protein motifs subjected to MEME to obtain the best possible matches. The phylogenetics tree represented three major clusters and similar DCLs protein sequences of different plant species clustered together. The obtained results could be used for further in silico analysis and homology modeling studies.

Keywords: Dicer, DCLs, miRNA, RNase, *In silico* analysis.
***Corresponding Author:** Ertuğrul Filiz (e-mail: ertugrulfiliz@gmail.com).
(Received: 25.05.2012 Accepted: 24.01.2013)

Gelişmiş bitki türlerinde dicer-benzeri protein (DCLs) dizilerinin *in silico* analizi

Özet

Dicer ve Dicer benzeri (DCLs) proteinler küçük RNA biyogenezi yolunun bir parçasıdır ve uzun dsRNA (pre-miRNA)'yı küçük RNA parçalarına sindiren RNase III tipi proteinlerdir. NCBI'da kullanılabilir farklı organizmalara ait 20 tam uzunlukta Dicer benzeri proteinler (DCL1, DCL2, DCL3 and DCL4) biyoenformatik araçlar yardımıyla DCLs'lerin özellik ve yapıları, domain analizleri, çoklu dizi hizalanmaları ve filogenetik ağaç yapımının araştırılması için değerlendirilmiştir. Bütün DCLs protein dizileri, *RNaseIII* domainin kapsadığı helikaz ATP-bağlanma tip–1, helikaz C-terminal, dicer çift zincir RNA, PAZ, ribonükleaz III, çift zincir RNA-bağlanma domain (dsRB) ailesinin bulunduğu *ribonükleaz III* protein ailesine sahiptir. Fizikokimyasal analizler bu enzimler hakkında pI, EC, Al, GRAVY ve instabilite indeksi gibi bilgileri sunmuştur. Bitki türerinde korunurlu varsayılan fosforilasyon bölgeleri belirlenmiştir ve sonuçlar DCLs proteinlerinde en sık fosforilasyon bölgesinin serin kalıntısında olduğunu göstermiştir. Motif ve profil analizlerinde Prosite, korunurlu protein motiflerinde en iyi muhtemel eşleşmelerin elde etmek için MEME kullanılmıştır. Filogenetik ağaç üç ana kümeyle temsil edilmiş ve farklı bitki türlerindeki DCLs protein dizileri beraber kümelenmiştir. Elde edilen sonuçlar, gelecekteki homoloji modelleme ve *in silico* analizlerde kullanılabilir.

Anahtar kelimeler: Dicer, DCLs, miRNA, RNase, *In silico* analiz.

Introduction

Initiation of RNA based gene silencing mechanism includes generation of double stranded RNA (dsRNA) in different ways which are bidirectional transcription of DNA and self-complementary RNA foldbacks (Matzke and Birchler 2005). The complementary dsRNAs are treated by the RNaseIII-type activities of Dicers into small RNAs (siRNA or miRNA) containing ~19–31 nucleotides (Kapoor et al. 2008). A microRNA (miRNA) is a 21–24 nucleotide (nt) small RNA that is the ultimate product of a non-coding RNA gene (Kim 2005; Chen 2005). The miRNA is integrated the ribonucleoprotein complex called RISC which serves the cleavage or translational repression of its target mRNAs (Bartel 2004).

Dicer or Dicer-like (DCLs) proteins are basic components in the miRNA biogenesis pathways in converting long double-stranded RNAs into mature small RNAs (Groβhans and Filipowicz 2008). Dicers are classified by the presence of six types of domains, DExD-helicase, helicase-C, Duf283, PAZ, RNaseIII and double stranded RNA-binding (dsRB) domain. (Margis et al. 2006). RNase III proteins have various lengths from ~200 to ~2000 amino acids, and have three classes based on domain composition (Lamontagne et al. 2001). Class I RNase III enzymes have a single ribonuclease domain and a dsRNA-binding domain (dsRBD). Class II RNase proteins contain a dsRBD and two ribonuclease domains. Class III proteins typically include two ribonuclease domains, a dsRBD and an N-terminal DExD/H-box helicase domain and a small domain of unknown function (DUF283) and a PAZ domain are found in Class 3 proteins family. Class 3 RNase III proteins are also known as the Dicer family of enzymes (MacRae and Doudna 2007; Bernstein et al. 2001). In *Arabidopsis thaliana*, four Dicer-like proteins (DCL1–DCL4) were found including different roles (Moissiard et al. 2007). DCL1 is associated with miRNA production and also affords important role producing small RNAs from endogenous inverted repeats and DCL2 makes siRNAs from natural cis-acting antisense transcripts. DCL3 produces siRNAs related to chromatin modification and DCL4 is connected with tasiRNA (trans acting siRNA) metabolism and performs posttranscriptional silencing process (Liu et al. 2009).

Computational tools provide opportunities to researchers for understanding the physicochemical and structural properties of proteins and many computational tools can be obtained from different sources (Sivakumar et al. 2007). Determining and characterizing the molecule's function, physical and chemical properties of proteins were obtained from the protein sequence. The statistics about a protein sequence such as number of amino acid, sequence length and the physicochemical properties of proteins such as molecular weight, atomic composition, extinction coefficient, GRAVY, aliphatic index, instability index, etc. can be computed by computational tools for the prediction and characterization of protein structure. In this paper, in silico analysis and characterization studies on 20 Dicer (DCLs) proteins of higher plant species were used.

Materials and methods

Data collection and analysis

All amino acid sequences for the selected plant species were collected from NCBI Entrez protein database (Table 1). The Dicer protein sequences were collected in FASTA format and used for further analysis.

Table 1. List of dicer protein sequences reported from some plant species.

Species	NCBI Accession no.
Brachypodium distachyon	XP_003558946, XP_003581414, XP_003566182, XP_003559709
Glycine max	XP_003553805, XP_003523657, XP_003550797, XP_003535104
Vitis vinifera	XP_002269915, XP_002264486, XP_002268369
Populus trichocarpa	EEE81952
Oryza sativa Japonica	ABS32306
Oryza sativa Indica	ABB20894
Medicago truncatula	XP_003603236
Ricinus communis	XP_002515097
Brassica rapa	ACE60552
Arabidopsis thaliana	AEE73926, AEE77843, NP_001190348

The multiple sequence alignment was performed by using ClustalW2 sequence alignment program and the phylogenetic tree was constructed by using the Neighbor-Joining (NJ) method in MEGA 5.1 (Fig. 2) (Tamura et al. 2011). For physicochemical characterization, theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient (Gill and Von Hippel 1989), instability index (Guruprasad et al. 1990), aliphatic index (Ikai, 1980) and grand average hydropathy (GRAVY) (Kyte and Doolottle 1982) were calculated using the Expasy's ProtParam server (Gasteiger 2005) (http://us.expasy.org/tools/protparam.html). (Table 2.) The amino acid sequences of the selected plants were analyzed for the putative phosphorylation sites at the NetPhos 2.0 Server (http://www.cbs.dtu.dk/services/NetPhos/) (Table 3.) (Blom et al. 1999). Prosite is a public database of protein families and domains (Falquet et al. 2002) and it was used to analyze amino sequences of protein with specific profiles and patterns (Table 4). Analysis of domain and conserved protein motifs was performed using MEME (http://meme.sdsc.edu/meme/meme.html) (Timothy et al. 2009). The conserved protein motifs inferred by MEME were analyzed to biological functional analysis using protein BLAST and domains were characterized using Interproscan supporting the best possible match based on highest similarity score.

Results and Discussion

Dicer protein sequences of some plant species (Table 1) were analyzed in this study and these protein sequences were collected from Genbank. Physio-chemical properties were examined to find differences between twenty dicer protein sequences using Expasy's ProtParam tool (Table 2). The isoelectronic point is the pH at which the protein does not migrate in an electric field. It plays important role for protein purification. The computed pI value that was less than 7 (pI<7) indicates that proteins were considered as acidic as or greater than 7 (pI>7) reveals that these dicer proteins were basic in character. The pI value of XP_003535104 and ABS32306 that are greater than 7 (pI>7) reveals that these proteins were basic in character.

Table 2. Parameters computed using Expasy's ProtParam tool.

Organism and protein name	Accession Number	Sequence Length	M.wt	pl	-R	+R	EC	II	Al	GRAVY
PREDICTED: endoribonuclease Dicer homolog 1-like [Brachypodium distachyon]	XP_003558946	1888	210368.5	6.17	255	232	180555	41.63	82.76	-0.367
PREDICTED: endoribonuclease Dicer homolog 1-like [Glycine max]	XP_003553805	1942	217362.5	6.13	270	246	193755	37.58	80.19	-0.454
PREDICTED: endoribonuclease Dicer homolog 3a-like [Glycine max]	XP_003523657	1671	187678.7	6.30	218	200	139490	47.53	91.06	-0.267
PREDICTED: Dicer-like protein 4-like [Glycine max]	XP_003550797	1636	184048.3	6.30	204	186	142345	47.09	90.82	-0.203
PREDICTED: endoribonuclease Dicer homolog 2-like [Vitis vinifera]	XP_002269915	1394	157771.1	6.99	162	159	139980	44.20	95.70	-0.143
PREDICTED:Dicer-like protein 4 [Vitis vinifera]	XP_002264486	1622	183082.3	6.45	203	188	139240	47.34	91.27	-0.220
PREDICTED: endoribonuclease Dicer homolog 1-like [Vitis vinifera]	XP_002268369	1971	221324.8	5.96	280	250	216660	41.33	81.31	-0.447
PREDICTED: endoribonuclease Dicer homolog 2a-like [Brachypodium distachyon]	XP_003559709	1402	157797.6	6.63	156	149	158235	46.82	97.03	-0.109
PREDICTED: endoribonuclease Dicer homolog 4-like [Brachypodium distachyon]	XP_003581414	1627	184917.6	6.20	213	191	132960	42.45	91.55	-0.225
PREDICTED: endoribonuclease Dicer homolog 2-like [Glycine max]	XP_003535104	1414	160386.2	7.38	167	167	163720	38.31	93.47	-0.186
Dicer-like protein [Populus trichocarpa]	EEE81952	1817	204250.9	5.77	270	235	199450	40.24	82.46	-0.487
Dicer-like protein [Oryza sativa Japonica Group]	ABS32306	1657	186989.1	7.20	201	199	127780	44.36	91.56	-0.220
PREDICTED: endoribonuclease Dicer homolog 3a-like [Brachypodium distachyon]	XP_003566182	1806	196595.4	5.99	224	197	160725	46.72	84.72	-0.300
Dicer-like protein [Oryza sativa Indica Group]	ABB20894	1116	126222.4	6.32	133	120	114775	44.96	94.52	-0.180
Endoribonuclease Dicer-like protein [Medicago truncatula]	XP_003603236	1758	195891.6	6.37	230	215	151645	41.51	91.50	-0.253
Dicer-1, putative [Ricinus communis]	XP_002515097	1543	172681.0	6.02	208	186	153000	41.02	88.77	-0.301
Dicer-like protein 2 [Brassica rapa]	ACE60552	1392	157415.5	7.25	155	154	147555	41.71	93.34	-0.117
Dicer-like protein 2 [Arabidopsis thaliana]	AEE73926	1388	156864.6	6.32	160	144	145730	41.76	94.85	-0.129
Dicer-like protein 3 [Arabidopsis thaliana]	AEE77843	1580	177424.8	5.98	210	185	113870	43.52	88.35	-0.308
Dicer-like protein 4 [Arabidopsis thaliana]	NP_001190348	1688	189677.7	6.29	216	196	134435	44.84	94.02	-0.209

Table 3. Putative phosphorylation residues in some plant species with score above 0.8.

Protein name	Accession No.	Putative Phosphorylation site prediction		
PREDICTED: Endoribonuclease Dicer homolog 1-like [Brachypodium distachyon]	XP_003558946	Ser: 61	Thr: 21	Tyr: 15
PREDICTED: Endoribonuclease Dicer homolog 1-like [Glycine max]	XP_003553805	Ser: 61	Thr: 30	Tyr: 16
PREDICTED: Endoribonuclease Dicer homolog 3a-like [Glycine max]	XP_003523657	Ser: 79	Thr: 15	Tyr: 19
PREDICTED: Dicer-like protein 4-like [Glycine max]	XP_003550797	Ser: 61	Thr: 16	Tyr: 12
PREDICTED: Endoribonuclease Dicer homolog 2-like [Vitis vinifera]	XP_002269915	Ser: 44	Thr: 16	Tyr: 16
PREDICTED: Dicer-like protein 4 [Vitis vinifera]	XP_002264486	Ser: 65	Thr: 12	Tyr: 12
PREDICTED: Endoribonuclease Dicer homolog 1-like [Vitis vinifera]	XP_002268369	Ser: 64	Thr: 28	Tyr: 21
PREDICTED: Endoribonuclease Dicer homolog 2-like [Brachypodium distachyon]	XP_003559709	Ser: 43	Thr: 16	Tyr: 15
PREDICTED: Endoribonuclease Dicer homolog 4-like [Brachypodium distachyon]	XP_003581414	Ser: 47	Thr: 12	Tyr: 16
PREDICTED: Endoribonuclease Dicer homolog 2-like [Glycine max]	XP_003535104	Ser: 32	Thr: 17	Tyr: 13
Dicer-like protein [Populus trichocarpa]	EEE81952	Ser: 61	Thr: 18	Tyr: 15
Dicer-like protein [Oryza sativa Japonica Group]	ABS32306	Ser: 61	Thr: 10	Tyr: 12
PREDICTED: Endoribonuclease Dicer homolog 3a-like [Brachypodium distachyon]	XP_003566182	Ser: 61	Thr: 28	Tyr: 13
Dicer-like protein [Oryza sativa Indica Group]	ABB20894	Ser: 61	Thr: 28	Tyr: 13
Endoribonuclease Dicer-like protein [Medicago truncatula]	XP_003603236	Ser: 68	Thr: 15	Tyr: 19
Dicer-1, putative [Ricinus communis]	XP_002515097	Ser: 48	Thr: 18	Tyr: 15
Dicer-like protein 2 [Brassica rapa]	ACE60552	Ser: 38	Thr: 14	Tyr: 18
Dicer-like protein 2 [Arabidopsis thaliana]	AEE73926	Ser: 42	Thr: 17	Tyr: 17
Dicer-like protein 3 [Arabidopsis thaliana]	EE77843	Ser: 70	Thr: 23	Tyr: 12
Dicer-like protein 4 [Arabidopsis thaliana]	NP_001190348	Ser: 65	Thr: 18	Tyr: 11

Table 4. Functional characterization of dicer proteins in some plant species at Prosite.

Protein name	Accession No.	Pattern	Profile
PREDICTED: Endoribonuclease Dicer homolog 1-like [Brachypodium distachyon]	XP_003558946	Helicase ATP-binding type-1, Helicase C-terminal, Dicer double-stranded RNA-binding fold, PAZ, Ribonuclease III , Double stranded RNA-binding domain	Ribonuclease III
PREDICTED: Endoribonuclease Dicer homolog 1-like [Glycine max]	XP_003553805	Helicase ATP-binding type-1, Helicase C-terminal, Dicer double-stranded RNA-binding fold, PAZ, Ribonuclease III , Double stranded RNA-binding domain	Ribonuclease III
PREDICTED: Endoribonuclease Dicer homolog 2-like [Glycine max]	XP_003553782	Helicase ATP-binding type-1, Helicase C-terminal, Dicer double-stranded RNA-binding fold, PAZ, Ribonuclease III , Double stranded RNA-binding domain	Ribonuclease III
PREDICTED: Endoribonuclease Dicer homolog 3a-like [Glycine max]	XP_003523657	Helicase ATP-binding type-1, Helicase C-terminal, Dicer double-stranded RNA-binding fold, PAZ, Ribonuclease III , Double stranded RNA-binding domain	Ribonuclease III
PREDICTED: Dicer-like protein 4-like [Glycine max]	XP_003550797	Helicase ATP-binding type-1, Helicase C-terminal, Dicer double-stranded RNA-binding fold, PAZ, Ribonuclease III , Double stranded RNA-binding domain	Ribonuclease III Serine proteases, subtilase family, aspartic acid active site
PREDICTED: Endoribonuclease Dicer homolog 2-like [Vitis vinifera]	XP_002269915	Helicase ATP-binding type-1, Helicase C-terminal, Dicer double-stranded RNA-binding fold, PAZ, Ribonuclease III , Double stranded RNA-binding domain	Ribonuclease III
PREDICTED: Dicer-like protein 4 [Vitis vinifera]	XP_002264486	Helicase ATP-binding type-1, Helicase C-terminal, Dicer double-stranded RNA-binding fold, PAZ, Ribonuclease III , Double stranded RNA-binding domain	Ribonuclease III
PREDICTED: Endoribonuclease Dicer homolog 1-like [Vitis vinifera]	XP_002268369	Helicase ATP-binding type-1, Helicase C-terminal, Dicer double-stranded RNA-binding fold, PAZ, Ribonuclease III , Double stranded RNA-binding domain	Ribonuclease III
PREDICTED: Endoribonuclease Dicer homolog 2-like [Brachypodium distachyon]	XP_003559709	Helicase ATP-binding type-1, Helicase C-terminal, Dicer double-stranded RNA-binding fold, PAZ, Ribonuclease III , Double stranded RNA-binding domain	Ribonuclease III
PREDICTED: Endoribonuclease Dicer homolog 4-like [Brachypodium distachyon]	XP_003581414	Helicase ATP-binding type-1, Helicase C-terminal, Dicer double-stranded RNA-binding fold, PAZ, Ribonuclease III , Double stranded RNA-binding domain	Ribonuclease III Serine proteases.
PREDICTED: Endoribonuclease Dicer homolog 2-like [Glycine max]	XP_003535104	Helicase ATP-binding type-1, Helicase C-terminal, Dicer double-stranded RNA-binding fold, PAZ, Ribonuclease III , Double stranded RNA-binding domain	Ribonuclease III Phosphopantetheine attachment site
Dicer-like protein [Populus trichocarpa]	EEE81952	Helicase ATP-binding type-1, Dicer double-stranded RNA-binding fold, PAZ, Ribonuclease III, Double stranded RNA-binding domain	Aldo/keto reductase family putative active site, Ribonuclease III
Dicer-like protein [Oryza sativa Japonica Group]	ABS32306	Helicase ATP-binding type-1, Dicer double-stranded RNA-binding fold, PAZ, Ribonuclease III, Double stranded RNA-binding domain	Serpins signature,Ribonuclease III, Serine proteases, subtilase family, aspartic acid active site
PREDICTED: Endoribonuclease Dicer homolog 3a-like [Brachypodium distachyon]	XP_003566182	Dicer double-stranded RNA-binding fold, PAZ, Ribonuclease III , Double stranded RNA-binding domain	Ribonuclease III
Dicer-like protein [Oryza sativa Indica Group]	ABB20894	Dicer double-stranded RNA-binding fold, PAZ, Ribonuclease III	Ribonuclease III
Endoribonuclease Dicer-like protein [Medicago truncatula]	XP_003603236	Helicase ATP-binding type-1, Dicer double-stranded RNA-binding fold, PAZ, Ribonuclease III , Double stranded RNA-binding domain	Ribonuclease III

Table 4. Continued.

Protein name	Accession No.	Pattern	Profile
Dicer-1, putative [Ricinus communis]	XP_002515097	Helicase ATP-binding type-1, Helicase C-terminal, Dicer double-stranded RNA-binding fold, PAZ, Ribonuclease III , Double stranded RNA-binding domain	Ribonuclease III
Dicer-like protein 2 [Brassica rapa]	ACE60552	Helicase ATP-binding type-1, Helicase C-terminal, Dicer double-stranded RNA-binding fold, PAZ, Ribonuclease III , Double stranded RNA-binding domain	Ribonuclease III
Protein dicer-like 2 [Arabidopsis thaliana]	AEE73926	Helicase ATP-binding type-1, Helicase C-terminal, Dicer double-stranded RNA-binding fold, PAZ, Ribonuclease III , Double stranded RNA-binding domain	Ribonuclease III
Protein dicer-like 3 [Arabidopsis thaliana]	AEE77843	Helicase ATP-binding type-1, Helicase C-terminal, PAZ, Ribonuclease III , Double stranded RNA-binding domain	Ribonuclease III,
Dicer-like protein 4 [Arabidopsis thaliana]	NP_001190348	Helicase ATP-binding type-1, Helicase C-terminal, Dicer double-stranded RNA-binding fold, PAZ, Ribonuclease III , Double stranded RNA-binding domain	Ribonuclease III

Table 5. Different motifs commonly observed in dicer protein sequences with best possible match amino acid sequences.

Motif number	Width Sequence	Protein sequences	Occurrence in Dicer sequences
1	50	CYQRLEFLGDAVLDYLITWHLYFTYPDLPPGQLTDLRSASVNNENFAQVA	20
2	50	GRVNLLVATSVGEEGLDIQTCNCVIRFDPPKTVCSFIQSRGRARMPNSDY	17
3	35	LEAITTERCQETFCYERLETLGDAFLKWVVSRHLF	20

The pI value of XP_003558946, XP_003553805, XP_003523657, XP_003550797, XP_002269915, XP_002264486, XP_002268369, XP_003559709, XP_003581414, EEE81952, XP_003566182, NP_001190348, ABB20894, AEE73926, XP_002515097, XP_003603236 and AEE77843 were less than 7 (pI<7) indicates that these dicer proteins were considered as acidic. The extinction coefficient (EC) indicates an amount of light absorbing of proteins at a certain wavelength. Extinction coefficient of DCLs at 280 nm is ranging from 126222.4 to 221324.8 $M^{-1}cm^{-1}$. The high extinction coefficient of DCLs (XP_002268369, XP_003553805 and XP_003558946) indicates presence of high concentration of Cys, Trp and Tyr. These amino acids (Trp, Tyr, Cys) are considered to be an important parameter in the calculation of extinction coefficient of proteins (Kumar and Bhalla 2011). The instability index is used to determine whether it will be stable in a test tube. If the index is less than 40, it is probably stable in the test tube. If the value is greater than 40, it is probably not stable (Guruprasad et al. 1990). The instability index value for the dicer proteins were found to be ranging from 21.90 to 47.14. The results imply XP_003553805 as stable protein (Table 2). The aliphatic index of a protein is a measure of the relative volume occupied by aliphatic side chain of the following amino acids; alanine, valine, leucine and isoleucine. The aliphatic index values of DCLs proteins ranging from 80.19 to 97.03. The very high aliphatic index of all DCLs proteins supports the view that DCLs proteins may be stable for a wide range of temperatures. The lower thermal stability of XP_003553805, XP_002268369 and EEE81952 is suggestive of a more flexible structure when compared to other DCLs proteins. The GRAVY (Grand Average of Hydropathy) value for protein is calculated as the sum of hydropathy values of all the amino acids. A hydropathy scale which is based on the hydrophobic and hydrophilic properties of the 20 amino acids is used. GRAVY values of DCLs proteins were ranging from -0.109 to -0.487. The very low GRAVY index of DCLs EEE81952, XP_003553805 and XP_002268369 implies that these DCLs could result in a better interaction with water.

The functions of DCLs proteins were analyzed by submitting the amino acid sequence to Prosite server. Particular cluster of residue types named pattern, motif, signature or fingerprint which are typically 10-20 amino acids in length. These regions thought to be important for biological function such as enzyme catalytic sites, prostethic group attachment sites, amino acids involved a metal ion and disulphide bonds (Sigrist et al. 2002). Prosite analysis suggested the functionality of these proteins with profiles and patterns identified for characteristic functionality were represented in Table 4. Profile analysis of DCLs protein sequences revealed the presence of Ribonuclease III domain coherently. A total of three motifs were determined in 20 DCLs protein sequences by using MEME. The motifs with width and best possible match amino acid sequences are shown in Table 5. Using the NetPhos 2.0 Server the putative phosphorylation sites were identified for the plant species with a strong prediction score above 0.8 (Table 3). The output score was given in 0.000-1.000 range and the score above the threshold (0.500) shows the confidence rate of true phosphorylation site by the server. Several putative phosphorylation sites are completely conserved in plant species, interestingly more phosphorylation sites were found in *Glycine max* (Fig. 1), *Vitis vinifera* and *Arabidopsis thaliana.*

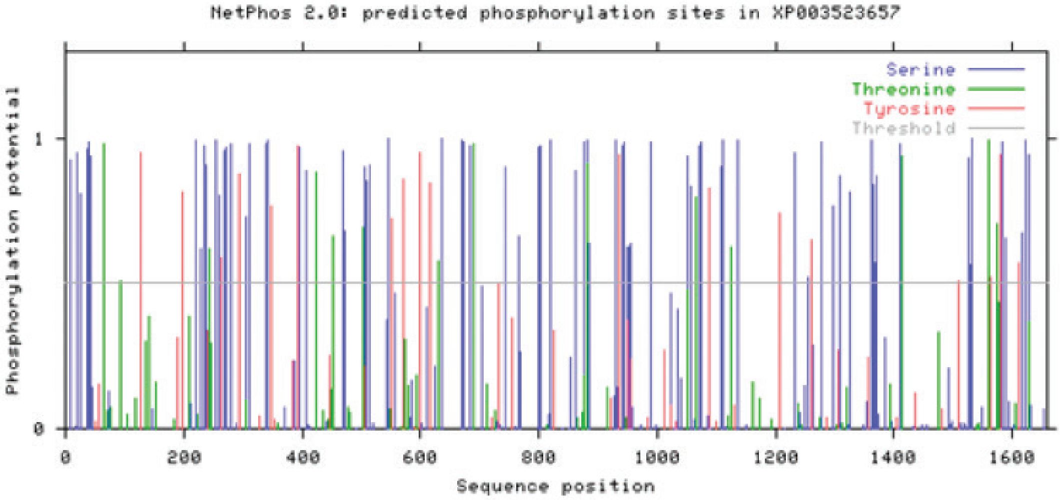

Figure 1. Putative phosphorylation sites in *Glycine max* predicted by NetPhos 2.0 Server. The same phosphorylation site prediction is done for the other plant species used in the study (data is not shown). The graph shows the prediction score just above the threshold (0.5) as putative phosphorylation sites.

The evolutionary relationships between the plants were evaluated by phylogenetic analysis of the aligned amino acids sequence of DCLs protein with neighbor-joining (NJ) method (Fig. 2).

This result agrees with previous studies (Zhang et al. 2004; Liu et al. 2009). Convergence and divergence are two essential phylogenetic properties, which can be useful to identify the closely as well as distantly related group containing plant DCLs proteins. The minimum degree of divergence was found to be 0.082 between *Populus trichocarpa* and *Ricunus communis*, while the maximum degree of divergence was found to be 1.244 between *Glycine max* and Arabidopsis *thaliana*.

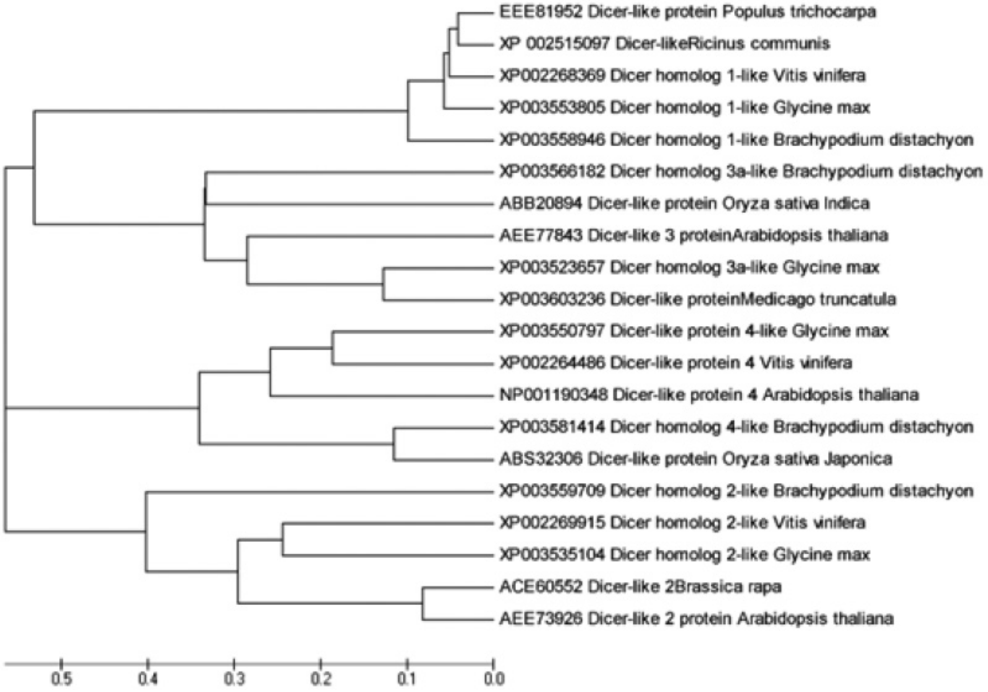

Figure 2. The phylogenetic tree of DCLs proteins from some plants species.

The phylogenetic tree showed that there are three major clusters and same plant DCLs proteins are not grouped together like *Brachypodium distachyon, Glycine max, Vitis vinifera* and *Arabidopsis thaliana.* Furthermore, similar DCLs protein sequences (DCL1, DCL2, DCL3 and DCL4) of different plant species were clustered together; such as for Dicer homolog 1- like *Vitis finifera*, *Glycine max* and *Brachypodium distachyon*, for Dicer homolog 2- like *Glycine max*, *Arabidopsis thaliana*, *Brassica rapa*, *Vitis finifera* and *Brachypodium distachyon*, for Dicer homolog 3-like *Brachypodium distachyon*, *Glycine max* and *Arabidopsis thaliana*, for Dicer homolog 4-like *Glycine max*, *Vitis finifera*, *Arabidopsis thaliana* and *Brachypodium distachyon.* This finding suggests that Dicer-catalyzed dsRNA processing is conserved and all Dicers evolved from a common ancestral enzyme (Macrae et al., 2006). All four classes of DCL genes (DCL1, DCL2, DCL3 and DCL4) were identified in rice and also *Arabidopsis* and *Populus trichocarpa* share more than 90% similarity at the amino acid level in DCL2 genes (Kapoor et al., 2008). Margis et al. (2006) reported that a fifth type of Dicer seems to have evolved in monocots. Also, in the present study the first major cluster consists of two subgroups which have DCL1 and DCL3 proteins. Liu et al. (2009) suggested that the function of plant DCLs were significantly divergent from each other and especially DCL1 is found to be strongly divergent from other DCL family members but this hypothesis is not consistent with these study results. In conclusion, in silico sequence analysis of DCLs proteins showed that these higher plants have been related together evolutionarily and conserved regions. In silico analysis of DCLs protein sequences would contribute to a better understanding of functional divergence within DCLs protein families in other plants.

References

Bartel D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 116: 281–297.

Bernstein E., Caudy A.A., Hammond S.M. and Hannon G.J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, 409: 363-366.

Blom N., Gammeltoft S. and Brunak S. (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of Biology*, Vol: 294 (5) pp: 1351-1362.

Chen X. (2005) microRNA biogenesis and function in plants. *FEBS Letters*, 579: 5923–5931.

Dhar P., Ganguli S. and Datta A. (2010) Molecular modeling of DICER and identification of phosphorylation sites. *Bioinformation*, 4(9): 412-416.

Falquet L., Pagni M., Bucher P., Hulo N., Sigrist, C.J. A. Hofmann K. and Bairoch A. (2002) The PROSITE database, its status in 2002. *Nucleic Acids Research*, 30: 235-238.

Gasteiger E. (2005) Protein Identification and Analysis Tools on the ExPASy Server. In: John M. Walker ed. *The Proteomics Protocols Handbook Humana Press*, 571-607.

Gill S.C. and Von Hippel P.H. (1989) Extinction coefficient. *Analitical Biochemistry*, 182: 319-328.

Groβhans H. and Filipowicz W. (2008) The expanding world of small RNAs. *Nature*, 451:414–416.

Guruprasad K., Reddy B.V.P. and Pandit M.W. (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Engineering Design and Selection*, 4: 155-164.

Ikai A.J. (1980) Thermo stability and aliphatic index of globular proteins. *The Journal of Bichemistry*, 88: 1895-1898.

Kapoor M., Arora R., Lama T., Nijhawan A., Khurana J.P., Tyagi A.K. and Kapoor S. (2008) Genome-wide identification, organization and phylogenetic analysis of Dicer-like, Argonaute and RNA-dependent RNA Polymerase gene families and their expression analysis during reproductive development and stress in rice. *BMC Genomics*, 9: 451 doi:10.1186/1471-2164-9-451.

Kim V.N. (2005) MicroRNA biogenesis: coordinated cropping and dicing. *Nature Reviews Moleculaer Cell Biology*, 6: 376–385.

Kumar N. and Bhalla T.C. (2011) In silico analysis of amino acid sequences in relation to specificity and physiochemical properties of some aliphatic amidases and kynurenine formamidases. *Journal of Bioinformatics and Sequence Analysis*, Vol. 3(6), pp. 116-123.

Kyte J. and Doolottle R.F. (1982) A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, 157: 105-132.

Lamontagne B., Larose S., Boulanger J. and Elela S.A. (2001) The RNase III family: a conserved structure and expanding functions in eukaryotic dsRNA metabolism. *Current Issues in Molecular Biology*, 3: 71-78.

Liu Q., Feng Y. and Zhu Z. (2009) Dicer-like (DCL) proteins in plants. *Functional & Integrative Genomics*, 9: 277–286.

MacRae I.J. and Doudna J.A. (2007) Ribonuclease revisited: structural insights into ribonuclease III family enzymes. *Current Opinion in Structural Biology*, 17: 1–8.

MacRae I.J., Zhou K., Li F., Repic A., Brooks A.N., Cande W.Z., Adams P.D. and Doudna J.A. (2006) Structural Basis for Double-Stranded RNA Processing by *Dicer. Science*, Vol 311: 195-197.

Margis R., Fusaro A.F., Smith N.A., Curtin S.J., Watson J.M., Finnegan E.J. and Waterhouse P.M. (2006) The evolution and diversification of Dicers in plants. *FEBS Letters*, 580: 2442–2450.

Matzke M.A. and Birchler J.A. (2005) RNAi-mediated pathways in the nucleus. *Natural Reviews Genetics*, 6(1): 24-35.

Moissiard G., Parizotto E.A., Himber C. and Voinnet O. (2007) Transitivity in *Arabidopsis* can be primed, requires the redundant action of the antiviral Dicer-like 4 and Dicer-like 2, and is compromised by viral-encoded suppressor proteins. *RNA*, 13: 1268–1278.

Sahay A. and Shakya M. (2010) In silico Analysis and Homology Modelling of Antioxidant Proteins of Spinach. *Journal of proteomics & bioinformatics*, Volume 3(5): 148-154.

Sivakumar K., Balaji S. and Radhakrishnan G. (2007) In silico characterization of antifreeze proteins using computational tools and servers. *The Journal of Chemical Sciences*, Vol. 119, No. 5: 571–579.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M. and Kumar S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*, 28: 2731-2739.

Timothy L., Mikael Bodén B., Buske F.A., Frith M., Grant C.E., Clementi L., Ren J., Li W.W. and Noble W.S. (2009) MEME SUITE: tools for motif discovery and searching *Nucleic Acids Research*, 37: 202-208.

Zhang H., Kolb F.A., Jaskiewicz L., Westhof E. and Filipowicz W. (2004) Single processing center models for human Dicer and bacterial RNase III. *Cell*, 118: 57–68.