The present study was conducted at Institute of Biotechnology (IBT), Rajendranagar, Hyderabad, Telangana, India in kharif, 2022 (July–October) to characterize yield enhancing genes in rice by in Silico Approach. Increasing the rice grain yield is a major constraint to feed the present burgeoning population which is anticipated to reach nine billion by 2050. In this regard many yield enhancing genes have been characterized for improving the yield potential in rice. Present study was undertaken on eight yield enhancing genes of rice for their structural and functional characterization by in silico approach. The conserved motif sequences shared across eight yield enhancing genes were identified and their positions on the genome were reported. The Spatio-temporal expression pattern of the yield genes in different stages of rice was studied, of which Gn1a and SPL14 genes shown higher level of expression at embryo and inflorescence stages respectively. Further, the expression patterns of these genes to different hormonal exposures were recorded in which Gn1a gene showed higher level of expression to cytokinin hormone in root and shoot tissues. The structure analysis for the above eight genes recorded the exon number between 2–10 along with their positions on the gene. Along with the protein characters of the above eight proteins, their potential interacting partners were identified, which shown their interactions to proteins controlling the flowering time, floral organ formation and regulating the grain size in rice.

KEYWORDS: Gene structure, motif, rice, yield enhancing genes
1. INTRODUCTION

Majority of the world’s population relies on rice for primary source of food. Despite of its increased production following green revolution and hybrid rice technology, rice yield remains a hot topic in rice breeding due to change in climatic conditions globally, steadily rising population and declining arable land (Ashikari et al., 2005, Song et al., 2007). According to research, from 1961 to 2008, the top three rice producers, India, China and Indonesia, had seen yield stagnation in more than 37%, 78%, and 81% of their respective rice-growing regions (Ray et al., 2012). Therefore, it is difficult, but in the future molecular breeding approaches have an inordinate potential for increasing rice yield.

Among the major cereal crops rice has the smallest genome with predicted size of 430 Mb which is entirely characterised (Arumuganathan and Earle, 1991). The potential source for the functional genomic research in rice is provided by the reported genome sequence, which covers about 95% of the 389 Mb genome. Furthermore, the genetic and molecular bases of quantitative traits can be investigated by advancements in genomics and transcriptomics. Rice yield is a complex quantitative trait determined majorly by number of panicles plant\(^{-1}\), number of grains panicle\(^{-1}\) and grain weight (Sakamoto and Matsuoka, 2008, Xing and Zhang, 2010). In the past few decades, significant progress has been made in the study of these crucial agronomic traits (Huang et al., 2013, Zuo and Li, 2014, Li and Li, 2016, Yu et al., 2013, Zhou et al., 2013) and abundant number of genes and QTLs related to yield have been identified and characterized (Chen et al., 2013) in which rice grain yield is directly affected by 20 QTLs (Bai et al., 2012). Major QTLs for yield traits i.e., Grain number, plant height, and heading date\(^7\) (Ghd7) (Xue et al., 2008, Yan et al., 2013, Weng et al., 2014), Dense and Erect Panicle (DEPI) (Li et al., 2016, Xu et al., 2016, Huang et al., 2022), Grain number 1a (Gn1a) (Feng et al., 2017, Reyes et al., 2021), OsSPL14/WFP/IPA1 (Jiao et al., 2010, Miura et al., 2010, Kim et al., 2018), Grain size (GS5) (Li et al., 2011, Kim et al., 2016), Thousand grain weight (TGW6) (Ishimaru et al., 2013), STRONG CULM2 (SCM2) (Ookawa et al., 2010, Ñano et al., 2015, Pandit et al., 2021), and SPIKELET NUMBER (SPIKE) (Fujita et al., 2013) have been identified.

The Gn1a gene (Ashikari et al., 2005) encodes cytokinin oxidase/dehydrogenase (OsCKX2), an enzyme that degrades cytokinin. Reduced OsCKX2 expression causes cytokinin accumulation in inflorescence meristems, which increases the number of reproductive organs, grains, and grain yield. Rice panicle architecture is regulated by phosphatidylethanolamine-binding protein-like domain which is encoded by DEP1 which produces shorter internodes in inflorescence and higher grain yield because of enhanced meristematic activity. OsSPL14/WFP encodes squamosa promoter-binding protein-like 14 in young panicles and is an ideal plant architecture (IPA1) gene that promotes panicle branching, resulting in more grains panicle\(^{-1}\) and increased yields (Miura et al., 2010).

Furthermore, Narrow Leaf1 (NAL1), which encodes a plant-specific protein with an unidentified biochemical function that boosts grain yield in indica cultivars, is allelic to SPIKE, LSCHL4 and GPS (Fujita et al., 2013, Takai et al., 2013, Zhang et al., 2014) that regulates leaf width and plant height. The yield-related variables heading date, plant height, and grain number panicle\(^{-1}\) are regulated by the CCT domain protein Ghd7. Through putative serine carboxypeptidase, GS5 controls grain width, filling and weight in addition to acting as a positive regulator of a portion of the G1-to-S transition genes in the cell cycle. APO1 (Aberrant Panicle Organization 1) is an F-box protein that controls culm diameter, tiller outgrowth, and panicle rachis branching. Furthermore, the mild allele SCM2 (Strong Culm 2) of APO1 helps to increase culm diameter and grain number panicle\(^{-1}\) while reducing tiller number (Ookawa et al., 2010, Terao et al., 2010). TGW6, encodes a new protein with hydrolase activity, controls both source ability and sink size to regulate grain weight and is a single-exon gene (Ishimaru et al., 2013). This study aims to investigate the structural and physicochemical characteristics of rice yield proteins. Additionally, investigations on protein-protein interactions and prediction of conserved patterns uncover functional protein domains and their interacting partners in the regulatory network.

2. MATERIALS AND METHODS

The current in silico analysis was conducted at Institute of Biotechnology, Rajendranagar, Hyderabad, Telangana, India for structural and functional characterization of yield enhancing genes in rice during kharif, 2022 (July–October).

2.1. Sequence retrieval

The sequences of eight rice yield genes and their proteins were retrieved from the NCBI and Uniprot databases respectively.

2.2. Conserved motif identification

The MEME server is used for the discovery of conserved motifs in all the above-downloaded sequences with a minimum of 5 motifs along with their locations in selected crop species (Bailey et al., 2009).

2.3. Spatio-temporal expression of genes

Global expression outline of genes viewed in two major dataset categories, namely, field/ development and plant hormones in the Ricepro database. These were the 
microarray data derived from the tissues or organs of various development stages under natural field conditions and from the various plant hormones treated shoot and roots.

2.4. Gene structure display

The positions of exon-intron or gene structure of genes were identified in GSDS 2.0 server (http://gsds.cbi.pku.edu.cn) (Hu et al., 2015), by aligning the gene sequences to their corresponding coding sequences.

2.5. Protein features identification

Using ProtParam tool, the physico-chemical features of yield protein were analyzed (https://web.expasy.org/protparam) and the prediction of subcellular localization of these yield enhancing proteins was carried out by cello server (http://cello.life.nctu.edu.tw) (Yu et al., 2006).

2.6. Interactions of yield proteins

The String server was utilized for the prediction of interactome network between the proteins (http://string-db.org) and was visualized using Cytoscape tool (Franceschini, et al., 2012).

3. RESULTS AND DISCUSSION

3.1. Sequence analysis of yield genes and proteins

Gene sequences of eight rice yield genes, Os01g0197700 (Gn1a), Os09g0441900 (DEP1), Os08g0509600 (SPL14), Os04g0615000 (SPIKE), Os05g0158500 (GS5), Os07g0261200 (Ghd7), Os06g0665400 (SCM2) and Os06g0623700 (TGW6) retrieved from the NCBI database were used for identifying the conserved motifs across them.

Protein sequences of rice yield proteins, Gn1a, DEP1, SPL14, SPIKE, GS5, Ghd7, SCM2 and TGW6 retrieved from the database UniProtKB/SwissProt of NCBI (Table 1) were utilized for estimating the subcellular localization of these proteins in the cell.

3.2. Conserved motif identification

The short continuous patterns in genomic DNA, which are assumed to have or involve in a biological function are motifs (Bailey et al., 2006). The five most conserved motif sequences were identified in yield genes (Figure 1) using the MEME tool. Motif 1 (CGGGCGRCGGGCCGCRCGGGCGGCRCGGGCGG), Motif 2 (GGGSCGGCCGCRCGGGCGGCRCGGGCGG), Motif 3 (GGGSCGGCCGCRCGGGCGGCRCGGGCGG), Motif 4 (CCGGCCGGCCGCRCGGGCGGCRCGGGCGG), and Motif 5 (GKHCGHBCGCYHVSHYGBGCCGGCCGCBRGSBS) were 25–47 nucleotides long (Table 2). The locations of the above five identified motifs in all the eight yield genes were identified (Figure 2). The identified motif sequences help in predicting the function of unknown genes.

3.3. Spatio-temporal gene expression analysis

The RAP locus ID of eight rice yield related genes were used as a query in the Ricexpro database for tissue-specific gene expression analysis. The expression pattern overview of these genes at various developmental stages or developmental cycle were analyzed which helps in knowing the target function and localization of a particular protein in plant tissues. The expression profile of the eight genes in 48 samples of various tissues were represented in Figure 3. Of the 8 genes, 5 reported tissue-specific expression at diverse developmental stages in 48 different tissues/ organs (Figure 3) and recorded high level of expression in developmental tissues namely root (reproductive-12:00; 00:00), embryo (14, 28 and 42 DAF), endosperms (28 and 42 DAF). SPL14 exhibited higher expression in stem (reproductive-12:00) inflorescence (0.6–1.0, 3.0–4.0, 5.0–10 mm). DEP1, SPL14 and GS5 recorded negligible expression in different developmental tissues in different stages. High downregulation of GS5 was seen in embryo (28 and 42 DAF) and endosperms (28 and 42 DAF). SPL14 highly downregulated in leaf blade (vegetative-12:00, reproductive-12:00, ripening- 12:00, 00:00), leaf sheath (reproductive- 00:00).

3.4. Phyto-hormonal expression profiling

The plant hormones act as a signaling molecules in inducing the defense response in plants. Understanding the inter-connectivity between plant hormones and the expression of above yield genes aids in further studies for improving the yield of plant. Here, the expression pattern
### Table 1: Protein characteristics of yield enhancing genes

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Gene ID</th>
<th>Locus ID</th>
<th>Protein ID</th>
<th>NAA</th>
<th>MW</th>
<th>TP</th>
<th>II</th>
<th>AI</th>
<th>GAHP</th>
<th>SCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gn1a</td>
<td>Os01g0197700</td>
<td>LOC_ Os01g10110</td>
<td>Q4ADV8</td>
<td>565</td>
<td>6021.22</td>
<td>6.15</td>
<td>37.76</td>
<td>87.98</td>
<td>0.026</td>
<td>Extracellular</td>
</tr>
<tr>
<td>DEP1</td>
<td>Os09g0441900</td>
<td>LOC_ Os09g36160</td>
<td>Q67UU9</td>
<td>315</td>
<td>3201.54</td>
<td>8.97</td>
<td>59.72</td>
<td>51.9</td>
<td>-0.361</td>
<td>Plasma membrane and Nucleus</td>
</tr>
<tr>
<td>SPL14</td>
<td>Os08g0509600</td>
<td>LOC_ Os08g39890</td>
<td>Q7EXZ2</td>
<td>417</td>
<td>4237.83</td>
<td>9.46</td>
<td>55.66</td>
<td>50.48</td>
<td>-0.495</td>
<td>Nucleus</td>
</tr>
<tr>
<td>SPIKE</td>
<td>Os04g0615000</td>
<td>LOC_ Os04g52479</td>
<td>B4XT64</td>
<td>582</td>
<td>6325.17</td>
<td>5.06</td>
<td>41.2</td>
<td>86.43</td>
<td>-0.242</td>
<td>Nucleus and Cytoplasm</td>
</tr>
<tr>
<td>GS5</td>
<td>Os05g0158500</td>
<td>LOC_ Os05g06660</td>
<td>Q5W727</td>
<td>483</td>
<td>5371.85</td>
<td>6.11</td>
<td>41.27</td>
<td>77.83</td>
<td>-0.241</td>
<td>Extracellular</td>
</tr>
<tr>
<td>Ghd7</td>
<td>Os07g0261200</td>
<td>LOC_ Os07g15770</td>
<td>E5RQA1</td>
<td>287</td>
<td>3065.44</td>
<td>6.1</td>
<td>59.37</td>
<td>51.11</td>
<td>-0.516</td>
<td>Nucleus</td>
</tr>
<tr>
<td>SCM2</td>
<td>Os06g0665400</td>
<td>LOC_ Os06g45460</td>
<td>Q655Y0</td>
<td>407</td>
<td>4273.55</td>
<td>9.95</td>
<td>63.51</td>
<td>94.32</td>
<td>0.3</td>
<td>Inner membrane</td>
</tr>
<tr>
<td>TGW6</td>
<td>Os06g0623700</td>
<td>LOC_ Os06g41850.1</td>
<td>Q69U01</td>
<td>350</td>
<td>3816.83</td>
<td>9.67</td>
<td>34.94</td>
<td>70.77</td>
<td>-0.367</td>
<td>Vacuole</td>
</tr>
</tbody>
</table>

NAA: Number of amino acids; MW: Molecular weight; TP: Theoretical pI; II: Aliphatic index; GAHP: Grand average of hydropathicity (GRAVY); SCL: Sub-cellular localization

### Table 2: Details of identified motif sequences

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Motifs</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Motif 1</td>
<td>CGGCCGRCGGMGCCGCGGC- GCCGCCGCCG</td>
</tr>
<tr>
<td>2.</td>
<td>Motif 2</td>
<td>GGSBCGCCCGCCGCCGCYC- GSSBCG</td>
</tr>
<tr>
<td>3.</td>
<td>Motif 3</td>
<td>GCCGKCGCCBGCGBKGCGBGCGB- SCCDS</td>
</tr>
<tr>
<td>4.</td>
<td>Motif 4</td>
<td>CCHCTTCTTCYTYCTCYTCYT- STCCHTC</td>
</tr>
<tr>
<td>5.</td>
<td>Motif 5</td>
<td>GCVGCCGSMGKHGCHBGCGVH- SYSBGCNCGGCGGRBGSBS</td>
</tr>
</tbody>
</table>

of these protein coding genes to different plant hormonal treatments in root and shoot tissues were carried out using the RiceXpro database. The expression pattern of rice yield genes in shoot and roots of seedlings exposed to different hormones namely, indole-3-acetic acid (IAA), abscisic acid (ABA), brassinolide (BL), gibberellic acid (GA3), jasmonic acid (JA), and trans-zeatin (tZ) were recorded. The expression pattern of yield genes in roots to different hormones was analyzed at 15 m, 30 m, 1 h, 3 h, 6 h of exposure and 1 h, 3 h, 6 h, 12 h of exposure on root and shoot of seedlings respectively. In the shoot, Gn1a gene recorded high level of expression in cytokinin. Gn1a, DEP1 exhibited low level of expression to abscisic acid, gibberellin and jasmonic acid treatments respectively. Gn1a gene exhibited negligible expression to abscisic acid, jasmonic acid treatments at 6 h and 12 h respectively. GS5 displayed low expression to abscisic acid, auxin and jasmonic acid treatments at different times (Figure 4).

In the root, Gn1a recorded high level of expression to abscisic acid, auxin, cytokinin treatments. Gn1a exhibited low level of expression to Brassinosteroid and Jasmonic acid at different time points. GS5 displayed negligible expression to abscisic acid, auxin and cytokinin at different times (Figure 5).

#### 3.5. Gene structure display

The structure displays the exon-intron regions and their positions along with the upstream and downstream elements. The blue colour region indicates the upstream and downstream elements of a particular gene. The yellow region represents the CDS sequences and the black line indicates the intronic regions of a gene (Figure 6). The structure analysis for the above eight genes recorded the exon number between 2–10 along with their positions on the gene.
The knowledge on organization of the exon-intron in genomic DNA is crucial in understanding the structural organization of genes, evolutionary changes, and their protein functionality among different species (Jo and Choi, 2015). The information on structure of genes can be utilized for phylogenetic analysis in understanding the gene structural changes over evolutionary period which helps in understanding the molecular evolution mechanism of genes and genomes. It also helps in the identification of splicing sites in pre-mRNA during mRNA formation (Lee and Rio, 2015).

3.6. Structural and functional analysis of proteins

The eight-rice yield enhancing proteins reported with 287-582 residues of amino acids having molecular weight of 30.6–63.2 kDa. The instability and isoelectric point values of proteins ranged from 5.06–9.95 and 34.94–63.51 (stable) respectively (Table 1). Using the CELSER server, it is known that all the SPL14 and GhD7 proteins were localized to the nucleus, GS5 to the extracellular membrane, TGW6 protein to the vacuole and SCM2, DEP1 proteins to the inner membrane and plasma membrane of the cell respectively.

3.6.1. Protein interacting partners

The studies on protein-protein interactions will help in understanding the functional pathways and molecular mechanisms carried out by the proteins (Rao et al., 2014). The eight proteins were used for studying protein-protein interactions.
partners of the selected proteins were identified. Apart from the interactions between the above eight yield related proteins, these proteins were predicted to have interaction partner proteins viz., EF7, HD3A, EHD1, MADS14 and MADS16 which are involved in controlling the flowering time and floral organ formation and also to other proteins viz., GW2, GS3 and DN1 proteins involved in regulating the grain size (Figure 7).

4. CONCLUSION

The structural, expression patterns of the rice yield enhancing genes and functional properties, protein interacting partners of proteins were identified using bioinformatics tools. The information reported in this study will be useful as preliminary data for much other functional analysis of these genes. The studies on the interactions with other proteins will be useful in increasing the crop yield by the genetic engineering approach.

5. REFERENCES

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., Noble, W.S.,


Reyes, V.P., Angeles-Shim, R.B., Lapis, R.S., Shim, J., Sunohara, H., Jena, K.K., Ashikari, M., Doi, K., 2021. Improvement of Asian rice cultivars through marker-assisted introgression of yield QTLs, Grain Number 1A (GN1A) and Wealthy Farmer’s Panicle...
Terao, T., Nagata, K., Morino, K., Hirose, T., 2010. A gene controlling the number of primary rachis branches also controls the vascular bundle formation and hence is responsible to increase the harvest index and grain yield in rice. Theoretical and Applied Genetics 120(5), 875–893.