Journal of Research High School

Vol. 2026 (6) 451 – 456
ISSN 2688-3651 [online]

JRHS Outstanding Research Paper Award

Increasing Drought Resistance in Rice through Targeted Promoter Engineering

Luke Jiang[1] *

[1]Waunakee Community High School, Waunakee, WI, USA
*Corresponding Author: lukejiang09@gmail.com

Advisor: Zhao Wang, wangzhao@genetics.ac.cn

Received June 19, 2025; Revised November 29, 2025; Accepted January 7, 2026

Abstract

Drought is a significant threat to global food production. Genetic modification presents a swift alternative to artificial selection, producing desirable traits by targeted modulation of gene expression. Many genes that impart drought resistance have been identified in rice by transgene overexpression studies. Leveraging variations in regulatory regions of such genes presents an opportunity to further increase expression, creating enhanced drought-resistant rice varieties. This study seeks to first review previous literature to identify genes that increase drought resistance when overexpressed, then find upstream variations that cause the greatest increase in chromatin accessibility. Possible target sites for clustered regularly interspaced short palindromic repeats (CRISPR)-associated endonuclease 9 (Cas9) editing are located near these identified variation sites. A library of possible insertions and deletions generated by simulated Cas9 editing are generated to modify regulatory regions of these genes, producing new alleles that increase chromatin accessibility beyond any natural variation documented within the RiceVarMap 2 database. This collection of *in silico* regulatory alleles provides a basis for fine-tuning gene expression to increase drought resistance in rice through future *in planta* editing, allowing for more rapid adaptation to the increasing prevalence of drought due to global climate change.

Keywords: Biology, Genetics, Drought resistance, CRISPR, Rice, Agriculture

1. Introduction

Agricultural output must increase to accommodate a rapidly growing world population, projected by the United Nations to increase to 9.7 billion by 2050 (United Nations, 2024). However, this is increasingly challenged by elevated levels of abiotic stresses from global climatic change. Many current rice varieties adapted to traditional cultivation are especially sensitive to drought-related stresses. Breeding techniques that rely on slow cross-generation artificial selection may not be able to deliver more stress resistant varieties to be sustainably farmed in the future in a sufficient timeframe. Rice is widely grown and provides a significant portion of the world's food supply (Ritchie et al., 2023). Consequently, any significant threat to the production of this staple crop would likely impact the food supply of a large proportion of the world population. A potential solution is direct genetic modification of the rice genome to create novel drought-resistant strains. This may be an effective method to combat abiotic stresses, as traditional artificial selection methods may not be rapid enough to accommodate growing demands and increasing stressors. Direct genetic engineering for resistance to abiotic stresses in rice allows it to remain viable in a changing world.

This study aims to first identify genes that, when transgenically overexpressed in *Oryza sativa* Japonica, lead to increased drought resistance. Genetic variations in regions upstream of these genes that increase chromatin accessibility through improved promoter binding lead to increased gene expression. These variations and their surrounding regions were then selected to characterize the potential for direct genetic modification that could lead to increased drought resistance. Potential target sites for genome editing with CRISPR-Cas9 were identified at and near

Vol. 2026 (6) 451 – 456
ISSN 2688-3651 [online]

each variation. The potential alleles generated by Cas9 at these target sites were then predicted to serve as an "allele library" to test their effects in increasing drought resistance in rice with future editing. It is hypothesized that new variations generated in the promoter regions of these genes will produce artificial alleles with a greater range of chromatin accessibility scores than existing natural alleles, allowing for the selection of plants with elevated expression levels of these genes and thus greater drought resistance.

2. Materials and Methods

The RiceVarMap 2.0 database (Zhao et al., 2014) was utilized for searching for variations in the selected genes and their chromatin accessibility scores in various tissues of the plant, in particular identifying the potential change of accessibility caused by natural variations. The inDelphi editing result predictor (Shen et al., 2019) was used to scan the variation and surrounding regions for potential Cas9 target sites, and to predict alleles to be generated by targeting these regions.

First, genes that contribute to drought resistance when transgenically overexpressed in rice were identified and vetted through lists compiled in previous studies (Oladosu et al., 2019; Yang et al., 2022). These genes were selected due to the detailed literature present regarding them, as well as robust documentation of variations within the RiceVarMap 2.0 database. Four genes were selected for their variations to be analyzed, LOC_Os06g04070, LOC_Os01g07120, LOC_Os02g50970, and LOC_Os03g12820. Afterward, upstream variations for these genes were searched and identified in the RiceVarMap 2.0 database.

Using this information, the average chromatin accessibility scores of the variations across all tissues were calculated, and the single variation with the greatest accessibility increase relative to the reference allele was identified for each gene. The average score across tissue types was chosen, instead of the tissues where expression of the gene is the highest, because tissue types were not strictly defined and inconsistent across data platforms, such as the expression atlas, variant database, as well as *in silico* prediction of chromatin accessibility changes. A segment of the genome 100 bp upstream and downstream of these selected variations was extracted from the rice reference genome on the JBrowse (Diesh et al., 2023) genome browser. Each 200-bp sequence was placed into the inDelphi editing result predictor, a machine learning model that predicts the results of DNA repair with "messy" end-joining after a double-strand break by Cas9. The default options were used: including 1-bp insertions in the predicted alleles, and all deletions regardless of if they were microhomology-mediated or not (produced from joining the homologous sequences revealed by peeling back one end of either side of the double-strand break), which identified all instances of the protospacer adjacent motif NGG, then generated predictions on the potential genotypes resulting from editing.

Statistical analysis was performed using R Statistical Software (v4.3.3; R Core Team 2024), a programming language. Packages were downloaded to extend the capabilities of the software, with the readr package used for file type conversions (v2.1.5; Wickham et al., 2024), and the ggplot2 package for data visualization (v3.5.2; Wickham, 2016). After collecting chromatin accessibility change scores (alternative vs. reference allele) of the upstream variations in each of the genes, the mean, median, as well as the maximum and minimum values, were calculated. These data, containing a list of possible insertions and deletions, as well as their frequencies, were summarized from inDelphi predictions, including the number of alleles per site, unique insertions and deletions per site, and the lengths of those insertions and deletions.

The downloaded genotype editing predictions on each selected site were combined into tables for all variations to summarize the total number of possible alleles predicted. The frequency, or the predicted chance that a cut at the NGG site would result in a particular allele after repair (Figure 1), and length distributions of the alleles on each gene (Figure 2) were visualized using ggplot2 (v2.1.5; Wickham et al., 2024).

3. Results

Four genes were selected because of their reported effects to increase drought resistance when overexpressed (Table 1). LOC_Os06g04070, also known as *ADC* (arginine decarboxylase), is a gene from oat that has been proven to protect chlorophyll in drought conditions when overexpressed in transgenic rice, and is mostly expressed in young

Vol. 2026 (6) 451 – 456
ISSN 2688-3651 [online]

panicle, stamen, and pistil tissues (Capell et al., 1998).

LOC_Os01g07120, or *DREB2*, is a sorghum transcription factor that also improves drought tolerance when transgenically expressed in rice with a stress-induced promoter, leading to a greater amount of panicles compared to a control. It is primarily expressed in the root, stamen, and pistil tissues (Bihani et al., 2010).

DsM1, numbered LOC_Os02g50970, is a mitogen-activated protein kinase that increases resistance to oxidative stress, leading to greater drought tolerance. Overexpression increases dehydration tolerance, especially in seedlings. Expression is greatest in the root and young leaf tissues (Ning et al., 2009).

Finally, radical-induced cell death protein OsSRO1c LOC_Os03g12820, is expressed in guard cells under drought stress, leading to stomatal closure. When overexpressed, water loss is decreased, though an undesirable phenotype of increased oxidative stress is also present. *OsSRO1c* is most prominently expressed in the root, young leaf, lemma and palea tissues (You et al., 2012). These genes all contribute to abiotic stress resistance, especially drought and salinity.

Table 1. Product, phenotypic influence, and number of catalogued upstream variations of the four selected genes.

Gene	Accession number	Product	Function	Variations	Citation
ADC	LOC_Os06g04070	Arginine decarboxylase	Reduces chlorophyll loss in drought conditions	122	Capell et al., 1998
DREB2	LOC_Os01g07120	Transcription factor	Improves yield of grain during water stress	190	Bihani et al., 2011
DsM1	LOC_Os02g50970	Mitogen-activated protein kinase	Regulates drought responses though regulation of reactive oxygen species (ROS) scavenging	66	Ning et al., 2010
OsSRO1	LOC_Os03g12820	Similar to radical-induced cell death one (SRO) protein	Signals for stomatal closure and hydrogen peroxide gathering, reducing water loss	63	You et al., 2013

The upstream variants of these genes (within 2000 bp of the 5' end) were identified in the RiceVarMap 2.0 database and screened for improvements in chromatin accessibility. As the RiceVarMap did not have consistent tissue naming and classification between the expression atlas and chromatin accessibility, specific tissues in which an increase in accessibility might be most effective cannot be identified. Consequently, the variation with the largest mean increase of accessibility scores across all tissues was identified in each gene as candidates for future genetic modification (Table 3).

Table 2. Summary statistics for mean chromatin accessibility across tissues per variation and across by gene.

Gene	Accession number	Mean	Minimum	Maximum
ADC	LOC_Os06g04070	0.003	-0.130	0.210
DREB2	LOC_Os01g07120	-0.003	-0.207	0.200
DsM1	LOC_Os02g50970	0.000	-0.060	0.198
OsSRO1	LOC_Os03g12820	0.030	-0.130	0.145

Table 3. Selected variations and their associated chromatin accessibility score changes

LOC_Os06g04070				
Variation	**Position**	**Nucleotide change**	**Chromosome**	**Number of NGG sites within 100 bp**
vg0601678306	1678306	CATATCACATCAG - C	6	11
LOC_Os01g07120				
Variation	**Position**	**Nucleotide change**	**Chromosome**	**Number of NGG sites within 100 bp**
vg0103356052	3356052	ATATAT - A	1	4
LOC_Os02g50970				
Variation	**Position**	**Nucleotide change**	**Chromosome**	**Number of NGG sites within 100 bp**
vg0231170670	31170670	AAACGCCACGTC - A	2	19
LOC_Os03g12820				
Variation	**Position**	**Nucleotide change**	**Chromosome**	**Number of NGG sites within 100 bp**
vg0306894180	6894180	T - TC	3	22

Vol. 2026 (6) 451 – 456
ISSN 2688-3651 [online]

Summarizing chromatin accessibility change scores (alternative vs. reference alleles, Table 2), it was found that negative and positive changes in chromatin accessibility were roughly equal in frequency and magnitude, resulting in mean and median scores close to zero. In three genes, *ADC*, *DREB2*, and *DsM1*, the maximum score was significantly greater than any other variation. All selected variations were insertion or deletion mutations. Note that on the gene LOC_Os01g07120, three unique variations on the same location produced similarly high scores.

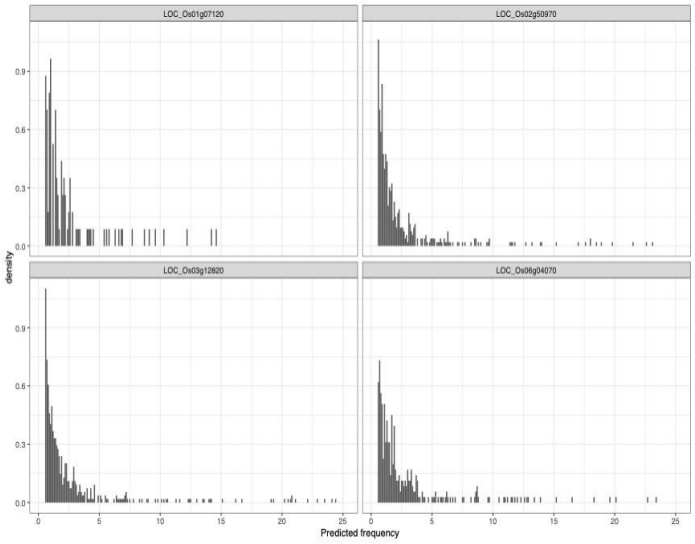

Figure 1. Distribution of frequencies (in percentages) of all simulated editing alleles by gene. Alleles with frequency below 1% were omitted.

Summarizing frequencies of predicted edited alleles, it was observed that the vast majority of predicted alleles (97.7%) would have a frequency below 0.5%, most of which had frequencies below 0.1%. Alleles with this frequency are unlikely to be generated and observed in real *in planta* editing experiments. The distribution of remaining frequencies showed a clear trend, with there only being several alleles with a predicted frequency above 5%, and the vast majority of all potential alleles having very low frequencies.

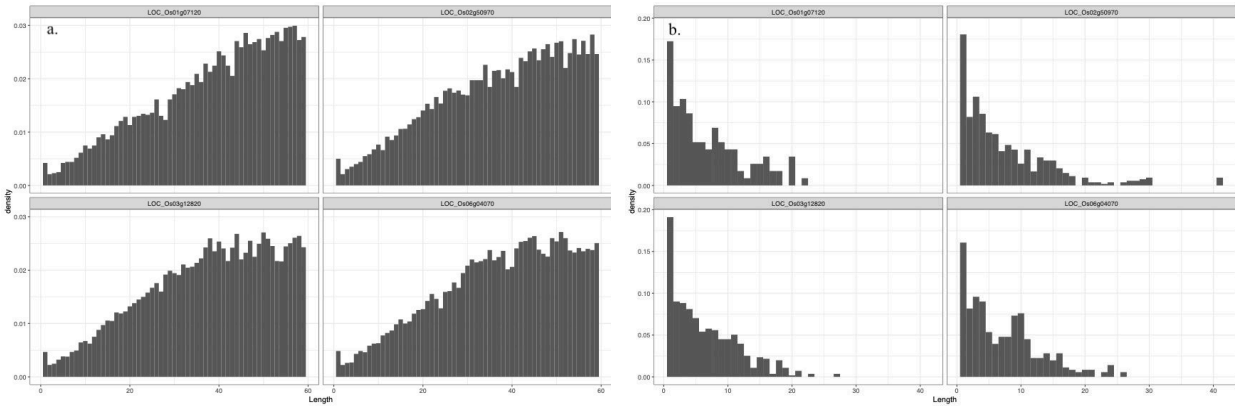

Figure 2. Density plot of the distribution of lengths of all generated alleles by gene. Note that the higher-than-expected frequencies of alleles with a length of 1 is due to the inclusion of the four 1-bp insertions. a. Distribution for all alleles. b. Distribution of all alleles with frequencies over 0.5%.

Summarizing the length distribution of the predicted alleles, the indel alleles generated in a range of lengths between 0 to 50bp, with a clear bias towards the smaller size. Alleles between 10 to 30bp are the most abundant. As length increases, the number of unique alleles increases as well. Combining both size and frequency, longer alleles were more numerous, but almost always had very low frequencies. Single-bp insertions and deletions were uniformly common, almost always having frequencies above 0.5%.

4. Discussion

Across these variations, those with the greatest effects on chromatin accessibility were insertion or deletion mutations that impaired the function of the promoter regions, significantly impacting chromatin accessibility by preventing the binding of transcription factors. However, insertions and deletions were also responsible for the most significant positive effects on chromatin accessibility, as modifying the distance from the promoter and the gene potentially allows for more optimal binding of transcription factors and RNA polymerase, increasing expression levels

Vol. 2026 (6) 451 – 456
ISSN 2688-3651 [online]

of the protein and resulting in a more drought-resistant phenotype. As demonstrated by several studies (Bihani et al. 2010; Capell et al., 1998), overexpression of the selected genes allow rice plants to better tolerate the stressors of heat and salinity. While longer insertions and deletions generally were more likely to result in significantly increased scores, the single base pair insertion on LOC_Os03g12820 with the highest total score demonstrates that there are outliers to this trend. Additionally, very large variations did not seem to produce a greater effect on chromatin accessibility compared to relatively shorter ones, with 60~80bp changes having similar scores to 10~20bp variations.

CRISPR-Cas9 editing system suffers from several limitations; the agrobacterium-mediated transformation of guide RNAs and Cas9 is not efficient and not optimized in some plant systems (but the process is relatively mature and efficient in rice). Additionally, due to the randomness in the repair process after a double-strand break, there is a lack of control over the alleles produced, as is apparent from the frequency distributions. This, however, is not a major limitation, as these alleles are generated in regulatory regions rather than coding regions; it is unlikely for any of these alleles to be highly deleterious. In contrast, the more alleles generated, the more likely beneficial alleles will arise. Still, this would realistically only be likely to produce a small proportion of the possible changes on each site due to general low predicted allele frequency; however, when compared to the catalogued natural variations, even this limited range of artificial alleles would still represent much greater genetic diversity.

The rarity of alleles with larger size changes likely will not impact the increased variation in chromatin accessibility resulting from actual editing, as the length of natural variations did not strongly correlate with chromatin accessibility score changes, especially for the alleles that are over 20 bp in length. Removing alleles with frequencies below 0.5%, it becomes clear that longer alleles are generally much less common, minimizing the potential deleterious effects of edited alleles.

These results indicate that *in planta* editing will be able to generate a number of unique alleles from hundreds of possible alleles per site, and that the vast majority of these alleles would be less than 5bp in length. This is vastly more genetic diversity and potential changes in chromatin accessibility than existing natural variation, as the majority of those were single-base pair substitutions. Editing at any one site close to the identified variations would be more than sufficient to screen potential increases in chromatin accessibility. To optimize for maximum generation of genetic diversity, efforts should be focused on single genes at a time. Once reporter assays for these genes are established and the edited alleles with significant increases in chromatin accessibility and expression are screened, “stacking” of these multiple edits in multiple genes can be further explored, combining the edited promoter regions with the greatest increase in chromatin accessibility across all genes into a single genotype. As specific alleles are identified, CRISPR-Cas9 can introduce them through specific homology-directed repair, instead of the nonhomologous end joining used to generate a variety of alleles. Through successful “stacking”, the final resulting plant should have increased chromatin accessibility beyond any natural variation on all four genes. These strains then can be subjected to field trials of drought resistance to develop the varieties with enhanced drought resistance.

5. Conclusion

Using databases and predictive tools, a library of potentially drought-resistant alleles was generated, and potential insertions and deletions from cleavage by Cas9 catalogued. This library serves as a resource for potential future editing with CRISPR-Cas9 to experimentally create more drought-resistant rice plants. Editing at target sites within the promoter regions will be able to generate a wide range of insertion and deletion mutations, causing significant variation in chromatin accessibility and potentially producing a more drought-resistant phenotype. Edited rice lines, after confirming the edited alleles at the targeted promoter regions by sequencing, can be subjected to phenotypic screening to select for strains that display enhanced drought resistance beyond what natural variations could provide.

References

Bihani, P., Char, B., & Bhargava, S. (2010). Transgenic expression of sorghum *dreb2* in rice improves tolerance and yield under water limitation. *The Journal of Agricultural Science*, *149*(1), 95–101. https://doi.org/10.1017/s0021859610000742

Vol. 2026 (6) 451 – 456
ISSN 2688-3651 [online]

Capell, T., et al. (1998). Over-expression of the oat arginine decarboxylase cdna in transgenic rice (Oryza sativa L.) affects normal development patterns in vitro and results in putrescine accumulation in transgenic plants. *Theoretical and Applied Genetics*, *97*(1–2), 246–254. https://doi.org/10.1007/s001220050892

Diesh, C., et al. (2023). JBrowse 2: A modular genome browser with views of Synteny and structural variation. *Genome Biology*, *24*(1). https://doi.org/10.1186/s13059-023-02914-z

Ritchie, H., Rosado, P., & Roser, M. (2023) - "Agricultural Production" Published online at OurWorldinData.org. Retrieved from: 'https://ourworldindata.org/agricultural-production' [Online Resource]

Ning, J., et al. (2009). A RAF-like MAPKKK gene *dsm1* mediates drought resistance through reactive oxygen species scavenging in rice   *Plant Physiology*, *152*(2), 876–890. https://doi.org/10.1104/pp.109.149856

Oladosu, Y., et al. (2019). Drought resistance in rice from conventional to molecular breeding: A Review. *International Journal of Molecular Sciences*, *20*(14), 3519. https://doi.org/10.3390/ijms20143519

R Core Team (2024). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Shen, M., et al. (2019). Author correction: Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature*, *567*(7746). https://doi.org/10.1038/s41586-019-0938-4

United Nations. (n.d.). *Population*. United Nations. https://www.un.org/en/global-issues/population
H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Wickham, H., Hester, J., & Bryan, J. (2024). _readr: Read Rectangular Text Data_. R package version 2.1.5, <https://CRAN.R-project.org/package=readr>.

Yang, Y., et al. (2022). Enhancement of heat and drought stress tolerance in rice by genetic manipulation: A systematic review. *Rice*, *15*(1). https://doi.org/10.1186/s12284-022-00614-z

You, J., et al. (2012). The SNAC1-targeted gene OSSRO1C modulates stomatal closure and oxidative stress tolerance by regulating hydrogen peroxide in rice. *Journal of Experimental Botany*, *64*(2), 569–583. https://doi.org/10.1093/jxb/ers349

Zhao, H., et al. (2014). RiceVarMap: A comprehensive database of rice genomic variations. *Nucleic Acids Research*, *43*(D1). https://doi.org/10.1093/nar/gku894