Identification and Functional Characterization an Ortholog of OsENOD93-1 Gene in Wheat Using in-silico Approach

Anuj Kumar¹,², Avneesh Kumar³, Pankaj Tyagi⁴* and Krishna Pal Singh¹,⁵

¹Advanced Centre for Computational and Applied Biotechnology, Uttarakhand Council for Biotechnology, Dehradun, Uttarakhand-248007, India
²Dr. A.P.J. Abdul Kalam Technical University, Lucknow, Uttar Pradesh-226031, India
³Department of Botany, Akal University, Talwandi Sabo, Bathinda, Punjab-151302, India
⁴Department of Biotechnology, Meerut Institute of Engineering and Technology, Meerut, Uttar Pradesh-250005, India
⁵Department of Molecular Biology, Biotechnology and Bioinformatics, College of Basic Sciences and Humanities, CCS Haryana Agricultural University, Hisar, Haryana-125004, India

*Corresponding author

A B S T R A C T

Nitrogen is an essential macromolecule necessary for suitable growth and development of plants by playing a vital role in biochemical and physiological functions throughout the life cycle. Early node 93 (ENOD93-1) gene is well known for significantly responding to both N induction and N reduction in plants, but little is known about the structure and function of the ENOD93-1 protein in cereal crops. In this study, a putative ortholog of rice OsENOD93-1 gene was identified in the wheat genome. BLASTP analysis showed the OsENOD93-1 shared the 80% similarity with wheat TRIAE_CS42_4AL_TGACv1_288261_AA0942630.1 scaffold. Present study describes the location of TaENOD93-1 gene on4AL chromosome of wheat. MicroRNA target analysis revealed that TaENOD93-1 gene targeting by miR6201. Phylogeny tree suggested that identified putative ortholog shared the evolutionary relationship with monocots. The three dimensional structure of TaENOD93-1 protein was predicted using computer-aided molecular modeling technique followed by structure evaluation using Ramachandran plot. The primary goal of this study was to mapping the chromosomal location and functional annotation of TaENOD93-1 gene in wheat using bioinformatics approach. The identification of putative ortholog of OsENOD93-1 gene in wheat has opened up new opportunities to improve the nitrogen use efficiency (NUE) in wheat through molecular breeding approach.

K e y w o r d s
Nitrogen, Early node, In-silico, Chromosomal mapping, Functional annotation, Homology modeling
Introduction

Nitrogen (N) is a valuable contributor to total plant biomass and a limiting factor in plant growth and development (Kraiser et al., 2011). It is an essential component of most biological molecules including proteins, enzymes, and metabolic products (Good et al., 2004). Nitrogen also acts as a signaling molecule that plays a central role in several physiological and developmental processes including: root development, seed dormancy, leaf expansion and flowering time (Bouguyon et al., 2012; Balyan et al., 2016). Along with several phenotypic changes such as expanded root architecture, reduced shoot biomass production, chlorosis, leaf discoloration, and impaired reproduction, N limitation contributes to reduce crop growth and yield (Hawkesford and Barraclough, 2011). In order to ensure sufficient N in soil, farmers frequently supply fertilizers containing nitrate (NO$_3^-$), ammonium (NH$_4^+$), or urea [(CONH)$_2$] (Miller et al., 2007). Crop plants uptake about 30-40% of total applied N. Remaining fertilizer is lost via leakage into the atmosphere, groundwater, lakes and rivers (Raun and Johnson, 1999; Glass, 2003). Waste amount of fertilizer is responsible for serious environmental pollution in the form of aerosolized nitrous oxides and leaching of soluble nitrates into waterways (Johnson et al., 2007; Doney, 2010; Montzka et al., 2011; Dechorgnat et al., 2011). In soil, the content of organic and inorganic N sources are available, which are heterogeneous and dynamic under natural conditions. Contents of N also depend on several abiotic (temperature, pH, chemical properties) and biotic (presence of microorganisms) factors (Balyan et al., 2016). In higher plants, regulation of N-metabolism is highly complex and influenced by various physiological and metabolic processes including sucrose synthesis and transport, circadian rhythms, key metabolic level [i.e. glutamine (Gln), and NO$_3$ itself] (McAllister et al., 2012). Improving nitrogen use efficiency (NUE) in crop plants is a challenging task for scientific community for the development of sustainable agriculture (Good et al., 2004; McAllister et al., 2012). NUE improvement through genetics approaches may play a significant role to fulfill the yield gap and beneficial for caloric intake of billions of people over the world (Han et al., 2015). A fairly large number of candidate genes are involved in NUE. These genes play a critical role in different biological processes associated with NUE such as uptake, assimilation, and utilization efficiency (Han et al., 2016). Early nodulation 93 (ENOD93) gene play a significant role in cell wall reorganization during nodulation in legumes (Li et al., 2016). ENOD93 is an active member of early nodulin-like proteins family. Members of this family play an important role during several developmental stages including: pollen formation, secondary cell wall deposition, germination, senescence, etc (Denancé et al., 2014). Seven nodulin-like protein families have been predicted in non-leguminous plant species. Prediction of these families reflects that they probably play an important role in a wide variety of physiological processes (Borner et al., 2003; Khan et al., 2007; Mashiguchi et al., 2009). In rice genome, six ENOD domain containing genes have been reported on same chromosomal region along with similar molecular weight and pI. All these rice ENOD93 proteins expected to be localized in mitochondria. OsENOD93-1 gene consists of three exons and two introns, encodes a 116 amino acids long protein. Zhu et al., 2003 revealed that OsENOD93-1 showed a high level of expression in root, especially at the panicle emergence stage. Bi et al., 2009 suggested that OsENOD93-1 gene significantly responded to both N induction and N reduction. Overexpression of OsENOD93-1 in transgenic rice plant increased the shoot dry biomass and seed

3241
yield. Li et al., (2016) suggested that ENOD93 gene showed up-regulation and promoted nitrogen fixation in faba beans. Despite the critical interplay of ENOD93-1 gene in N regulation, no efforts were made in past to identify ENOD93-1 gene in wheat. The recent availability of wheat reference genome sequences allowed us to identify putative ortholog of OsENOD93-1 gene in the wheat genome.

In view of the above, the current study was aimed at the chromosomal mapping and functional annotation of TaENOD93-1 gene using bioinformatics approach. We also modeled the 3D structure of TaENOD93-1 using homology modeling followed by structural evaluation and regulatory networks.

**Materials and Methods**

**In-silico chromosomal mapping of TaENOD93-1 in wheat genome**

**In-silico** chromosomal mapping was performed to identify the putative ortholog of OsENOD93-1 gene in wheat genome. Full length protein sequence of OsENOD93-1 (MSU: LOC_Os06g05010) was extracted from Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/) in FASTA format. In order to conduct the similarity searches BLASTP algorithm was performed against wheat genome assembly TGACv1 publically available on EnsemblPlants platform (Bolser et al., 2016). The results of BLAST analysis were used to identify ‘putative’ wheat ortholog using the criteria followed by Kumar et al., (2016a, 2016b), Nagaraju et al., (2018), Supriya et al., (2018), Mathpal et al., (2018). Conserved domain search (CD-Search) program was used to predict functional domain inside the protein sequence of OsENOD93-1 and its wheat ortholog TaENOD93-1 (Marchler-Bauer et al., 2017). Sub cellular localization of TaENOD93-1 protein was predicted using PLANT-mPLoc server (http://www.csbio.sjtu.edu.cn/cgi-bin/Plant mLoc.cgi). In order to validate the novelty of identified TaENOD93-1, we carried out BLAST searches against wheat and other plant species non-redundant proteins sequence on NCBI.

**Physicochemical analysis**

ProtParam (Gasteiger et al., 2005) was used to calculate the various physicochemical properties such as molecular weight (Mw), extinction co-efficient, half-life, theoretical pI, instability index, aliphatic index (AI), and grand average of hydropathicity (GRAVY).

**Phylogeny analysis and miRNA targeting TaENOD93-1**

Phylogeny analysis was performed by multiple sequence alignment (MSA) using ClustalX server 2.1 (Larkin et al., 2007). MEGA 6.0 (Tamura et al., 2013) was used to obtain the phylogenetic trees using the neighbor-joining method (Saitou et al., 1987). The phylogenetic tree involved 20 ENOD93 proteins from different plants. Bootstrap values for the phylogeny tree was calculated as percentage of 1000 iterations (Felsenstein, 1985). The evolutionary distances (expressed as number of amino acid differences per site) were computed using the number of differences method (Nei and Kumar, 2000). The miRNA targeting the TaENOD93-1 gene was predicted employing web-based psRNA Target server (Dai and Zhao, 2011) using default parameters.

**Homology modeling and structure evaluation**

The three dimensional (3D) structure of TaENOD93-1 protein was deduced using homology modeling based approach. Firstly, PSI-BLAST was carried out against protein
data bank (PDB) (https://www.rcsb.org/) and SwissProt template library (SMTL) to find out the suitable homologous template structures. Homologous template was selected based on the criteria followed by Kumar et al., (2013), Kumar et al., (2017), Gajula et al., (2016) and Jee et al., (2017). Swiss-Model server (Biasini et al., 2014) was used to predict 3D structure of TaENOD93-1 protein. Modeled 3D structure of TaENOD93-1 protein was calculated by using phi (Φ) and psi (Φ) torsion angles and covalent quality using PROCHECK through PSVS server (http://psvs-1.5-dev.nesg.org/). Evaluated structure was further rendered by UCSF CHIMERA 1.10 (Pettersen et al., 2004) in different 3D coordinates.

**Prediction of regulatory partners of ENOD93-1 protein**

Since only information on rice OsENOD93-1 protein and its target was available in STRING database (https://string-db.org/), one protein-protein interaction network was prepared. The associations in STRING database include direct (physical) interactions as well as indirect (functional) interactions (Szklarczyk et al., 2017).

**Results and Discussion**

**In-silico chromosomal mapping of TaENOD93-1 gene**

A putative wheat ortholog of OsENOD93-1 gene was identified and functionally annotated using different computational biology based algorithms. The protein sequence of OsENOD93-1 protein was aligned against the wheat sequences available on EnsemblPlants. BLASTP results showed maximum 80% similarity with TRIAE_CS42_4AL_TGACv1_288261_AA0942630.1 of chromosome 4AL. Mapping results showed that TaENOD93-1 gene has complete gene structure including transcription starting site (TSS) along with polyA tail. TaENOD93-1 gene coded for 121 amino acid long protein. CD-search for predicted gene suggested that presence of ENOD93 domain (Pfam03386) belonging to ENOD93 protein superfamily is similar to the OsENOD93-1 gene (Figure 1). The machine-learning program PLNAT-mPLoc suggests nucleus localization of TaENOD93-1 gene encoded protein. BLASTP algorithm was performed against NCBI non-redundant protein database to find out the similarity between the identified TaENOD93-1 gene encoded protein and other plant protein sequences. Similarity results revealed that TaENOD93-1 protein did not share >80% sequence similarity with wheat and other plants species (Figure 2).

**Physicochemical analysis**

The molecular weight of TaENOD93-1 was found to be 12633.67 g/mol. With isoelectric point (pI) value at 10.43 indicating basic nature while instability index value was found to be at 35.49 which classifies that protein is stable. In this study, aliphatic index value was 85.04 (relatively higher value showed greater stability), hydropathicity value (return of GRAVY score) was found to be 0.148, revealing better interaction with a water molecule.

**Phylogeny analysis and miRNA targeting TaENOD93-1**

A phylogenetic tree was constructed to confirm the evolutionary relationship between identified TaENOD93-1 protein and other ENOD93 proteins reported from different plant species. TaENOD93-1 protein was found closely associated with AtENOD93 (Aegilops triuncialis), BdENOD93 (Brachypodium distachyon), and OsENOD93-1 (O. sativa) (Figure 3).
Fig.1 (a) Predicted function domain in protein sequence of TaENOD93-1; (b) Ribbon image of modeled structure; (c) Hydrophobic surface representation. The molecular graphic images were produced using the UCSF Chimera package from the resource for Biocomputing, Visualization, and Informatics (http://www.cgl.ucsf.edu/chimera) at the University of California, San Francisco (supported by NIH P41 RR-01,081); (d) Predicted Ramachandran plot for evaluated structure.

Fig.2 Circos representation of BLASTp result for TaENOD93-1 protein conducted against non-redundant protein database of NCBI.
**Fig. 3** Phylogenetic tree constructed by utilizing ENOD93 protein sequences from different plant species green colour represent the identified TaENOD93-1 protein

**Fig. 4** Protein-protein interaction network of OsENOD93-1 protein and its targets predicted using STRING database
Table 1 Summary of structure quality factors of TaENOD93-1 protein structure generated using PSVS 1.5 server

<table>
<thead>
<tr>
<th></th>
<th>Mean score</th>
<th>SD</th>
<th>Z-score&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procheck G-factor&lt;sup&gt;e&lt;/sup&gt; (phi / psi only)</td>
<td>-0.24</td>
<td>N/A</td>
<td>-0.63</td>
</tr>
<tr>
<td>Procheck G-factor&lt;sup&gt;e&lt;/sup&gt; (all dihedral angles)</td>
<td>-0.23</td>
<td>N/A</td>
<td>-1.36</td>
</tr>
<tr>
<td>Verifv3D</td>
<td>0.06</td>
<td>0.0000</td>
<td>-6.42</td>
</tr>
<tr>
<td>ProsaII (-ve)</td>
<td>-0.14</td>
<td>0.0000</td>
<td>-3.27</td>
</tr>
<tr>
<td>MolProbity clash score</td>
<td>13.25</td>
<td>0.0000</td>
<td>-0.75</td>
</tr>
</tbody>
</table>

<sup>a</sup>Residues selected based on: Dihedral angle order parameter, with $S(\phi)+S(\psi)<1.8$

<sup>b</sup>Selected residue ranges: 66A-104A

<sup>c</sup>With respect to mean and standard deviation for a set of 252 X-ray structures < 500 residues, of resolution <= 1.80 Å, R-factor <= 0.25 and R-free <= 0.28; a positive value indicates a 'better' score

But wide variation was noticed with that of legume and dicot plants, indicating that *TaENOD93-1* gene encoding protein is derived from its common ancestor *B. distachyon* and *O. sativa*. MicroRNA target analysis carried by psRNA Target server revealed that *TaENOD93-1* gene by miR6201-3p. It has been reported that miR6201 potentially involved in embryogenic callus formation and somatic embryogenesis (Chu et al., 2016)

**Homology modeling and structure evaluation**

Automated homology modeling based Swiss-Model server was used to build a 3D structure of TaENOD93-1 protein as mentioned in the methods section. TaENOD93-1 protein shared the higher similarity with the known template (PDB: 2mdt.A.1) derived from the SMTL library as a result of PSI-BLAST. The predicted model showed the closest Ca RMSD (root-mean-square deviation) with respect to the corresponding template upon superposition. Modeled 3D structure of TaENOD93-1 protein was submitted to Protein Model Database (PMDB) (Castrignano et al., 2005) and assigned id: PM0081519 for further structural annotations. Generated model was rendered by UCSF CHIMERA 1.10 for the visual inspection in different 3D shapes (Figure 1a, b). Dihedral statistics of Ramachandran plot for 3D structure of TaENOD93-1 protein revealed a total of 86.1, 8.3, and 5.6% residues in most favoured, additionally allowed, and generously allowed, respectively (Figure 1c). The Ramachandran plot predicted by Richardson's lab also showed the excellent geometry of the protein model. The overall G-score -0.23 suggested that the modeled
structure was acceptable (Table 1). Verify3D server gave the score 0.06 which is indicated that generated model belonged to good quality.

**Regulatory networks**

Since the data for wheat interactome (the whole set of molecular interactions) are not available in STRING database, we believe on the available interactome data of rice for conducting the network analysis of ENOD93-1 protein and its regulatory partners. Predicted regulatory interaction network suggested that ENOD93-1 has four regulatory partners (Figure 4). These partners have been predicted based on the shared functional domains, physical interactions, and co-expression analysis. Predicted regulatory partners of ENOD93-1 protein play a central role in several physiological processes. Further, the interaction between predicted regulatory partners of ENOD93-1 could be validated in wet-lab by using co-immunocoprecipitation and in-situ hybridization methods.

Nitrogen is one of the most important macromolecules for plant growth and development. The identification and functional annotation of favorable gene variants for NUE is a fundamental strategy to improve crop plants. Annotations of gene structure and basic function in N regulation is central to functional genomics, both with respect to molecular breeding and engineering crop plants for important traits. In this study, we have identified and mapped a TaENOD93-1 gene on the 4AL chromosome of wheat. We conducted the sequence and structure based functional annotation of identified TaENOD93-1 gene followed by molecular modeling and regulatory networks. This study, with the in-silico based annotation being placed on a consensus chromosomal map, and anticipates that will contribute to a more thorough study of its physiological significances on NUE improvement in wheat crop. In the future identified TaENOD93-1 could be utilized as a framework for a similar functional annotation of other NUE associated genes in the major complex cereals including barley, rice, maize etc.

**Conflict of interest**

The authors declare that the research was conducted in the absence of any commercial or financial relationship that could be constructed as a potential conflict of interest

**References**


Montzka, S., Dlugokencky, E. and Butler, J. 2011. Non-CO2 greenhouse gases and

How to cite this article:

doi: https://doi.org/10.20546/ijcmas.2018.705.379