Plant Archives Vol. 18, Special Issue (ICAAAS-2018), 2018 pp. 61-68 ***ISSN 0972-5210***

GENOME WIDE *IN SILICO* CHARACTERIZATION OF DIRIGENT PROTEIN FAMILY IN FLAX (*LINUM USITATISSMUM* L.)

Pooja Archana Sahani[1], Megha Ujinwal[1], Sangeeta Singh[2] and Nisha Singh[1*]

[1]Department of Computational Sciences (Bioinformatics), Central University of Punjab, Bathinda-151 001 (Punjab), India.
[2]National Research Centre on Plant Biotechnology, IARI, New Delhi - 110 012, India.

Abstract

Flax (*Linum usitatissimum* L.) is an important agricultural crop grown worldwide both for its seed oil as well as its stem fiber. Flax seeds are well-known for their nutritional and nutraceutical attributes such as omega-3 fatty acids in the oil and lignans and mucilage from the seed coat. It has high health benefits against certain allergies, cardiovascular diseases and cancer. Flax dirigent proteins (*DIR*) have roles in biosynthesis of lignan and are inducible by various types of abiotic and biotic stress factors, including secondary metabolism, fibers biosynthesis, and pathogen resistance. Lignan is the key component of seed protein and has vital role in seed development.In present study, first report on *insilico* analysis of six dirigent protein (DP) *(DIR1*, *DIR2*, *DIR3*, *DIR4*, *DIR5* and *DIR6*) family in flax and comparative analysis with other oil seed family DPs protein including *Brassica campestris*, *Helianthus annuus*, *Glycine max* and *Arabadopsis thaliana*. DP gene family members encoding (–) and (+) pinoresinol-forming DPs and their associated downstream metabolic processes involved in early seed coat development stages. *In silico* analysis of *DIRs* protein revealed that it comprises high level of alanine, Phenylalanine and threonine across six *DIR* proteins. Whereas, in *DIR4* and *DIR6* possess high level of leucine. These *DIRs* protein have high hydrophobicity and there secondary and tertiary structure shown that it has more beta pleated sheet both parallel and antiparallel. There are 3 significant and common motif found in these six *DIRs* proteins. These conserved motifsand active sites can be further target for drug designing aspect. The phylogenetic analysis of flax *DIRs* family protein shown kinship with *Arabadopsis thaliana* and *Helianthus annuus.* This study will be highly informative and useful for effective genetic enhancement of flax seed development, oil profiling and development of disease resistance flax varieties.

Key words : Abiotic stress, dirigent protein, flax, lignan, pinoresinol.

Introduction

Flax (*Linum usitatissmum* L.) is a self-pollinated diploid annual (2n=2x=30) crop that belongs to family *Lineacae* also known as linseed. Flax seed is one of the ancient crops, had cultivated since the beginning of civilization. Now more than 200 species in *Linum* have been reported with small flat seeds varying from radish brown colour to golden yellow colour (Rubilar *et al.*, 2010). The genome size of flax comprises of ~700 Mbps (Wang *et al.*, 2012). Flax seed is one of important functional food because the presence of plenty amount of alpha-linolenic acid (ALA) *i.e.*, short chain poly-unsaturated omega-3 fatty acid, soluble and insoluble fiber, proteins and an array of anti-oxidants (Riediger *et al.*, 2009). The flax seed are oil rich in unsaturated fatty acids, which have high level of tocopherols are now added into various food products and available as neutraceutical supplements in many countries (Lampi *et al.*, 2002). The majority of the oilseed crops contain LIO but LIN is present only in oils from certain fish, microalgae and crops such as canola (rapeseed or oilseed rape) and linseed. The stem of flax yields good quality of fibre having durability and high strength (Singh *et al.*, 2011). Flax seed contain total dietary fiber about 400 gm/kg. These fibers are rich in pentosans and contain 2-7% mucilage. The seeds mainly composed of ~45% oil, 25% protein, 30% dietary fibres and 75% unsaturated fatty acid (Rabetafika *et al.*, 2011). It is having anti-cancerous and phyto-estrogenic property. Other major constituents are the storage protein that can range from 10-30%; albumin 20–42% and globulins are the major storage protein of flax seed contains 66% of total seed storage protein.

Some epidemiological studies indicated that phyto-estrogen rich foods reduced the risk of osteoporosis, heart diseases, hormone dependent, cancers, diabetics and constipation (Krajcova *et al.*, 2009; Toure and Xueming, 2010). So, Flax seed has a huge role in health benefit as it is a good source of phyto-estrogen. Flax seed was also used in manufacturing of drying oil, paints, coatings and printing inks. But mainly the flax seed is used to produce oil and it is edible because of its high neutraceutical values. Among the presence of varieties of protein in flax seed, the lignin (phyto-alexin) protein play a vital role. Lignan are phenolic compound of 2-cinnamic acid residues (Oomah, 2001). It acts as both anti-oxidant and phyto-estrogens. Flax contains approximately 700-800 times more lignan than other plant food. Lignan content in flax seed is basically composed of matairesinol (0.54mg/100g), Pinoresinol (3.33mg/100g), secoisolariciresinoldi-glucoside (295-700mg/100g), p-coumaric and feriulic acid and other phenolic compounds (Morris, 2007; Toure and Xueming, 2010). As pinoresinol is one of the minor product of lignan, but having a great dietary value in flax seed. The content of pinoresinol depends on dirigent protein(DIR), because DIR is biosynthetic acid of pinoresinol. The dirigent was discovered first in *Forsythia intermedia*. It impart stereo-selectivity on the phenoxy radical-coupling reaction, which yield optically active lignan from two molecules of coniferyl alcohol in the biosynthesis of lignin alkaloids and flavonolignan it play a major role in plant secondary metabolism. DIR also resembles with the DRR 206 (Disease Resistant response protein 206). This DRR 206 is induced by the metabolite pinoresinolmonoglucoside and is involved in lignin biosynthesis in a defence response.

The discovery of this unique protein specify precisely about the biochemical outcome of phenoxy radical coupling. The term dirigent derived from the name “DIRIGERE” (a Latin word means to guide or align) was coined to define the first example of an apparently new class of protein. Therefore here, analysed the sequence, structure, function, phylogeny and interaction of dirigent family proteins in flax with different oil seed crops. With the advancement of next generation sequencing technologies (NGS), whole genome sequences are becoming publicly available for *in silico* analysis. With the help of these techniques such as NGS and bioinformatics analysis gene mining data have been successfully applied to identify and characterize plant gene families. Here, we report on genome-wide *in silico* and comparative analyses of six dirigentfamily protein in flaxwhich are essential in lignan biosynthesis pathways and in a defence response for effective genetic improvement of flax seed oil profiles.

Materials and Methods

Identification of dirigent proteins family sequences in flax

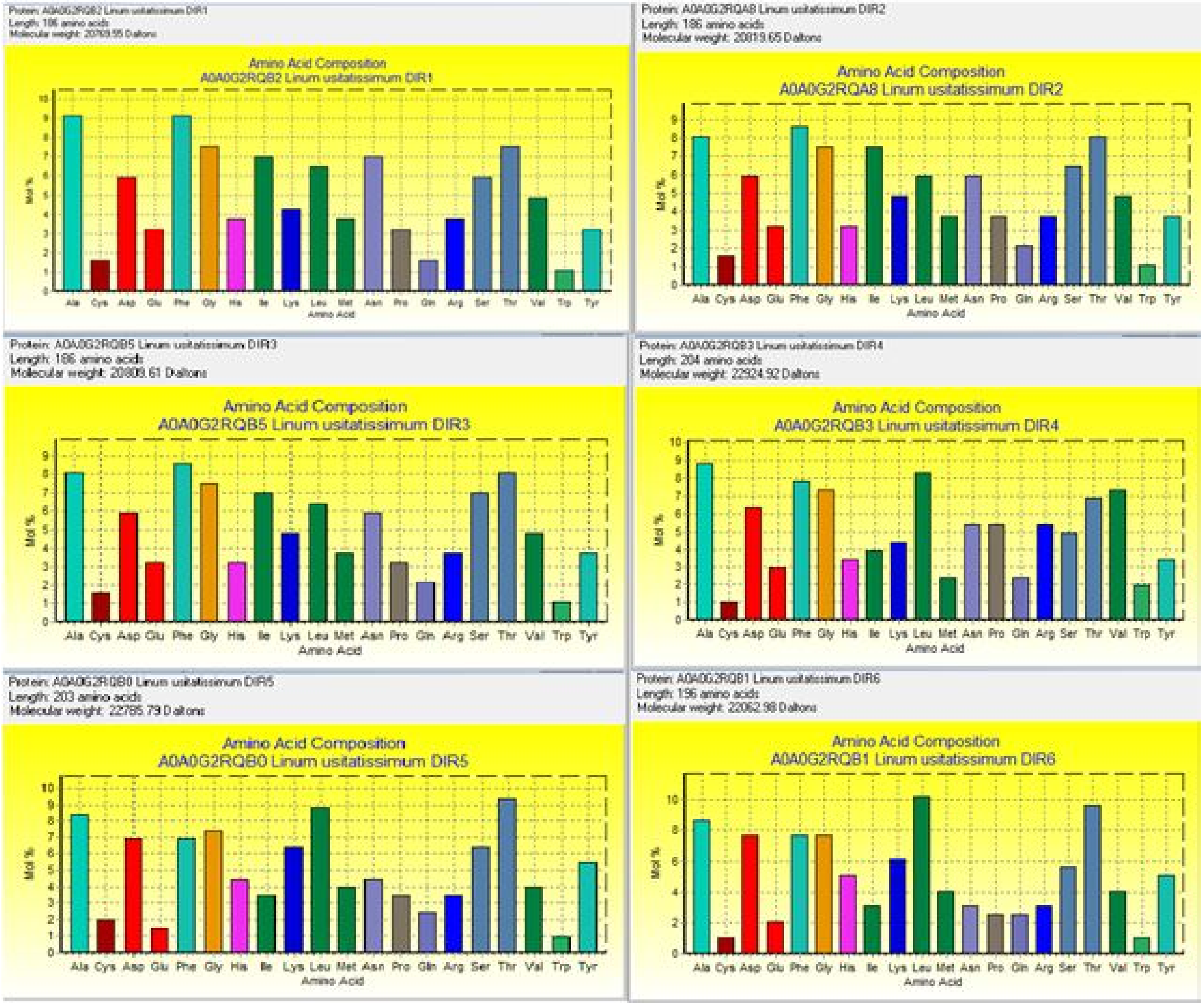
The amino acid sequence of sixdirigent protein (*DIR1*, *DIR2*, *DIR3*, *DIR4*, *DIR5* and *DIR6*) of flax seed were downloaded from UniProt database (http://www.uniprot.org/). Analysed the primary structure of all these six *DIRs* family proteins including its amino acid composition, molecular weight, hydrophobicity and hydrophilicity, using BioEdit sequence alignment tool (Hall *et al*., 1999). Amino acid composition of six DIRs proteins have been calculated to give sets of values representing single or average determination on 80 separate proteins, as residues of amino acid/100 residues in each protein. These values can be plotted on histograms, recommended that the occurrence of an individual amino acid in proteins could be specified by a mean value and its standard deviation. Hydrophobicity and hydrophilicity of amino acid of different dirigent protein (*DIR1*, *DIR2*, *DIR3*, *DIR4*, *DIR5* and *DIR6*) estimated by Kyte and Doolittle scale mean hydrophobic scale and Bokyo scale mean hydrophobicity profile method in Bio Edit. Graph is plotted with X-axis and Y-axis. X-axis show position of each amino acid in sequence and Y-axis show mean hydrophobicity and mean hydrophilicity, respectively. The more positive the value, the more hydrophobic are the amino acids situated in that region of the protein.

Multiple sequence alignment and phylogenetic tree construction

Clustal Omega was used for Multiple Sequence Alignment (MSA) of six diverse dirigent proteins of flax seed with four other oil seed dirigent proteins crops including *Arabidopsis thialiana*, *Glycine max*, *Brassica campestris*, *Helianthus annuus*. There was a remarkable difference in sequence after performing MSA, similarity along with the sequence length of these proteins. Dirigent proteins found in Flax (*Linum*) are *DIR1*, *DIR2, DIR3, DIR4, DIR5* and *DIR6* were compared with other four oil seed crops including *Arabidopsis thialiana, Brassica campestris, Helianthus annuus* and *Glycine max.* A phylogenetic tree were constructed based on MSA to illustrated their evolutionary relationship with each other and characterized them in a group.

Motif analysis and protein-protein interaction

To identify significant motif or conserve domain of *DIRs* flax proteins MEME tools (meme-suite.org/tools/meme) was used (Bailey *et al.*, 2009). MEME represents motifs which explain the probability of each position in sequences. The predicted distinct motifs with similarity


Protein: A0A0G2RQB2 Linum usitatissimum DIR1
Length: 186 amino acids
Molecular weight: 20769.55 Daltons

Protein: A0A0G2RQA8 Linum usitatissimum DIR2
Length: 186 amino acids
Molecular weight: 20819.65 Daltons

Protein: A0A0G2RQB5 Linum usitatissimum DIR3
Length: 186 amino acids
Molecular weight: 20809.61 Daltons

Protein: A0A0G2RQB3 Linum usitatissimum DIR4
Length: 204 amino acids
Molecular weight: 22924.92 Daltons

Protein: A0A0G2RQB0 Linum usitatissimum DIR5
Length: 203 amino acids
Molecular weight: 22785.79 Daltons

Protein: A0A0G2RQB1 Linum usitatissimum DIR6
Length: 196 amino acids
Molecular weight: 22062.98 Daltons

Fig. 1 : Graphical representation ofamino acid composition of six dirigent protein (*DIR1, DIR2, DIR3, DIR4, DIR5, DIR6*) showing its position in sequence and its molecularweight.

<60% are categorised in "Significant motif" section. The predicted motif was searched against sequence database for the occurrence of each motif per sequence by MAST (Motif-based sequence analysis tool) tool. In MAST, each sequence assigned a *p*-value. Based on *p*-value an individual motif found in sequence. The significant matches were listed against each motif in "Significant motifs" section. For protein-protein interaction, STRING (Search Tool for the Retrieval of Interacting Genes/ Proteins) tool (https://string-db.org/) was used against *Brassica campestris*.

Homology modelling of DIRs family protein

For DIRs family protein modelling Modeller was used (Sali *et al*., 2017). DIRs protein sequences were aligned to be model having template structure, the atomic co-ordinates of templates and a script file. An alpha-helix, beta-plated sheet and loop structure is obtained from the align sequence.

Results and Discussion

Analysis of primary, secondary and tertiary structureof flax dirigent family proteins

After analysing the primary structure the level of phenylalanine,alanine are more in *DIR*s proteins except *DIR4* and *DIR6*. Phenyalanine is more in case of *DIR1, DIR2*, *DIR3* but in case of *DIR4* there are more amount of alanine where as in *DIR5* its threonine and *DIR6* its leucine. In case of hydrophobicity and hydrophilicity the DIRs family is more hydrophobic and hydrophilic also due to the presence of both hydrophobic and hydrophilic amino acids like alanine, leucine and phenylalanine.

Analysing secondary and tertiary structure it conclude that DPs family is highly composed of beta pleat, so it is a beta sheet dominated protein family. By comparing different homology modelling structures we can analyse the number ofparallel and antiparallel beta sheets present

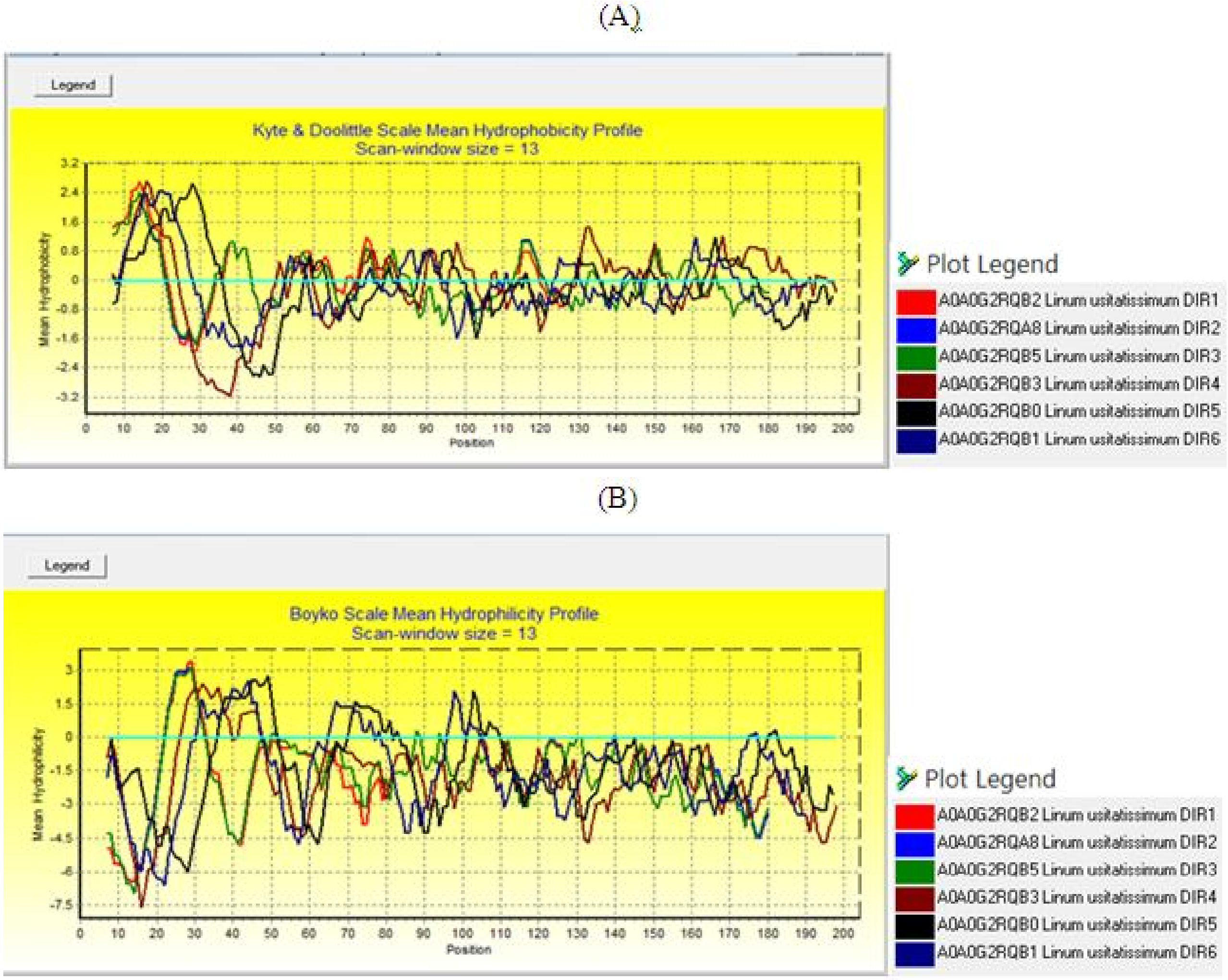


Fig. 2 : Graphical representation of hydrophobicity and hydrophilicity of six DIR family protein. (A) using, Kyte and Doolittle scale for mean hydrophobicity profile of dirigentprotein. (B) using, Bokyo scale for mean hydrophilicity profile. Six different dirigent proteins of flax were represented with red-*DIR1,* blue-*DIR2,* green-*DIR3,* purple-*DIR4,* black-*DIR5*and navy blue-*DIR6.*

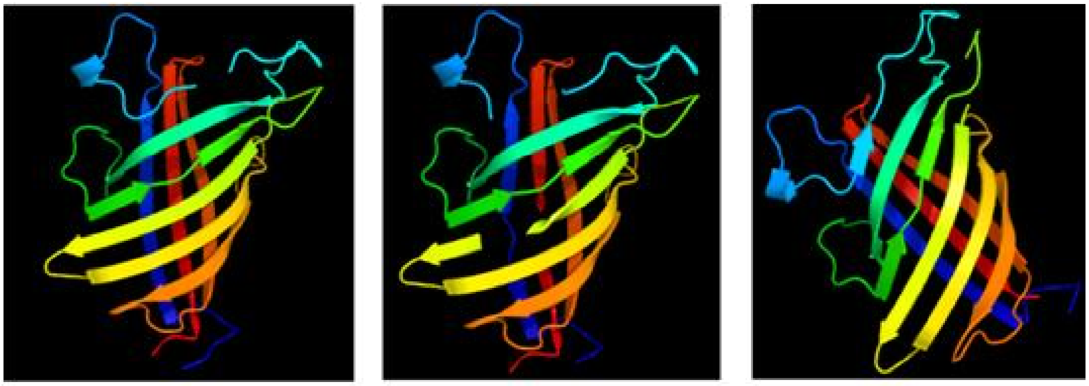
DPs family of different species.

Multiple sequence and Phylogenetic analysis

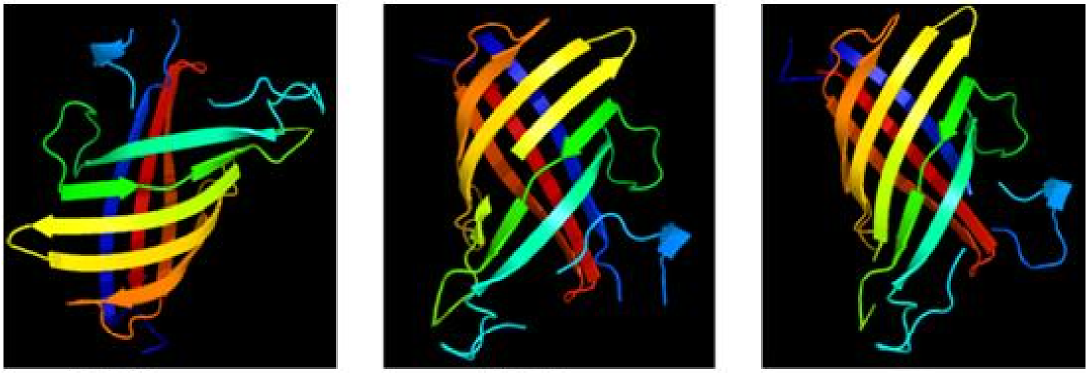
In multiple sequence alignment of dirigent protein of flax with other 4 oil seed dirigent protein, which shared the similarities and differences that indicate the evolutionary relationships between them like they shared common ancestors. For further analysis of the evolutionary relationships between these species we construct the phylogeny, the 5 different groups having common ancestry. In group 1 we found 5 similar protein sequences. The percentage of amino acid sequence similarities emphasize that the sequences like *DIR4* protein of *Linum*, DP protein of *Helianthus*, *DIR1*, *DIR2*, *DIR3* protein of Linum are similar to each other and they share their common ancestry. In group 2, we found also 5 similar protein sequences. The percentage of aminoacid sequence similarities emphasize that the sequences like *DIR5* and *DIR6* in *Linum*, *DIR5* and *DIR6* in Arabidopsis, *DIR* protein in Brassica are similar to eachother and share theircommon ancestry. In group 3 both the dirigent proteins are from Glycine family so they show similar sequence and belongs to same common ancestry.In group 4 the sequence similarities we can see in between *DIR3* of Arabidopsis and DP protein of Brassica.The 5th group having twoDP of Glycine and *DIR2* and *DIR1* of *Arabidopsis*, which shows sequence similarities and belongs to common ancestry.

Motif analysis

The motif and conserved sites across all six DP were detected by using "MEME" is 10 motifs where 9 motifs are functional and 1 is the non functional motif. Among the 9 functional motifsonly 3 motifs are found significant and common across six DP of flax and these are conserved across other oil seed crops. The amino acid



(A) *DIR1* (B) *DIR 2* (C) *DIR3*



(D) *DIR4* (E) *DIR5* (F) *DIR6*

Fig. 3 : Homology modelling (A-F) show parallel and anti-parallel beta-sheet structure in diverse six dirigent protein (*DIR1*-*DIR6*) of flax.

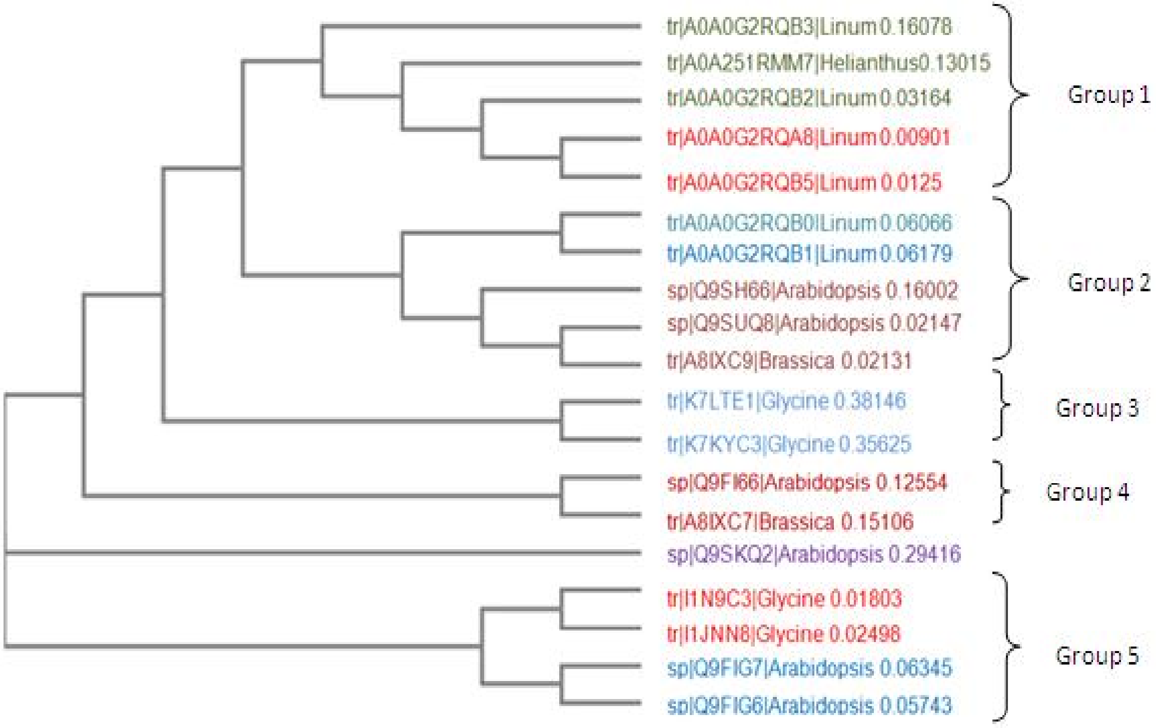
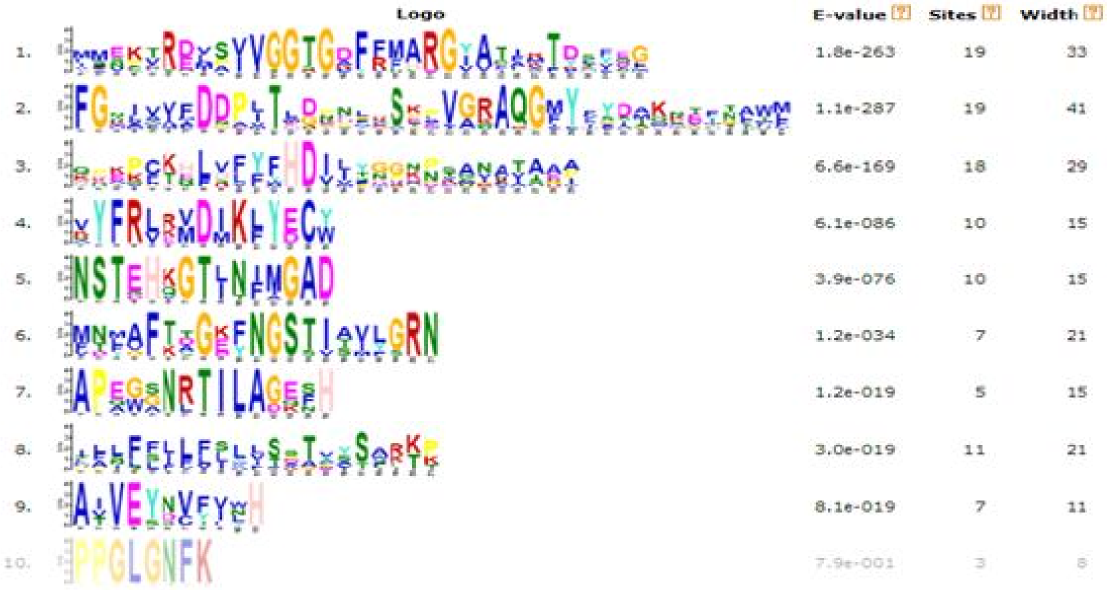


Fig. 4 : Phylogenetic analysis of DIR protein family across five different crop species (*Brassica campestris*, *Helianthus annuus*, *Glycine max* and *Arabadopsis thaliana*).



	Logo	E-value	Sites	Width
1.		1.8e-263	19	33
2.		1.1e-287	19	41
3.		6.6e-169	18	29
4.		6.1e-086	10	15
5.		3.9e-076	10	15
6.		1.2e-034	7	21
7.		1.2e-019	5	15
8.		3.0e-019	11	21
9.		8.1e-019	7	11
10.		7.9e-001	3	8

Fig. 5 : Representation of motifs and its conserved amino acid sequencesof all six DIRs in flax by different colours.

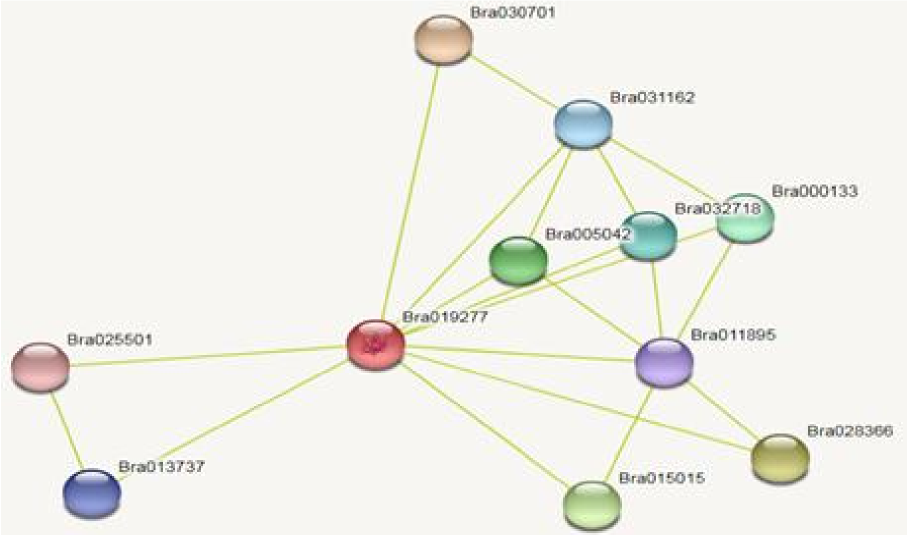


Fig. 6 : Protein protein interaction of six DIRs family protein of flax with *Brassica campestris*. It signifies high specificity between the dirigent protein of *Linum* and *Brasicae campestris*.

sequences are represented by different colours.

Analysis of protein-protein interaction

For protein-protein interaction, STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) tool (https://string-db.org/) were used. Protein sequence of all six DIRs proteins shown fine interaction against *Brassicacampestris*. According to bit-score, particular protein sequence is selected for interaction. One protein sequence of *B. campestris* shows great interaction with other protein sequence of other species. It plays a modulator key role in this interaction. Biological process, molecular function, cellular component and KEGG pathway are desired output that obtained from String. Biological process include number of protein folding is 5 in gene set. In carbohydrate metabolic process and primary metabolic process have 7 and 13 gene set. Molecular function includes binding and catalytic activity have 11 and 10 gene set. Cellular component include apoplast, cell wall and endoplasmic reticulum lumen have 9, 9 and 5 gene set.

Conclusion

Dirigent protein is one of the most important secondary metabolites in plant. It plays an important role in lignan biosynthesis. It helps to increase the amount of pinoresinol in lignan protein which is a nutritional product of flax seed. After the analysis of lignin preotein. Here, it shows that flax DIR famil proteins. *In silico* analysis of *DIRs* protein revealed that it comprises high level high hydrophobicity with high percentage of alanine, Phenylalanine, leucine and threonine across six *DIR* proteins of flax. Its secondary and tertiary structure expressed that it encompasses more beta pleated sheet both parallel and antiparallel then to helix and random coil. There are three significant and common motif found in these six *DIRs* proteins. These conserved regions and active sites of *DIRs* proteins will be useful for drug design against various kind of biotic stress. Based multiple sequence analysis and phylogenetic relationship of flax *DIRs* family protein revealed that *Arabadopsis thaliana* and *Helianthus annuus* are closer to flax. However, the evolution of dirigent protein and their function is not yet resolve as whether they only evolved during land plant adaptation (Pickel *et al.*, 2012). Furthermore, this analysis can be extensively useful for genetic improvement of flax seed development, and disease resistance flax varieties.

References

Bailey, T. L., M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi and W. S. Noble (2009). MEME SUITE : tools for motif discovery and searching. *Nucleic Acids Research*, **37** : W202-W208.

Bozan, B. and F. Temelli (2008). Chemical composition and oxidative stability of flax, safflower and poppy seed and seed oils. *Bioresource Technology*, **99(14)** : 6354-6359.

Chung, M. W. Y., B. Lei and E. C. Y. Li-Chan (2005). Isolation and structural characterization of the major protein fraction from NorMan flaxseed (*Linum usitatissimum* L.). *Food Chemistry*, **90(1-2)** : 271-279.

Davin, L. B. and N. G. Lewis (2000). Dirigent proteins and dirigent sites explain the mystery of specificity of radical precursor coupling in lignan and lignin biosynthesis. *Plant Physiology*, **123(2)** : 453-462.

Diederichsen, A. and K. Richards (2003). Cultivated flax and the genus *Linum* L. *Flax : the genus Linum*, 22-54.

Goyal, A., V. Sharma, N. Upadhyay, S. Gill and M. Sihag (2014). Flax and flaxseed oil : an ancient medicine & modern functional food. *Journal of Food Science and Technology*, **51(9)** : 1633-1653.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/ 98/NT. In : *Nucleic acids symposium series,* **41** : 95-98.

Kole, C. (2007). *Genome mapping and molecular breeding in plants*, **1** : 243-253.

Krajèová, A., V. Schulzová, J. Hajšlová and M. Bjelková (2009). Lignans in flaxseed. *Czech J. Food Sci.*, **27** : S252-S255.

Lampi, A. M., A. Kamal-Eldin and V. Piironen (2002). Tocopherols and tocotrienols from oil and cereal grains. *Functional Foods : Biochemical and Processing Aspect*, **2** : 1-38.

Mazza, G., J. Shi and M. Le Maguer (2016). *Functional Foods : Biochemical and Processing Aspects,* Volume **2**. CRC Press.

Morris, D. H. (2007). *Flax: A health and nutrition primer*. Flax Council of Canada.

Moyad, M. A. (2005). An introduction to dietary/supplemental omega-3 fatty acids for general health and prevention: Part II. *Urologic Oncology: Seminars and Original Investigations,* **23(1)** : 36-48. doi:10.1016/ j.urolonc.2005.03.001

Oomah, B. D. (2001). Flaxseed as a functional food source. *Journal of the Science of Food and Agriculture*, **81(9)** : 889-894.

Oomah, B. D. and G. Mazza (1993). Flaxseed proteins–a review. *Food chemistry*, **48(2)** : 109-114.

Oomah, B. D. and G. Mazza (1999). Health benefits of phytochemicals from selected Canadian crops. *Trends in Food Science & Technology*, **10(6-7)** : 193-198.

Paniagua, C., A. Bilkova, P. Jackson, S. Dabravolski, W. Riber, V. Didi and T. Hamann (2017). Dirigent proteins in plants: modulating cell wall metabolism during abiotic and biotic stress exposure. *Journal of Experimental Botany*, **68(13)** : 3287-3301.

Pickel, B. and A. Schaller (2013). Dirigent proteins: molecular characteristics and potential biotechnological applications. *Applied Microbiology and Biotechnology*, **97(19)** : 8427-8438.

Pickel, B., J. Pfannstiel, A. Steudle, A. Lehmann, U. Gerken, J. Pleiss and A. Schaller (2012). A model of dirigent proteins derived from structural and functional similarities with allene oxide cyclase and lipocalins. *The FEBS Journal*, **279(11)** : 1980-1993.

Rabetafika, H. N., V. Van Remoortel, S. Danthine, M. Paquot and C. Blecker (2011). Flaxseed proteins: food uses and health benefits. *International Journal of Food Science & Technology*, **46(2)** : 221-228.

Riediger, N. D., R. Othman, E. Fitz, G. N. Pierce, M. Suh and M. H. Moghadasian (2008). Low n-6:n-3 fatty acid ratio, with fish- or flaxseed oil, in a high fat diet improves plasma lipids and beneficially alters tissue fatty acid composition in mice. *European Journal of Nutrition*, **47(3)** : 153-160. doi:10.1007/s00394-008-0709-8

Rubilar, M., C. Gutiérrez, M. Verdugo, C. Shene and J. Sineiro (2010). Flaxseed asa Source of Functional Ingredients. *Journal of Soil Science and Plant Nutrition*, **10(3)**. doi:10.4067/s0718-95162010000100010

Sandhya, N. K. S. and V. Rai (2018). Identification of cis-regulatory elements associated with salinity and drought stress tolerance in rice from co-expressed gene interaction networks. *Bioinformation*, **14(3)** : 123-131.

Singh, K. K., D. Mridula, J. Rehal and P. Barnwal (2011). Flaxseed: a potential source of food, feed and fiber. *Critical Reviews in Food Science and Nutrition*, **51(3)** : 210-222.

Touré, A. and X. Xueming (2010). Flaxseed lignans: source, biosynthesis, metabolism, antioxidant activity, bio active components, and health benefits. *Comprehensive Reviews in Food Science and Food Safety*, **9(3)** : 261-269.

Vaisey-Genser, M. and D. Morris (1997). Flaxseed–Health, Nutrition and Functionality Flax Council of Canada Revised Edition. *Boehringer Mannheim: Winnipeg, MB*.

Vaisey-Genser, M. and D. H. Morris (2003). Introduction: history of the cultivation and uses of flaxseed. In : *Flax* (pp. 13-33). CRC Press.

Venglat, P., D. Xiang, S. Qiu, S. L. Stone, C. Tibiche, D. Cram and F. Bekkaoui (2011). Gene expression analysis of flax seed development. *BMC Plant Biology*, **11(1)** : 74.

Webb, B. and A. Sali (2017). Protein Structure Modeling with MODELLER. In : *Functional Genomics* (pp. 39-54). Humana Press, New York, NY.

Westcott, N. D. and A. D. Muir (1996). Variation in the concentration of the flax seed lignan concentration with variety, location and year. *Proc. of the Flax Institute of the United States, Fargo, ND: Flax Institute of the United States*, **56** : 77-80.

Westcott, N. D. and A. D. Muir (1996). Variation in the concentration of the flax seed lignan concentration with variety, location and year. *Proc. of the Flax Institute of the United States, Fargo, ND: Flax Institute of the United States*, **56** : 77-80.

Westcott, N. D. and A. D. Muir (2003). Chemical studies on the constituents of Linum spp. *Flax the genus Linum. New York (USA): Taylor and Francis*, 55-73.