In Silico Identification of Abiotic Stress-Responsive Candidate Genes in Rice and Their Allelic Differences Between At 354 and Bg 352

[1]Saubhagya Rathnayake, [2,*]Nisha Kottearachchi, [3]S. P. C. Lankika

[1,3]Department of Applied Sciences, Faculty of Humanities and Sciences,
Sri Lanka Institute of Information Technology, Malabe, Sri Lanka
[2]Department of Biotechnology, Faculty of Agriculture & Plantation Management,
Wayamba University of Sri Lanka
Email of corresponding Author- * nisha@wyb.ac.lk

ARTICLE INFO

ABSTRACT

Rice (*Oryza sativa L*) feeds more than half of the world's population. One of the main elements that harm yield globally is abiotic stress. Therefore, it is important to develop abiotic stress tolerant rice varieties in order to increase rice productivity and to extend the cultivation. The lack of knowledge of the genetic mechanisms underlying abiotic stress tolerance is the primary issue with the traditional breeding technique. Hence, studying genes responsible for abiotic stress mechanisms is important to accelerate breeding by molecular marker - based detection techniques. Aiming at finding the candidate genes for abiotic stress tolerance ,two rice genome sequences of At 354 and Bg 352 varieties given by National Research Council, Sri Lanka -16 -16 project were analyzed. At 354 has some abiotic stress tolerance (salt) traits and Bg 352 has some susceptible traits. Next-generation sequencing-derived genome sequences were used to identify SNPs and Indels in the At 354 and Bg 352 varieties with reference to *Oryza sativa* japonica group cultivar Nipponbare. The STRING Database was used to extract the most correlated genes with abiotic stress. The allelic differences among Nipponbare, At 354 and Bg 352 sequences were detected from Multiple Sequence Alignment by using the Rice Annotation Project

database, UGENE software and MEGA 11 software. The mutations of the genes were validated if they were present in another germplasm in the NCBI database. Altogether 100 genes were used to examine, and 166 mutations were observed including 163 SNPs and 3 Indels while 09 genes were validated due to their presence in other rice accessions. The amino acid sequences of the validated sequences were determined by Expasy Translate tool. The Swiss model database and ProtParam tool were used to predict the protein structures and their parameters, which showed some structural differences among tested alleles. These mutant alleles further need to be assessed against abiotic stress and varietal turnover in order to use them in rice improving breeding programs to be used in abiotic stress-prone ecosystems.

1. INTRODUCTION

Rice (*Oryza sativa L*) is one of the most important staple crops and feeds more than half of the world's population (Molina et al., 2011). Although the world's population is predicted to reach 9.1 billion people by 2050, it is increasing at a faster rate than agricultural output. The global agricultural output needs to rise by 60–110% in order to feed the world's population, and by 2050, there will need to be 70% more supply of food to feed an additional 2.3 billion people (Tilman et al.,2011). One of the main elements that harm crop development and yield globally is abiotic stress (Romana et al.,2023). In order to supply the need for food, abiotic stress is one of the main areas of concern (Shanker & Venkateswarlu,2011). High salinity, drought, submergence, and cold are the main abiotic factors that cause risks to global food security. Therefore, it is important to develop abiotic stress tolerant varieties of rice in order to increase rice productivity and extend its area of cultivation. The lack of knowledge of the genetic mechanisms underlying abiotic stress tolerance is the primary issue with the traditional breeding technique. Therefore, it is important to study the

abiotic stress mechanism at molecular level to develop abiotic stress tolerant rice varieties. As a solution for those problems, NRC -16 -16 project has produced the whole genome sequences of At 354 and Bg 352 by next-generation sequencing (Buddini et al. 2020). At 354 is a salinity tolerant elite rice indica variety and Bg 352 is a salinity susceptible elite rice indica variety. These data were used to identify SNPs and Indels in the At 354 and Bg 352 varieties with reference to Oryza sativa japonica group cultivar Nipponbare, in this study. The findings of the study will be useful in development of abiotic stress tolerant rice varieties.

The broad objective of this research was to identify genes that controlling abiotic stress tolerance, to be used in rice improving breeding programs. The specific objectives of the study were to identify the abiotic responsive candidate genes in rice and to reveal polymorphic alleles between Bg 352 and At 354 in abiotic stress responsive genes by using in-silico methods.

2. MATERIALS AND METHODS

As the first step, a list of important genes that related to abiotic stress tolerance of rice were found using literature. In the second step, most correlated genes with abiotic stress from the network developed by STRING Database were extracted. Raw sequencing data files were retrieved from NGS sequences, and converted into chromosome wise Bam files. Then allelic differences among Nipponbare, At 354 and Bg 352 sequences were detected by multiple sequence alignment. A multiple sequence alignment was carried out using the genes screened with variants as follows. The relevant BAM file was opened in UGENE version 40.1, and the sequences of the respective genes of At 354 and Bg 352 were copied. Then, Nipponbare sequence obtained from RAP-DB and At 354 and Bg 352 were aligned together using MEGA 11 software and the variations that

are noted in VCF file were confirmed. According to the cDNA sequence of Nipponbare, At 354 and Bg 352 sequences were converted into cDNA sequences of the respective genes by removing the intron regions. The amino acid sequences were built for the genes of At 354 and Bg 352 that showed variants in exons, and the open reading frames (ORF) were obtained by Translate tool-Expasy. The validation process was done to confirm the presence of mutations in the in silico derived sequences of At 354 and Bg 352. First, the amino acid sequences were blast using NCBI Blast tool. When similar variations were observed between query sequences and other varieties, those variations were considered as confirmed. Then the protein structure prediction was done by using the Swiss model database for the confirmed ones in order to check the possible structural variations.

3. RESULTS AND DISCUSSION

Altogether 100 genes were used to examine, and we observed 163 variants including 160 SNPs and 3 Indels in 20 genes compared to Nipponbare reference. From them, 09 genes could be validated due to their presence in other rice accessions according to NCBI-Blast website. When similar variations were observed between query sequences and other varieties, those variations were considered as confirmed. Table 1 shows the validated polymorphic nucleotide variations of At 354 and Bg 352.

Table 1 : Validated genes of candidate genes

Gene ID	Gene Name	Amino acid position	Variety	NCBI accessions which show the same mutation
Os06g0130100	Stress induced protein kinase 1	20 – A<E 590 – I<V	Bg 352 At 354	CP012614.1 (*Oryza sativa* Indica Group cultivar RP Bio-226 chromosome 6 Sequence)
Os01g0307500	High-affinity k+ transporter 8	128 – D<N 184 – R<H	Bg 352 At 354	CP018157.1 (*Oryza sativa* Indica Group cultivar Shuhui498 chromosome 1, partial sequence)
Os03g0688300	Calcium-dependent protein kinase 9	570 – I<L	Bg 352	CP018159.1 (*Oryza sativa* Indica Group cultivar Shuhui498 chromosome 3, partial sequence)
Os03g0795900	Heat stress transcription factor A2e	117- G<A 205 – S<A	At 354 At 354	CP018159.1 (*Oryza sativa* Indica Group cultivar Shuhui498 chromosome 3, partial sequence)
Os09g0522200	Dehydration-responsive element-binding protein 1A	172 – D<E	Bg 352	CP012617.1 (*Oryza sativa* Indica Group cultivar RP Bio-226 chromosome 9 Sequence)
Os02g0766700	bZIP transcription factor 23	199 – F<L 260 – K<N	Bg 352	CP054677.1 (*Oryza sativa* Indica Group cultivar Minghui 63 chromosome 2)

Gene ID	Protein name	Mutation	Variety	Reference sequence
Os03g0322900	Late embryogenesis abundant protein 17	Frame shift	Bg 352	CP018159.1 (*Oryza sativa* Indica Group cultivar Shuhui498 chromosome 3, partial sequence)
Os02g0208500	Cadmium tolerant 4	4 - P < L	At 354	CP012610.1 (*Oryza sativa* Indica Group cultivar RP Bio-226 chromosome 2 Sequence)

The Swiss model database and ProtParam tool were used to predict the protein structures and their parameters, which showed some structural differences among tested alleles. Figure 01 and table 02 show the protein structure and parameters differences of Os05g0148600 (NHX2) gene caused by different alleles. OsNHX2, which is based on leaf mesophyll cells and bundle sheath cells, plays an important role in salt stress in rice plants and can be used to genetically engineer crop plants with enhanced salt tolerance (Xin – xia et al.,2017).
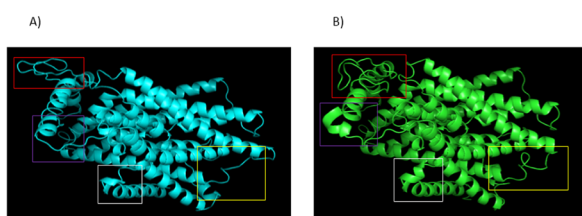

Figure 01:The protein structure of Os05g0148600 (NHX2) gene A) At 354 variety B) Bg 352 variety

Table 2: The parameters of Os05g0148600 gene

Parameters	Nippon-bare	At 354	Bg 352
Theoretical pI:	7.17	8.99	8.23
Aliphatic index	121.65	110.94	111.43
Grand average of hydropathici-ty (GRAVY)	0.680	0.551	0.564

4. CONCLUSION/S

This study aimed to find polymorphic nucleotide variants with potential of abiotic stress tolerance based on allelic differences between At 354 and Bg 352 varieties with reference to Nipponbare. As a result, we found 20 genes containing polymorphic variations in the exon region. 163 variants, including 160 SNPs and 3 Indels, were observed in those 20 genes. 12 mutations in 09 genes were validated due to their presence in other rice accessions, indicating their concordance. Amino acid sequences of At 354 and Bg 352 alleles showed missense and frame shift mutations leading to possible structural changes in their proteins. The protein structures were predicted for Os01g0307500 and Os05g0148600 genes and their parameters indicated some differences between At 354 and Bg 352 due to mutations. Further studies are necessary to examine the contribution of these mutations to abiotic stress tolerance.

REFERENCES

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.;*Protein Identification and Analysis Tools on the Expasy Server;* (In) John M. Walker (ed): The Proteomics Protocols Handbook, *Humana Press.*pp. 571-607

Molina, J., Sikora, M., Garud, N., Flowers, J. M., Rubinstein, S., Reynolds, A., Huang, P., Jackson, S., Schaal, B. A., Bustamante, C. D., Boyko, A. R., & Purugganan, M. D. (2011). Molecular evidence for a single evolutionary origin of domesticated rice. *Proceedings of the National Academy of Sciences of the United States of America, 108*(20), 8351–8356. https://doi.org/10.1073/pnas.1104686108

National Center for Biotechnology Information (NCBI). https://www.ncbi.nlm.nih.gov/

Shanker, A. (2011). Abiotic Stress in Plants - Mechanisms and Adaptations. *Crida.* http://www.academia.edu/1129311/Abiotic_Stress_in_Plants_-_Mechanisms_and_Adaptations

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N. T., Legeay, M., Fang, T., Bork, P., Jensen, L. J., & von Mering, C. (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic acids research*, *49*(D1), D605–D612. https://doi.org/10.1093/nar/gkaa1074

Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences of the United States of America, 108*(50), 20260–20264. https://doi.org/10.1073/pnas.1116437108

Unipro UGENE: a unified bioinformatics toolkit *Konstantin Okonechnikov, Olga Golosova, Mikhail Fursov,the UGENE team Bioinformatics* (2012)28: 1166-1167 doi: 10.1093/bioinformatics/bts091

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore &R., Schwede (2018)., T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res. 46*, W296-W303