# *In-silico* analysis of CDS for snorkel genes of *Oryza sativa* using bioinformatics tools and techniques in deep water rice

Pragnya Paramita Jena*[1], Pranati Swain[2], Gayatri Gouda[3]

[1,2,3] *Senior Research Fellow, Crop Improvement Division, Central Rice Research Institute (CRRI), Cuttack, Odisha 753006, India.*
jena.prajna@gmail.com[1], rosylora20@gmail.com[2], gayatrigouda90@gmail.com[3]

## Abstract

**Background**: Prolonged submergence stress is a serious consequence in rice. The CDS regions of Snorkel protein were considered for the study.
**Method:** A brief comparative analysis has been performed by using several bioinformatics tools and techniques in order to understand the physico-chemical properties, secondary structures of conserved protein regions, multiple sequence alignment, homology models, and backbone confirmation for generated homology models.
**Findings**: The objective of this study was to understand the similarity among all these four CDS regions. Different bioinformatics tools PROTPARAM, CFFFSP server, CLC genomic workbench, MODELLER, RAMPAGE has been used to understand both structural and comparative differences. This insilco analysis will helpful to researchers to study more on the networking mechanism between sub1 introgressed lines and snorkel introgressed lines and mechanism involve when both the gene are introgressed together in single variety.
**Application/Improvements**: understanding the network mechanism in snorkel introgressed lines.

***Keywords***: snorkel, submergence tolerance, protein, bioinformatics tool

## 1. Introduction

Rice landraces discovered till date can tolerate up to two weeks of submergence tolerance. Rainfed lowland and deepwater rice together account for approximately 33% of global rice farmlands (50 million hectares of the estimated 150 million hectares of rice fields worldwide in 2004–2006 [1]. Flooding by acquiring the ability to significantly elongate its internodes, which have hollow structures and function as snorkels to allow gas exchange with the atmosphere, and thus prevent drowning. Deepwater rice responds to partial submergence by enhancing cell division and elongation in the internodal regions of underwater stems, via a mechanism triggered by entrapment of ethylene, which promotes abscisic acid (ABA) degradation and increases gibberellic acids (GA) and their downstream effects [2]. Molecular mechanism for the deepwater rice response regulated by SNORKEL1 (SK1) and SNORKEL2 (SK2).Among the two progenitors (WO120 and WO 106) of *Oryza sativa* the progenitor WO120 contain both the genes. Gain-of-function analysis of the SK genes suggested that SK2 has a more pronounced effect than SK1.AmongThe SK genes were expressed in leaf blade, leaf sheath, and basal parts of the stem, including nodes and internodes, in which the deepwater response occurs [3]. This unusually robust underwater growth is controlled by three quantitative trait loci (QTLs). Of these, the SNORKEL QTL on chromosome 12 encodes two ethylene responsive factor (ERF) DNA binding proteins, SNORKEL1 (SK1) and SNORKEL2 (SK2), that are absent from the non-deepwater rice accessions evaluated to date [4,5]. In contrast to Sub1 rice, deepwater rice escapes stagnant partial flooding by promoting elongation of internodes. The deepwater rice genes SK1/SK2 and the submergence tolerance gene SUB1A regulate ethylene-mediated GA responsiveness in an opposing manner; it seems unlikely that they can be combined to generate genotypes resilient of both stagnant flooding and submergence. The CDS of all the four genes has been selected for in-silico comparison. A brief proteomic analysis was performed by using bioinformatics tools and techniques to find out the relationship among all the CDS of genes.

## 2. Material and methods

### 2.1. Sequence retrieval

Two deep water varieties of rice Bhadua and C9285 were considered for study. Both the varieties possessing SK1 and SK2 gene. The CDS of genes with the GenBank IDs AB510481.1 (snorkel2-BHADUA), AB510480.1 (snorkel1-BHADUA), AB510479.1 (snorkel2-C9285), AB510478.1 (snorkel1-C9285) were retrieved from NCBI. The fasta formatted sequences were further used for study.

### 2.2. Protein structure prediction

#### *2.2.1. Secondary structure prediction*

The secondary structures were generally predicted to find out the percentage of helices, sheets and turns [6].The secondary structures of the corresponding amino acids were predicted by using the CLC genomic workbench, which showed information about beta, helix and turns (Figure.1a,b, c,d) (Table 1).

*Figure 1a. secondary structure for CDS of AB510481.1*

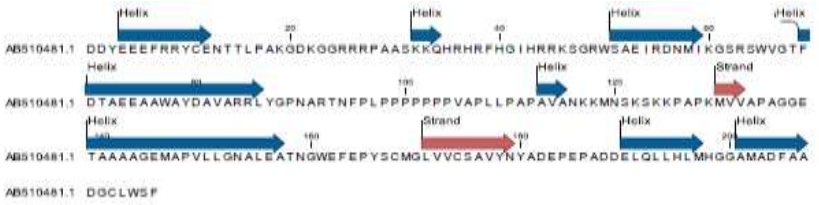



*Figure 1b. Secondary structure for CDS of AB510480.1*


*Figure1c. Secondary structure for CDS of AB510479.*

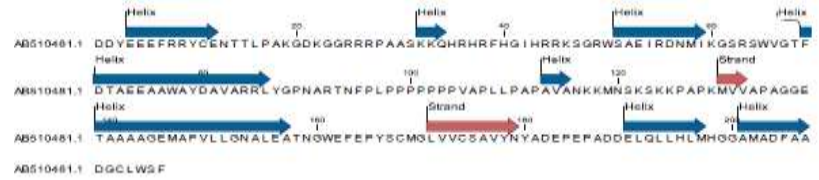



*Figure 1d. Secondary structure for CDS of AB510478.1*


*Table1. secondary structure of proteins*

| IDs | helix | sheet | turn | coil |
|---|---|---|---|---|
| AB510478.1 | 29.0% | 33.9% | 16.1% | 21.0% |
| AB510479.1 | 29.0% | 33.9% | 16.1% | 21.0% |
| AB510480.1 | 29.0% | 33.9% | 16.1% | 21.0% |
| AB510481.1 | 29.0% | 33.9% | 16.1% | 21.0% |

### 2.3. Physico-chemical analysis

The physico-chemical properties generally represents the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity19 (GRAVY) was computed by using ProtParam tool (http://web.expasy.org/protparam/). Both the sequences were submitted in protparam from where both physical and chemical properties were found successfully (Table 2).

*Table 2. physico-chemical properties of proteins*

| properties | AB510478.1 | AB510479.1 | AB510480.1 | AB510481.1 |
|---|---|---|---|---|
| Number of amino acids | 256 | 258 | 256 | 258 |
| Molecular weight | 27749.0 | 27948.6 | 27749.0 | 27948.6 |
| Total number of negatively charged residues | 35 | 33 | 35 | 33 |
| Total number of positively charged residues | 31 | 32 | 31 | 32 |
| Total number of atoms | 3803 | 3864 | 3803 | 3864 |
| Aliphatic index | 57.34 | 62.95 | 57.34 | 62.95 |
| Grand average of hydropathicity | -0.676 | -0.506 | -0.676 | -0.506 |
| Theoretical PI | 6.07 | 6.60 | 6.07 | 6.60 |

### 2.4. Multiple sequence alignment

A multiple sequence alignment is done among the CDS of four different varieties to find out the conserved residues among them. It is helpful to understand the phylogenetic relationship among different varieties. CLC workbench has been used for generating phlogram and cladogram trees as well as to find the common residues (Figure.2a, b, c).

*Figure 2a. Multiple sequence alignment among four different varieties showing the conserved residues*

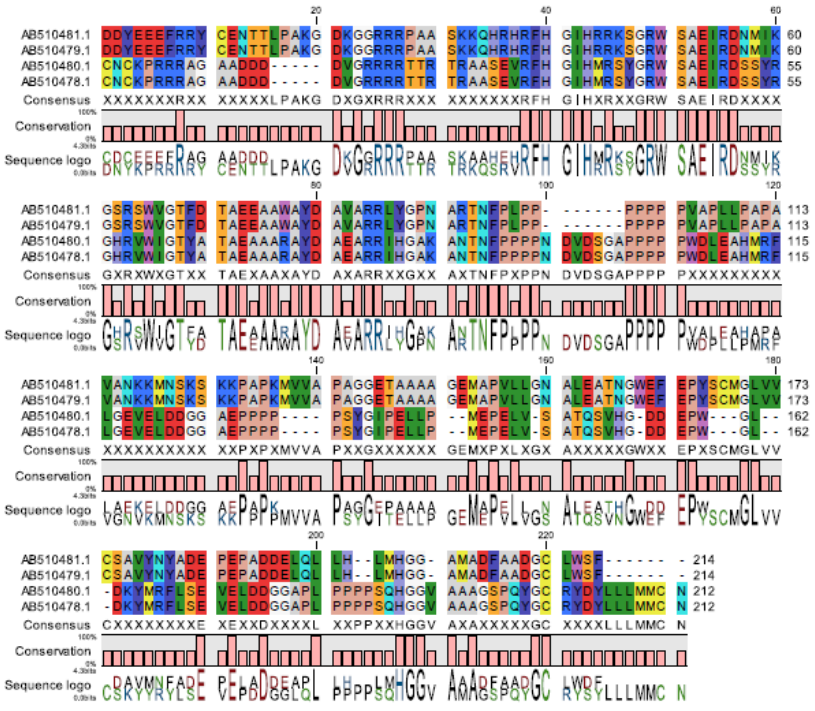


*Figure 2b. Circular phylogram tree*

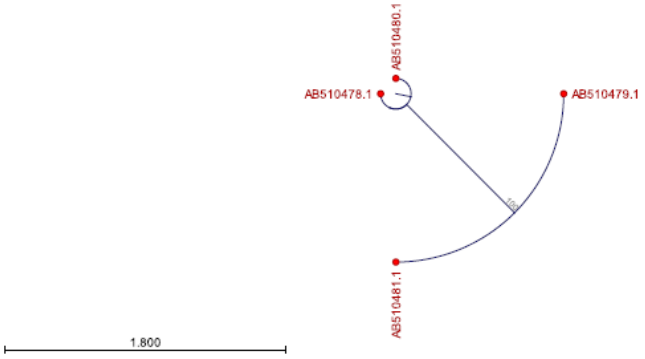
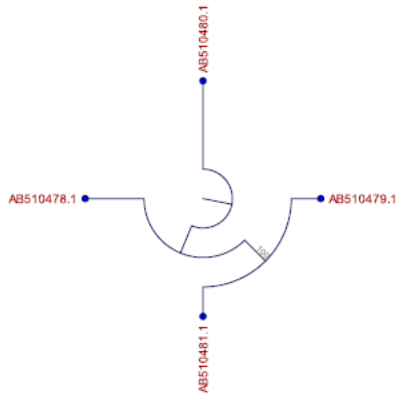



*Figure 2c. Circular cladogram tree*




**2.5. Homology modeling**

As the NMR crystallographic structures of the models are not available at PDB, so the models were generated by using model building servers. The tertiary structures of proteins were predicted by using the homology modeling concept in which the template is generated and the target is compared with the template to generate more appropriate model [7](Table 3). Modeller tool was used to generate the 3D structures of proteins. Models with highest DOPE scores were selected for further analysis. The final models were visualized by using PyMol visualization tool. The tertiary structures of proteins are given in (Figure 3) where helices were colored with red colors, sheets with yellow colors and loops with green colors respectively.

*Figure 3. 3D Structures of (A)AB510478.1 snorkel2 C9285, (B)AB510479.1 snorkel2-C9285, (C) AB510480.1 snorkel11-bhadua, (D) AB510481.1 snorkel11-bhadua*

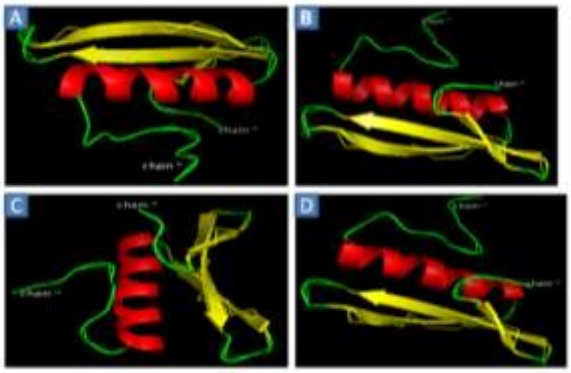


*Table 3. Result which found after the visualization of models*

| IDs | Atom count | Formal Charge sum | Molecular Surface area | Solvent successible Surface area | Stability of The object | VDW of The object |
|---|---|---|---|---|---|---|
| AB510478.1 snorkel2-C9285 | 522 | 5.0 | 6460.129A$^0$ | 5321.692A$^0$ | 64.39 kcal/mol | 25.778A$^0$ |
| AB510479.1 snorkel2-C9285 | 528 | 5.0 | 6557.098 A$^0$ | 5302.572 A$^0$ | 58.91 kcal/mol | 25.801A$^0$ |
| AB510480.1 snorkel11-bhadua | 522 | 5.0 | 6460.129A$^0$ | 5321.692A$^0$ | 64.39 kcal/mol | 25.778A$^0$ |
| AB510481.1 snorkel11-bhadua | 528 | 5.0 | 6557.098 A$^0$ | 5302.572 A$^0$ | 58.91 kcal/mol | 25.801A$^0$ |

## 2.6. Model validation and optimization

The models were then optimized by using RAMPAGE server (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php). The backbone confirmation of the proteins was checked. The allowed, favored regions along with the outliner regions were found. When most of the residues lie in the favored region and less in outlier region, then it indicates that the protein is suitable for further study (Figure.4a, b, c, d).

*Figure 4a. AB510479.1- snorkel2-C9285*

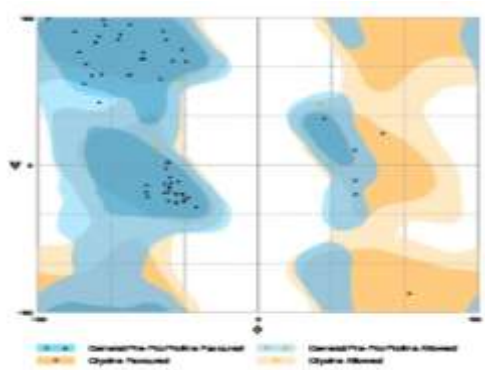


*Figure 4b. AB510478.1- snorkel2-C9285*

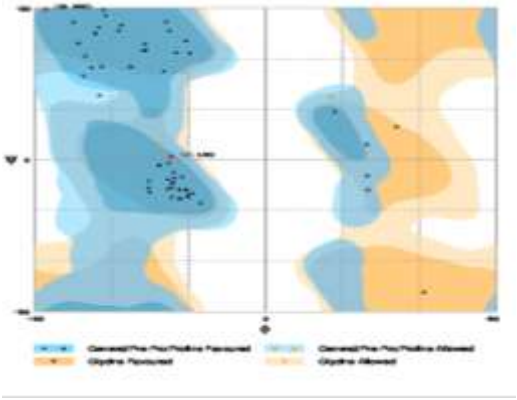


*Figure 4c. AB510480.1- snorkel11-bhadua*

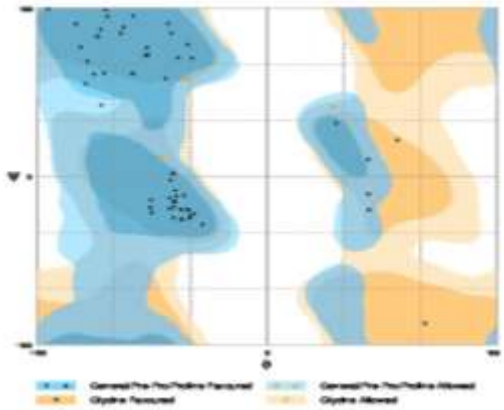
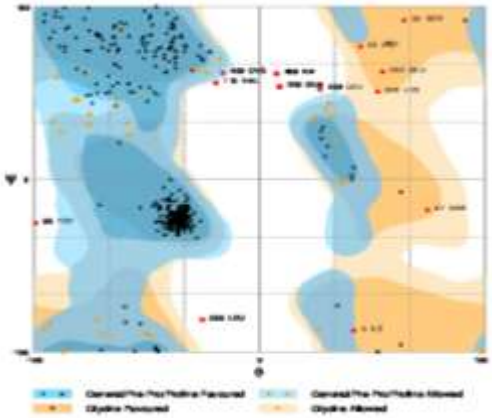


*Figure 4d. AB510481.1- snorkel11-bhadua*



## 3. Result

### 3.1. Sequence retrieval and secondary structure prediction

The amino acid sequences were retrieved from uniprot and subjected for secondary structure prediction. The secondary structure of proteins shows that all the proteins are similar on the basis of their secondary structure.

### 3.2. Physico-chemical analysis

From the physico-chemical analysis it is found that the protein of snorkel2-C9285 (AB510478.1) is similar to the protein of snorkel11-bhadua (AB510480.1), whereas the protein of snorkel2-C9285 (AB510479.1) is similar with the protein of snorkel11-bhadua (AB510481.1)

### 3.3. Multiple sequence alignment

The multiple sequence alignment showed the conserved residues are D, Y, E, R, C, N, T, K, S, H, E, I, V, L, W, P, Q, M.

### 3.4. Tertiary structure prediction analysis

The amino acid sequences of the protein were subjected for generating homology models and then visualized by using various visualization tools. The helices are denoted with red, sheets with yellow and loops in green colors respectively.

The visualization of protein structures were done by using pymol visualizer, from where the total number of atoms, formal charge sum, Molecular Surface area, Solvent successible Surface area, Stability of The object, VDW of The object. On the basis of these characters the models were selected for further analysis (Table 3).

Amino acids are the building blocks of the proteins and the backbone confirmation generally refers to phi-psi bonds, i.e. the regions where the amino acids are bind together by the peptide bonds due to electrostatic force, vander-walls force. So when more amino acid residues lie in the favored regions, then it gives better result on the further analysis (Table 4).

*Table 4. Backbone confirmation of proteins*

| IDs | Number of residues in favored region | Number of residues in allowed region | Number of residues in outlier region |
|---|---|---|---|
| AB510478.1 snorkel2-C9285 | 57 (95.0%) | 3 (5.0%) | 0 (0.0%) |
| AB510479.1 snorkel2-C9285 | 55 (91.7%) | 3 (5.0%) | 2 (3.3%) |
| AB510480.1 snorkel11-bhadua | 414 (90.0%) | 33 (7.2%) | 13 (2.8%) |
| AB510481.1 snorkel11-bhadua | 1179 (84.5%) | 123 (8.8%) | 94 (6.7%) |

## 4. Discussion

Snorkel genes play an important role in deep water rice tolerance evolved from the ancestor WO120 variety of *O.sativa*. The mechanism involves in sub1gene submergence tolerance and snorkel gene tolerance mechanism is antagonistic to each other. Sub1 gene mechanism involve storage of energy in the stressed condition and utilize it for plant regeneration, the antagonistic is escape from the level of water by utilizing current resource. Sub1 gene is useful when period of stress is for 1-2 weeks but when it comes to deep water or stagnant water rice snorkel genes are useful. The snorkel gene is an from the in-silico analysis it is found that the proteins are similar with each other on the basis of secondary structures, however the C9285 snorkel1 gene is similar with the Bhadua snorkel 2 gene and vice-versa. The model which generated by using the protein of C9285 snorkel1 gave the best structure and most of the residues lie in favored regions, so this protein is more suitable for further study. This model prediction will helpful in breeding programs and it will give breeders a better look to choose the gene of introgression. Introgression of more potential gene to the elite varieties lacking the snorkel genes will improve their stagnant water tolerance mechanism for low land deep water varieties. This in silico analysis will helpful for researchers to study more on the networking mechanism between sub1 introgressed lines and snorkel introgressed lines and mechanism involve when both the gene are introgressed together in single variety. The broad study of understanding the networking mechanism will give a new horizon of mechanism of submergence tolerance. From all the in-silico analysis it is found

that the CDS which were used for proteomic study showed that the snorkel1-C9285 is similar with snorkel1-BHADUA and snorkel2-C9285 is similar with snorkel2-BHADUA. Finally these sequences were compared on lalign server by using Waterman-Eggert algorithm, which showed that snorkel1-C9285 (AB510478.1) is 100% similar with snorkel1-BHADUA (AB510480.1) with 256 amino acids overlap whereas snorkel2-C9285 (AB510479.1) is 100% similar with snorkel2-BHADUA (AB510481.1) with 258 amino acids overlap.

## 5. References


1. D. Catling. Rice in deep water. London. MacMillan Press Ltd. 1992.
2. R.E. Huke, E.H. Huke. Rice area by type of culture. south, southeast, and east Asia. A revised and updated database. *International Rice Research Institute*.1997.
3. H. Kende, E .Van der Knaap, H.T. Cho. Deepwater rice: a model plant to study stem elongation. *Plant Physiology*. 1998; 118, 1105-1110.
4. Y. Hattori, K. Nagai, S. Furukawa, X.J. Song, R. Kawano, H. Sakakibara, Jianzhong Wu, Takashi Matsumoto, Atsushi Yoshimura, Hidemi Kitano, Makoto Matsuoka, Hitoshi Mori, Motoyuki Ashikari. The ethylene response factors SNORKEL1 and SNORKEL2 allow rice to adapt to deep water. *Nature*. 2009; 460, 1026-1030.
5. L.A.C.J. Voesenek, J. Bailey-Serres. Genetics of high-rise rice. *Nature*. 2009; 460, 959-960.
6. S. Pranati, B. Lambodar. Comparative structural analysis of cytokinin dehydrogenase enzymes of *O.sativa*, A. thaliana and Zea mays leading to predict the best enzyme. *Computational Molecular Biology*. 2014; 14(4), 1-9.
7. Samina Bilal, Shumaila Barkat Ali, Sahar Fazal, Asif Mir. Generation of a 3D model for human cereblon using comparative modeling. *Journal of Bioinformatics and Sequence Analysis*. 2013; 5(1), 10-15.


*The Publication fee is defrayed by Indian Society for Education and Environment (iSee). www.iseeadyar.org*

***Cite this article as:***

Pragnya Paramita Jena, Pranati Swain, Gayatri Gouda. *In-silico* analysis of CDS for snorkel genes of *Oryza sativa* using bioinformatics tools and techniques in deep water rice*. Indian Journal of Bioinformatics and Biotechnology.* Vol 3 (2), 70-76, February, 2014.