# Harnessing High-throughput Sequencing in Developing Drought-Resistant Rice

Abdul Fatah A. SAMAD, AFAS, And Samad*
Department of Biosciences, Faculty of Science Universiti Teknologi Malaysia
abdulfatah@utm.my

Iqmal Asyraf ILLIAS, IAI, And Illias
Department of Biosciences, Faculty of Science Universiti Teknologi Malaysia
iqmalasyraf91@gmail.com

## ABSTRACT

High-throughput sequencing is a groundbreaking tool in molecular biology, facilitating a comprehensive understanding of the genetic composition of various rice varieties and their responses to environmental stressors, particularly drought. This event aims to explore the innovative applications and methodologies of high-throughput sequencing techniques in identifying genes associated with drought resistance in rice. Six transcriptomic libraries were prepared from both control and drought-treated rice, with each group consisting of three biological replicates. Approximately 37 million reads were generated from the control group, whereas around 43 million reads were generated from the treated rice. Specifically, 2325 genes showed significant up-regulation, while 1635 genes were downregulated. Among the significantly regulated genes are those encoding regulatory and stress proteins. This data offers a comprehensive understanding of gene regulation in rice under drought stress, thereby providing a strategy to develop drought-resistant rice varieties through breeding programs.

## CCS CONCEPTS

• **Information systems**; • **Data management systems**; • **Database management system engines**;

## KEYWORDS

High-throughput sequencing, transcriptome, drought, rice

**ACM Reference Format:**
Abdul Fatah A. SAMAD, AFAS, And Samad and Iqmal Asyraf ILLIAS, IAI, And Illias. 2023. Harnessing High-throughput Sequencing in Developing Drought-Resistant Rice. In *2023 7th International Conference on Computational Biology and Bioinformatics (ICCBB 2023), December 11–13, 2023, Kuala Lumpur, Malaysia.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3638569.3638573

## 1 INTRODUCTION

As the global population continues to grow, the imperative to ensure food security becomes increasingly paramount [1]. However, this task is incessantly challenged by the escalating impacts of climate change, notably the rise in frequency and severity of drought events. In the context of agriculture, where rice stands as a staple crop feeding billions, the need for drought-resistant varieties becomes ever more pressing [2]. Addressing this challenge necessitates cutting-edge tools and technologies, and one such innovation at the forefront is high-throughput sequencing. This revolutionary genomic technique allows for the rapid and comprehensive analysis of the entire genetic makeup of an organism, providing researchers with unprecedented insights into the complex interactions that govern drought resistance in rice. Harnessing high-throughput sequencing in the development of drought-resistant rice not only holds the promise of bolstering global food security but also signifies a paradigm shift in the intersection of genomics and crop improvement.

The utilization of high-throughput sequencing technologies marks a pivotal advancement in the pursuit of resilient agricultural systems, particularly in the context of developing drought-resistant rice. Traditional breeding methods, while effective, often entail protracted timelines and significant resource investments [3]. High-throughput sequencing, with its ability to decode the intricate genomic signatures associated with drought tolerance, offers an expeditious and precise alternative. By unraveling the genetic intricacies that govern a plant's response to water scarcity, researchers can identify key genes and molecular pathways crucial for conferring drought resistance. This wealth of genomic information not only expedites the breeding process but also enables targeted modifications, allowing for the development of rice varieties with enhanced resilience to drought conditions.

In this study, we utilize high-throughput sequencing to identify significantly regulated genes responsive to drought in rice. The finding may provide a thorough insight into the regulation of genes in rice when subjected to drought stress. Consequently, it presents a roadmap for the creation of drought-resistant rice varieties through targeted breeding initiatives.

## 2 METHODOLOGY

### 2.1 Plant material, germination and treatment

Oryza sativa L. ssp. Indica cv MR303 was used in this study. The seeds were acquired from MARDI Seberang Prai, Pulau Pinang branch. The rice seeds underwent surface sterilization and were subjected to a 3-day period of stratification in darkness at 4℃ before being planted in soil. The planted seeds were germinated under conditions with a daytime temperature of approximately 27℃ and a nighttime temperature of 25℃, while maintaining humidity levels between 70% and 90% throughout the day. The seeds were watered

*Place the footnote text for the author (if applicable) here.


*ICCBB 2023, December 11–13, 2023, Kuala Lumpur, Malaysia*

ACM ISBN 979-8-4007-1633-1/23/12
https://doi.org/10.1145/3638569.3638573

twice daily for a duration of three weeks, resulting in their germination into plantlets, each displaying 3 to 4 leaves. During this stage, the plantlets were divided into two groups. The first group was labeled as the control, with regular watering maintained. In contrast, the second group was designated as the drought group, and watering was completely halted until the plants exhibited wilting [2].

## 2.2 RNA Extraction and Library Preparation

Leaf samples were collected from both the control and drought-treated groups at week 5 for RNA extraction. The RNA extraction procedure involved the use of TRIzol®reagent (Invitrogen, USA) with some modifications. Initially, frozen leaf samples were crushed in a mortar and then suspended in TRIzol reagent. Subsequently, the mixture was heated at 37℃ for 2 minutes, followed by a 5-minute incubation at room temperature. The sample was then centrifuged at 12000 rpm for 10 minutes at 4℃. The aqueous phase was carefully transferred into a new tube, and chloroform was added. After vortexing and a 3-minute incubation at room temperature, the tube was centrifuged again at 12000 rpm for 20 minutes at 4℃. The resulting aqueous phase was pipetted out, mixed with isopropanol in a 1:1 ratio, and left at room temperature for 10 minutes. Subsequently, the mixture was incubated overnight at -20℃.

Following incubation, the tube was centrifuged at 12000 rpm for 15 minutes at 4℃. The aqueous phase was carefully removed, and the pellet in the tube was washed with 80% ethanol. Subsequently, the tube was re-centrifuged at 7500 rpm for 5 minutes at 4℃. The ethanol was removed, and the pellet was air-dried for 5 minutes before being dissolved in RNase-free water. The RNA sample then underwent treatment with the TURBO DNA-freeTM kit (Invitrogen, USA) to remove the DNA genome according to the product manual. TURBO DNase buffer and DNase were added to the RNA sample, mixed gently, and incubated at 37℃ for 30 minutes. Following this, DNase inactivation reagent was added, and the solution was incubated at room temperature for 2 minutes before centrifugation at 12000 rpm for 1.5 minutes. The RNA was then transferred to a new tube. The quality of the RNA was assessed using a Nanodrop Spectrophotometer (Thermo Scientific, United Kingdom) and gel electrophoresis (Applied Biosystem, USA). The sample then sent to Azenta Life Science (China) for deep sequencing using paired end sequencing Illumina Novax (Illumina).

## 2.3 Bioinformatic analysis

In pre-analysis, Cutadapt was employed to process pass-filter data in fastq format, eliminating technical sequences such as adapters, polymerase chain reaction (PCR) primers, or their fragments, as well as bases with a quality score lower than 20. For alignment, reference genome sequences and gene model annotation files for the relevant species were obtained from genome websites such as UCSC, NCBI, and ENSEMBL. Subsequently, the reference genome sequence was indexed using Hisat2 (v2.0.1). Lastly, the clean data were aligned to the reference genome using the Hisat2 (v2.0.1) software.

For expression analysis, transcripts in fasta format were generated from a known gff annotation file and appropriately indexed. Subsequently, utilizing this file as a reference gene file, HTSeq (v0.6.1) calculated gene and isoform expression levels from the pair-end clean data. Finally, differential expression analysis was conducted using the DESeq2 Bioconductor package, which employs a model based on the negative binomial distribution. The estimates of dispersion and logarithmic fold changes incorporate data-driven prior distributions, and a threshold of Padj < 0.05 for genes was applied to identify differentially expressed ones.

# 3 RESULTS & DISCUSSION

## 3.1 Sequencing Data

Illumina sequencing of the rice transcriptome generated approximately 37,400,627 ± 8 million reads in the control and 43,382,383 ± 0.5 million reads in the drought-treated samples. After mapping, there were 30,176,961 ± 7 million reads in the control library and 35,173,835 ± 0.4 million reads in the treated library. Unique reads represent non-redundant sequences, indicating the number of individual sequences in the dataset. The details of the transcriptome statistics are presented in Table 1 below.

Based on Table 1, around 80.53% and 81.08% of control and drought-treated libraries has been uniquely map to the rice reference genome. The mapping cannot be done completely 100% to the reference due to the several possibilities. First, it may cause by alternative splicing. Eukaryotic genes undergo alternative splicing, a process where different combinations of exons can be included or excluded from the final mRNA transcript [4]. This results in multiple mRNA isoforms from a single gene. As a consequence, the transcripts may not perfectly match the reference genome sequence. Second, the transcriptome may contain novel transcripts since the library has been constructed using new varieties of rice. Not all transcripts present in a particular cell or tissue may be represented in the reference genome. These transcripts would not align to the reference genome. Third, it's may due to the genetic variations such as single nucleotide polymorphisms (SNPs) or small insertions/deletions, which can lead to differences between the transcript and the reference genome [5]. If the transcriptome data comes from individuals with genetic variations not present in the reference genome, this can affect the mapping. Next is technical challenges. The sequencing process itself may introduce errors, and the quality of the sequencing data can affect the accuracy of the mapping [6]. Short read lengths or sequencing errors can make it difficult to align reads to the reference genome.

## 3.2 Gene Expression Analysis

In this study, a total of 3960 genes were identified as drought-responsive. Of these, 2325 were upregulated whereas 1635 were downregulated. The top 10 upregulated and downregulated genes were demonstrated in Figure 1 and Figure 2 below.

Based on Figure 1, the most upregulated gene due to the drought treatment is OsLTPd6 (nonspecific lipid transfer protein 6). To the best of our knowledge this gene is understudied in rice. However, a study indicated that a lipid transfer protein promotes salt and drought stress tolerance in *N.tabacum* [7]. Besides a group of LEA (Late embryogenesis-abundant) genes also showed upregulation. The LEA gene is crucial in mediating the response to abiotic stress in various plants such as *O. sativa*, *A. thaliana*, *P. trichocarpa*, and

**Table 1: statistics of rice transcriptome in control and treated libraries**

| Type of library | Total reads | Total mapped | Uniquely mapped |
|---|---|---|---|
| **Control** | 37400627±8392606 | 32842876±7718513 **(87.65%)** | 30176961±7099263 **(80.53%)** |
| **Drought-treated** | 43382383±481399 | 38132655±471940 **(87.90%)** | 35173835±447644 **(81.08%)** |

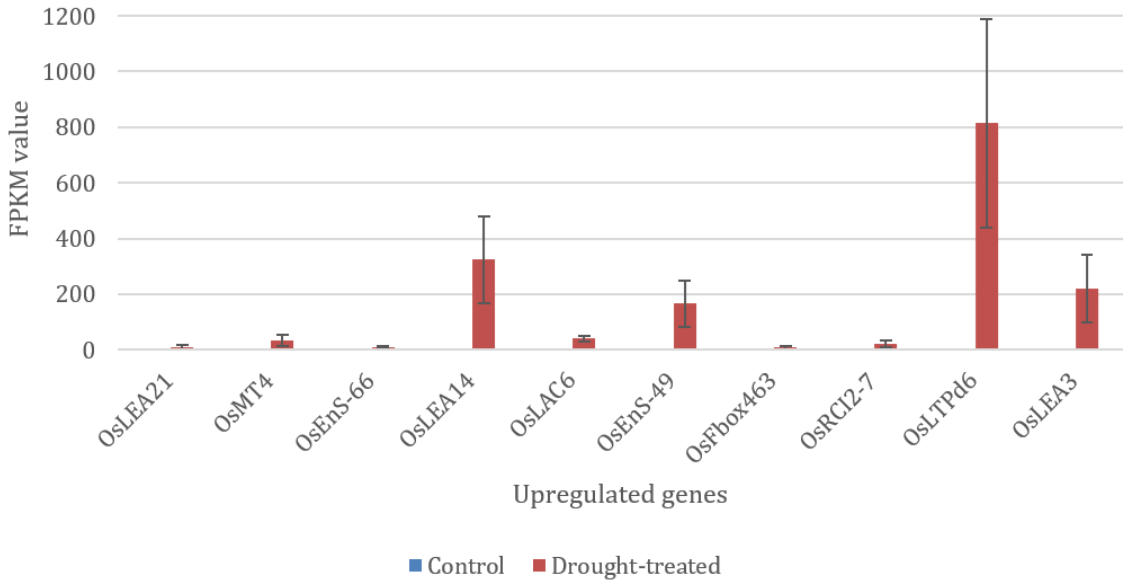


**Figure 1: Top 10 upregulated gene responsive to drought treatment**

others [8]. Other upregulated genes include those encoding L-ascorbate oxidase, plant EC metallothionein, and F-box proteins.

On the other hand, peroxidase 47 is the most downregulated gene due to drought treatment. Peroxidase is one of the superoxide dismutase enzymes critical in detoxifying oxidative damage during the abiotic stress response. Elevated levels of superoxide dismutase (SOD) during periods of drought stress have been observed in various plant species, including rice, peas, wheat, sunflower, beans, and sweet potatoes [2]. However, in rice not all peroxidases were upregulated. Several peroxidases in rice roots, including LOC_Os05g04470.1, LOC_Os07g48030.1, LOC_Os07g48060.1, and LOC_Os04g59260.1, experience repression [9]. Further study needs to be done in order to elucidate the downregulated peroxidase during drought. Additionally, other significantly downregulated genes include those encoding germin-like protein (GLP), leucine-rich repeat (LRR), and amino acid carrier (AAP).

## 4 CONCLUSION

In conclusion, this study employs next-generation technology to identify genes responsive to drought stress in rice. The data generated in this research can serve as a fundamental basis for the development of drought-resistant rice varieties.

## ACKNOWLEDGMENTS

The authors express their gratitude for the financial support received from the Malaysian Ministry of Higher Education (Malaysia) through the Fundamental Research Grant Scheme (FRGS/1/2020/STG01/UTM/02/6). Additionally, the authors extend their appreciation to Mr. Hassan Ahmad (MARDI Branch Tanjong Karang), Madam Syuhaidah Abu Bakar, and Miss Rahiniza Kamaruzzaman (MARDI HQ) for supplying the rice seeds for the study.

## REFERENCES

[1] Giller, K. E., Delaune, T., Silva, J. V., Descheemaeker, K., van de Ven, G., Schut, A. G. T., van Wijk, M., Hammond, J., Hochman, Z., Taulya, G., Chikowo, R., Narayanan, S., Kishore, A., Bresciani, F., Teixeira, H. M., Andersson, J. A. and van Ittersum, M. K. The future of farming: Who will produce our food? Food Security, 13, 5 (2021/10/01 2021), 1073-1099.

[2] Panda, D., Mishra, S. S. and Behera, P. K. Drought Tolerance in Rice: Focus on Recent Mechanisms and Approaches. Rice Science, 28, 2 (2021/03/01/ 2021), 119-132.

[3] Naqvi, R. Z., Siddiqui, H. A., Mahmood, M. A., Najeebullah, S., Ehsan, A., Azhar, M., Farooq, M., Amin, I., Asad, S., Mukhtar, Z., Mansoor, S. and Asif, M. Smart breeding approaches in post-genomics era for developing climate-resilient food crops. Frontiers in Plant Science, 13 (2022-September-16 2022).

[4] Baralle, F. E. and Giudice, J. Alternative splicing as a regulator of development and tissue identity. Nat Rev Mol Cell Biol, 18, 7 (Jul 2017), 437-451.

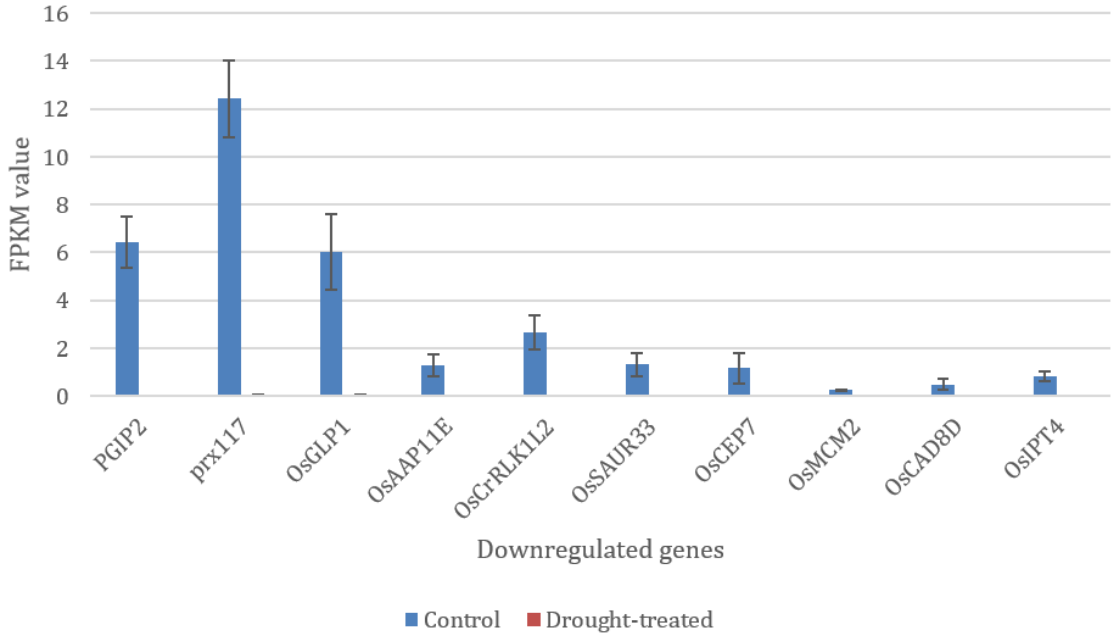


**Figure 2: Top 10 downregulated regulated gene responsive to drought treatment**

[5] Robert, F. and Pelletier, J. Exploring the Impact of Single-Nucleotide Polymorphisms on Translation. Front Genet, 9 (2018), 507.
[6] Cheng, C., Fei, Z. and Xiao, P. Methods to improve the accuracy of next-generation sequencing. Front Bioeng Biotechnol, 11 (2023), 982111.
[7] Xu, Y., Zheng, X., Song, Y., Zhu, L., Yu, Z., Gan, L., Zhou, S., Liu, H., Wen, F. and Zhu, C. NtLTP4, a lipid transfer protein that enhances salt and drought stresses tolerance in Nicotiana tabacum. Sci Rep, 8, 1 (2018/06/11 2018), 8873.
[8] Chen, Y., Li, C., Zhang, B., Yi, J., Yang, Y., Kong, C., Lei, C. and Gong, M. The Role of the Late Embryogenesis-Abundant (LEA) Protein Family in Development and the Abiotic Stress Response: A Comprehensive Expression Analysis of Potato (Solanum Tuberosum). Genes (Basel), 10, 2 (Feb 15 2019).
[9] Nadarajah, K. and Kumar, I. S. Drought Response in Rice: The miRNA Story. Int J Mol Sci, 20, 15 (Aug 1 2019).