Original Research Article

Exploring Genetic and Molecular Diversity of Indian Rice Landraces: A Molecular Marker - DrivenStudy Incorporating D^2 Analysis

Comment [H1]: There are several studies in which the genetic and molecular diverities of rice genotypes; while how author differentiate this study from other study is questioonable. The manuscript was well written and can be accpeted with minor correction. The MS alos need to be revied by subject specialist to varify the technical aspect of MS.

ABSTRACT

Comment [H2]: Adequate

Assessing genetic diversity is the key factor for enhanced crop breeding programme which aids in improving desirable characteristics in the cultivars. In the present study,200 rice landraces were initially screened using PEG at -7 bar concentration for drought tolerance. From the 200 landraces, 12 lines with 50 % germination were chosen to assess molecular and genetic divergence. A total of 50 SSR markers were utilized across the identified 12 drought tolerant genotypes along with two susceptible and two tolerant checks to assess divergence among the genotypes. Among the 50 SSR markers, 49 exhibited polymorphisms with a PIC value ranged from 0.82 (RM447) to 0.30 (RM418) with an average of 0.63. The number of alleles varied from two (RM408, RM418, RM512) to eight (RM404) with an average of 4.4 allele per marker and a total of 216 alleles were observed. Based on unweighted pair group of arithmetic mean (UPGMA) method a dendrogram was constructed and the genotypes were grouped in four main clusters. Among the four, cluster I holding more number of genotypes than cluster IV with high dissimilarity coefficient of 0.88. Further D^2 analysis for five drought traits at seedling stage revealed the similarity and diversity among the landraces by separating them in different clusters based on genetic distance. The highest inter cluster distance of 175.96 were noticed between cluster 2 and 3 with high divergence which aids in better selection of landraces. Both molecular and genetic diversity shows distinct divergence among the genotypes which exhibits broader genetic base with wider adaptability.

Keywords: Genetic Diversity, SSR, UPGMA, D^2 Analysis, Polymorphism, Rice

1. INTRODUCTION

Comment [H3]: Cite some earlier studies in which screening of rice for genetic and molecular diversity were done.

Rice holds a prominent position in Indian agriculture and contributes significantly to the country's economy being the primary source of nutrition for two thirds of the world's population. Rice has a small genome size of 430 Mb [1] therefore, it is accessible to sequence the entire genome with highly saturated molecular markers. Rice is one of the most perfect model plant for studying the grass genetics and genome organization. Rice is a highly diversified crop that is grown in various ecological conditions [2]. In India, about 60 per cent of the arable land is rain-fed, and the cultivation of crops is at serious risk due to climate change. Abiotic stress is one of the major factors that limiting crop growth and yield causing detrimental effects. However, the occurrence of extreme drought, poses a significant threat to rice yields and global food security [3]. Drought tolerance is a complex trait controlled by polygenes. Therefore, improvement for drought tolerance is a huge challenge. The severity of drought depends on several factors, including precipitation patterns, evaporation rates and soil moisture retention capacity [4]. PEG mediated drought screening is the *invitro* screening of plants, simulating drought conditions by reducing water potential during seed germination and growth to test their ability to withstand drought stress. To address the challenge of expanding rice cultivation in water-scarce areas, it is crucial to identify genotypes that can thrive under drought conditions [5]. Different germplasm with distinct genetic makeup promises a future improvement of rice cultivars against drought stress [6]. The adaptability of rice germplasm to abiotic stresses varies; whereas many genotypes are highly vulnerable, only some have the capacity to withstand intense drought stresses. Therefore, assessing the genetic diversity of rice cultivars becomes essential to establish relationships among different varieties to develop effective breeding programs. The success of crop improvement programs relies on the degree of genetic variation and the transmission of desirable traits.The knowledge about genetic divergence and heritability aids plant breeders in predicting the characteristics of the next generation and determining the extent of genetic improvement brought about by selection [7].$D^2$is a statistical analysis helps to identify the genetic variation among and within the genotypes and aids in better selection of highly divergent parents for the hybridization programs.

Studying genetic diversity through molecular markers facilitates the development of improved recombinants, a key component for crop improvement programs aids in identifying novel alleles. Abundant genetic diversity available within the rice landraces is highly critical to exploit [8] but analyzing at DNA level could effectively hasten the breeding programme aimed at commercially grown cultivars [9]. The focus of assessing genetic diversity has switched from using morphological markers to utilizing molecular markers as a result of the considerable developments in molecular biology. Being co-dominant and PCR based, SSRs are preferred over other markers for genetic study. To obtain desirable recombinants in the segregating generations, it is crucial to choose diverse, agronomically suitable parents for hybridization. Landraces have a high degree of genetic diversity and important genes for various abiotic stress tolerances since they have not been exposed to long-term selective breeding. Hence, this study mainly focused to identify genetic divergence in terms of drought traits and to consolidate the genotypes best suited for futurebreeding programme for drought tolerance.

2. MATERIAL AND METHODS

Comment [H4]: Information on crop growing condition and practices followed for growing rice need to given;

The present experimental material comprised of 200 rice landraces along with two susceptible varieties *viz*., Jaya and IR64 and two drought tolerant checks *viz*., Apo and Wayrarem.The 200 rice landraces were screened under laboratory condition using PEG-6000 at a concentrationof -7, in the Department of Plant Biotechnology, Centre for Plant Molecular Biology and Biotechnology, Tamil Nadu Agricultural University, Coimbatore.A total of 12 genotypes were identified for further studies along with checks (**Supplementary Table 1**). The genotypes along with checks wereraised in field during Summer, 2023 in two replications following Randomized Block Design at Paddy Breeding Station, TNAU, Coimbatore. The 20 days old young leaves were collected and subjected to DNA extraction. The DNA was extracted using the Cetyl Tri-Methyl Ammonium Bromide (CTAB) method [10] and immediately stored in -20^{o}C. The DNA purity and concentration were checked in microvolume spectrophotometer at 260/230 nm and diluted to working concentration of 50 ng/µL with sterile distilled water and stored at -20 ^{o}C.Fifty SSR markers covering the 12 linkage groups were utilized to assess polymorphism among the genotypes **(Supplementary Table 2)**. The sequences of primer pairs were obtained from the GRAMENE database (http://www.gramene.org).A polymerase chain reaction was carried out to selectively amplify the particular segment of genomic DNA *in vitro* to a billion-fold. The PCR reaction consisted of 10 µL volume comprising 2 µL (50 ng/µL) DNA, 1µL of both forward and reverse primers, 5µl of (1X) Master-mix and 2 µL of sterile distilled water. PCR was programmed with an initial denaturation of 94^{o}C for 5 min, followed by 35 cycles of denaturation at 94^{o}C for 1 min., 30 seconds at the annealing temperature of particular primer pair, extension at 72^{o}C for 1min. and final extension of 72^{o}C for 5 min. A constant voltage of 80V was maintained for 2 hrs and the fragments were visualized under UV gel documentation system. Polymorphic Information Content describes the informativeness of the marker.It is the sum total of polymorphism obtained by the markers and was calculated using the formula:

Comment [H5]: ?? It should be three.

PIC = 1-$\sum pi^2$ where, pi is the frequency of i^{th} allele.

The gels were scored based on allelic size obtained by each marker. Using DARwin software 5.0 [11]. Simple matching dissimilarity coefficient was formed, and this matrix was subjected to cluster analysis using UPGMA method. The allanbiotools software [https://allanbiotools.shinyapps.io/pbperfect/] was used to perform phenotypic clustering using D2 analysis to evaluate the genetic variation among genotypes for seedling drought traits. was utilized.The phenotypic information for the five traits recorded during PEG screening at -7 bar concentration**(Table. 1)**for all the genotypes along with checkswere subjected to D^2 analysis to study the genetic divergence among the landraces. A dendrogram and distance plot wereconstructedand given as figures 3 and 4.To explore the genotypic and phenotypic relationship among the landraces both genetic and molecular divergence has been undertaken.

Table.1. PEG screening data of identified landraces at -7 bar concentration

Genotypes	Germination percentage	Shoot length	Root length	Shoot/root ratio	Root/shoot ratio
IC 458581	77	1.4	2.5	0.6	1.8
IC 378202	83	1.0	2.3	0.4	2.3
IC 67496	97	2.3	3.3	0.7	1.4
IC 206282	67	1.1	2.1	0.6	2.1
IC 115406	87	2.6	3.4	0.8	1.3

IC 248033	73	0.7	1.4	0.5	2.4
IC 464685	93	3.5	3.2	1.1	0.9
IC 208155	83	1.6	2.7	0.6	1.7
IC 115439	97	2.6	3.4	0.8	1.3
IC 458210	67	2.1	3.6	0.6	1.9
IC 457996	93	3.4	3.7	0.9	1.1
IC 465008	87	1.7	1.2	1.5	0.7
Apo	80	2.8	3.4	0.8	1.2
Wayreram	83	1.8	5.0	0.4	2.8
IR64	67	0.7	0.6	3.1	0.9
Jaya	67	0.7	0.8	0.9	1.2

**Mean data of three replications for each trait*

3. RESULTS AND DISCUSSION

Comment [H6]: Adequate

3.1 Molecular diversity assessment using SSR markers

The results of screening of 200 rice landraces using -7 bar concentrations of PEG 6000 revealed a selection of 12 drought tolerant genotype which performed similarly to the tolerant checks. These 12 genotypes along with 4 checks were subjected to diversity analysis using 50 SSR markers. Among them, 49 markers exhibited polymorphism (**Fig.1**) and one marker expressed monomorphic banding pattern.

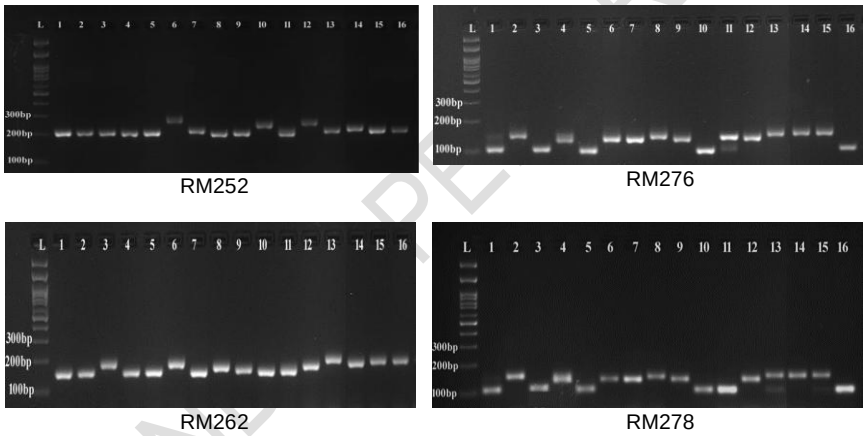

Fig.1. SSR marker profile of sixteen rice genotypes generated by the markers RM252, RM276, RM262 and RM278

L represents ladder and the Number 1 to 16 denotes the genotypes *viz*., 1-IC458210, 2-IC465008, 3-IC67496, 4-IC248033, 5-IC115439, 6-IC457996, 7-IC464685, 8-IC208155, 9-IC458581, 10-IC378202, 11-IC206282, 12-IC115406, 13-APO, 14-IR64, 15-WAYRERAM, 16-JAYA.

Microsatellite markers exhibit considerable genetic diversity per locusdue to their multi-allelism [12]. The PIC value describes the diversity and frequency of alleles among the genotypes. The highest PIC value was observed for the marker RM 447 (0.82) followed by the marker RM 276 (0.76) whereas, the lowest PIC was observed by RM 418 (0.30). The markers exhibiting higher PIC value ought to be used in taxonomical and germplasm characterization studies. The PIC value ranged from 0.82 to 0.30 with mean value of 0.63. Similarly, [13] observed average PIC value of 0.57 with values varying from zero (RM 115) to 0.890 (RM 202) among 40 cultivated varieties and five wild relatives of rice.

A total of 216 alleles were observed from polymorphic markers where, number of alleles ranged from 2 to 8 with an average of 4.46 alleles per marker. The number of alleles observed in the present study was less than the average number of alleles reported by [14] who observed an average of 11.85 alleles per locus in wild rice (*Oryza rufipogon*).Recent findings by [15] indicates that locus with a high PIC value has more alleles per locus in rice.

The degree of divergence present among the cultivars was calculated using a dissimilarity matrix. Genotypes with low dissimilarity ratio will have higher similarity i.e., they are closely related and *vice versa*. The dissimilarity coefficient was ranged from 0.41 to 0.88 (**Supplementary Table 3**). The highest dissimilarity of 0.88 was observed between IR64 and IC67496; Jaya and IC457996 followed by 0.86 between IR64 and IC458210; Jaya and IC67496; IR64 and IC464685. The lowest dissimilarity of 0.41 was observed between IC248033 and IC67496; IC115439 and IC248033, respectively. [16] studied high dissimilarity coefficient of 0.042 between the cultivar LC-4 and IR-82635-B-B-47-1 and identified as highly diverse genotypes.

Dendrogram based on Unweighted Pair Group Method with Arithmetic mean grouped the 16 genotypes into four main clusters. Among the four clusters, cluster I was subdivided into two sub clusters (**Fig.2**) and form the largest cluster consisting of nine genotypes *viz.,* IC115439, IC248033, IC67496, IC464685, IC457996, IC206282, IC378202, IC458581 and IC208155 followed by cluster II with three genotypes viz., Wayreram, Apo and IC115406. The cluster III consisted of IC 465008, IC458210 and cluster IV grouped the susceptible checks Jaya, IR64in one cluster. Among the four clusters, cluster I and cluster IV exhibits higher dissimilarity index among the genotypes. [17] found similar outcomes when they examined the genetic diversity for 34 rice genotypes using three polymorphic SSR markers and grouped them into four clusters.

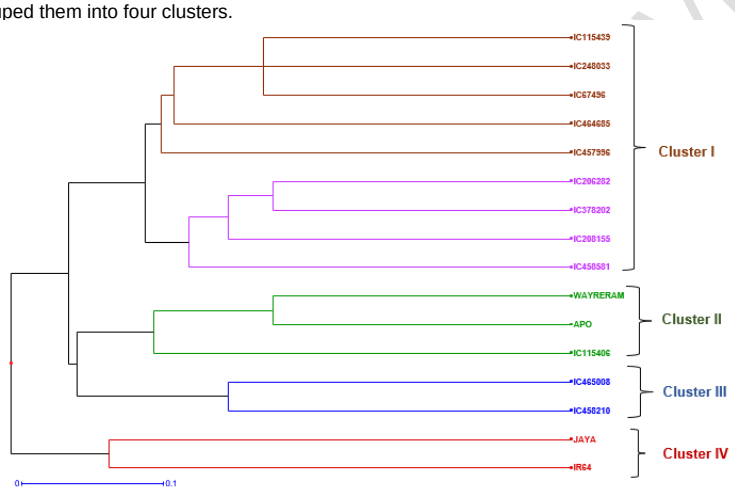

Fig. 2. Dendrogram based on UPGMA for sixteen genotypes

3.2 Clustering analysis to assess the phenotypic diversity for seedling drought traits

The selected 12 drought tolerant genotypes assessed for their ability to withstand drought was measured using five quantitative traits at seedling stage *viz*., germination percentage, shoot length, root length, shoot/root ratio, root/shoot ratio at -7 bar concentration of PEG-6000. The traits were further analyzed for their diversity to congregate the elite genotypes with utmost drought tolerance equivalent to the checks Apo and Wayreram using D^2 analysis. The 16 genotypes were grouped into 2 major clusters (**Fig.3)** differentiating the susceptible checks as one group and the remaining 14 genotypes including checks in three groups. Both the check varieties *viz*., Apo and Wayreram made closer grouping along with the rice landraces IC67496, IC115439, IC458210, IC115406.The selected traits recorded from the PEG screening that highly impacted the clustering pattern and clearly partitioned the tolerant genotypes towards the check varieties.A similarity in the grouping involving phenotypic traits and markers was observed for the genotype IC115406 i.e.,similarity in grouping using allelic dataand phenotypic data was observed for this genotype.

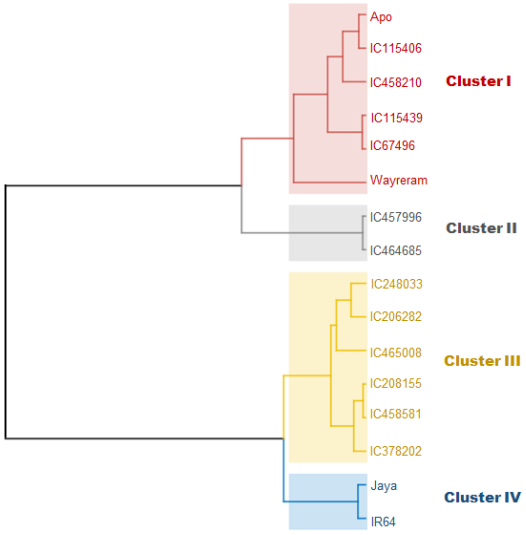

Fig. 3. Cluster Dendrogram of 16 genotypes based on the PEG screening data at -7 bar concentration

The attributes that lead to the greatest divergence should be given more importance when selecting the clusters for the purpose of parents for hybridization programs [18]. The distance plot separated the genotypes into 4 clusters**(Fig. 4)**. The highest inter cluster distance of 175.96 was observed between cluster 2 comprising the genotypes IC464685, IC457996, Apo, IC115406, IC115439, IC67496and cluster 3 with the susceptible checks viz., IR64, Jaya expressing the extremities of the susceptible and tolerant genotypes. The smallest distance of 30.74 was observed betweencluster 1with IC458581, IC208155, IC378202, IC206282, IC465008, IC248033 and cluster 3with IR64 and Jayaindicating the alignment of these selected genotypes towards the susceptible checks and this can be taken as an indicator for selection. The lowest intra cluster distance of 5.11 was observed in cluster 3 followed by cluster 1 and a high intra cluster distance of 29.71 was observed in cluster 4 **(Table.2**).Generally, the higher the inter-cluster, the genotypes present between clusters may display a broad range of genetic variation[19]. Identification of diverse genotypes may result in broadening the genetic base of the cultivated rice by way of creating MAGIC population or involving in pre-breeding programmes [20].

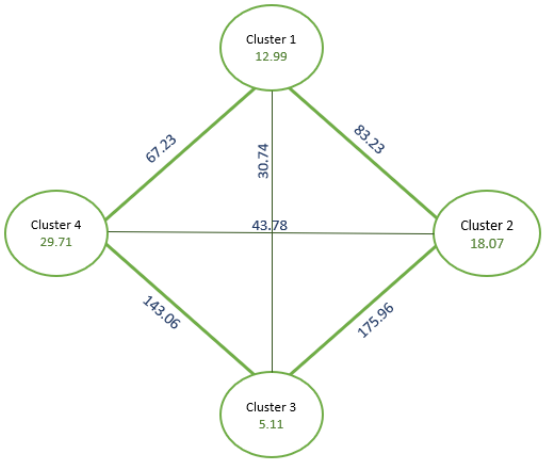

Fig. 4.Distance plot of 16 genotypes depicts intra and inter cluster distance

Table.2. The list of 16 genotypes clustered based on phenotypic traits in distance plot

S. No.	Cluster	Number of genotypes	Genotypes
1	Cluster 1	6	IC458581, IC208155, IC378202, IC206282, IC465008, IC248033
2	Cluster 2	6	IC464685, IC457996, Apo, IC115406, IC115439, IC67496
3	Cluster 3	2	IR64, Jaya
4	Cluster 4	2	IC458210, Wayreram

4. CONCLUSION

Comment [H7]: Conclusion should be concise nad based on the data presented in MS; Aviod writing general statement in conclusion.

This study emphasis the use of 50 SSR markers in differentiating the 16 genotypes. Out of 50 SSR markers, 49 markers exhibited polymorphism and 41 markers showed higher PIC value of greater than 0.5. This revealed that the genotypes and the markers have tight genetic relationships. With regard to the clustering pattern of both molecular and genetic diversity of identified landraces, the obtained molecular markers and the quantitative traits clearly partitioned the landraces *viz.*, IC464685, IC457996, IC115406, IC115439, IC67496 towards tolerant check Apo. Thus, the combined strategy of PEG screening and the molecular characterization of landraces proves to be an efficient and rapid method of selecting genotypes for future breeding programmes. Further,the obtained markers which are highly polymorphic could be utilized to select landraces with diverse genetic background. This investigation combining the allelic diversity and phenotypic diversity at seedling stage has set as an example of rigorous screening of larger population in a limited time period to end up in proven drought tolerant genotypes.

REFERENCES

Comment [H8]: Strictly follow the reference writing style of the journal for both citing references in the text and enlisting the in reference list.

Comment [H9]: ?

1. Causse, M. A., Fulton, T. M., Cho, Y. G., Ahn, S. N., Chunwongse, J., Wu, K., ... & Harrington, S. E. (1994). Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics, 138(4), 1251-1274.
2. Long, S. P., & Ort, D. R. (2010). More than taking the heat: crops and global change. Current opinion in plant biology, 13(3), 240-247.
3. Passioura, J. (2007). The drought environment: physical, biological and agricultural perspectives. Journal of experimental botany, 58(2), 113-117.
4. Wery, J., Silim, S. N., Knights, E. J., Malhotra, R. S., & Cousin, R. (1993). Screening techniques and sources of tolerance to extremes of moisture and air temperature in cool season food legumes. Euphytica, 73, 73-83.
5. Anis, G., Hassan, H., El-Sherif, A., Saneoka, H., & EL Sabagh, A. (2019). Evaluation of new promising rice hybrid and its parental lines for floral, agronomic traits and genetic purity assessment. Pak. J. Agric. Sci, 56, 567-576.
6. Hossain, M. M., Islam, M. M., Hossain, H., Ali, M. S., Teixeira da Silva, J. A., Komamine, A., & Prodhan, S. H. (2012). Genetic diversity analysis of aromatic landraces of rice (*Oryza sativa* L.) by microsatellite markers. Genes, Genomes and Genomics, 6(11), 42-47.
7. Tuhina-Khatun, M., Newaz, M. A., & Bari, M. A. A. (2007). Combining ability and heritability estimates in F2 diallel population of spring wheat under interacting environments. *Bangladesh Journal of Agricultural Sciences*, *34*, 75-82.
8. McCouch, S. R., Temnykh, S., Lukashova, A., Coburn, J., Declerck, G., Cartinhour, S., ... & Li, J. (2001). Microsatellite markers in rice: abundance, diversity, and applications. In Rice genetics IV (pp. 117-135).
9. Wang, B., & Chee, P. W. (2010). Application of advanced backcross quantitative trait locus (QTL) analysis in crop improvement. Journal of plant Breeding and crop Science, 2(8), 221-232.
10. Murray, M. G., & Thompson, W. (1980). Rapid isolation of high molecular weight plant DNA. Nucleic acids research, 8(19), 4321-4326.
11. Perrier, X., & Jaqeuemond-Collet, J. P. (2005). DARwin-5.0. Dissimilarity analysis and representation for windows, Equipe Mathematique et Informatique, France
12. Ferreira, ME, & Grattapaglia, D. (1996). Introduction to the use of molecular markers in genetic analysis.

Comment [H10]: ??

13. Ravi, M., Geethanjali, S., Sameeyafarheen, F., & Maheswaran, M. (2003). Molecular marker based genetic diversity analysis in rice (*Oryza sativa* L.) using RAPD and SSR markers. Euphytica, 133, 243-252.
14. Prathepha, P. (2012). Genetic diversity and population structure of wild rice,*'Oryza rufipogon'*from northeastern Thailand and Laos. Australian Journal of Crop Science, 6(4), 717-723.
15. Ayeella, P. G., Kanneboina, S., Senguttuvel, P., Revathi, P., Kemparaju, K. B., Balakrishnan, D., ... & Prasad, A. H. (2023). Estimation of indica-tropical japonica genome proportion in wide compatible restorer lines derived through inter sub specific hybridization and molecular diversity analysis among rice genotypes. Electronic Journal of Plant Breeding, 14(2), 616-624.
16. Rashmi, D., Bisen, P., Saha, S., Loitongbam, B., Singh, S., & Singh, P. K. (2017). Genetic diversity analysis in rice (*Oryza sativa* L.) accessions using SSR markers. International Journal of Agriculture, Environment and Biotechnology, 10(4), 457-467.
17. Shahriar, M. H., Robin, A. H. K., Begum, S. N., & Hoque, A. (2014). Diversity analysis of some selected rice genotypes through SSR-based molecular markers. Journal of the Bangladesh Agricultural University, 12(452-2016-35814), 307-311.
18. Md, A., & Babu, D. R. (2019). Genetic divergence studies in foxtail millet [Setaria italica (L.) Beauv.]. *Electronic Journal of Plant Breeding*, *10*(1), 18-29.
19. Palaniyappan, S., Arunachalam, P., Banumathy, S., Mini, M. L., & Muthuramu, S. (2020). Genetic divergence and clustering studies in advanced breeding lines of rice (Oryza sativa L.). *Electronic Journal of Plant Breeding*, *11*(02), 499-504.
20. Suvi, W. T., Shimelis, H., Laing, M., Mathew, I., & Shayanowako, A. I. T. (2020). Assessment of the genetic diversity and population structure of rice genotypes using SSR markers. *Acta Agriculturae Scandinavica, Section B—Soil & Plant Science*, *70*(1), 76-86.

Supplementary Table 1. List of identified drought tolerant genotypes used in this study

S. No.	Genotypes	Classification of Genotypes	Source
1	IC458210	Drought tolerant landraces	NBPGR, New Delhi
2	IC465008		NBPGR, New Delhi
3	IC67496		NBPGR, New Delhi
4	IC248033		NBPGR, New Delhi
5	IC115439		NBPGR, New Delhi
6	IC457996		NBPGR, New Delhi
7	IC464685		NBPGR, New Delhi
8	IC208155		NBPGR, New Delhi
9	IC458581		NBPGR, New Delhi
10	IC378202		NBPGR, New Delhi
11	IC206282		NBPGR, New Delhi
12	IC115406		NBPGR, New Delhi
13	APO	Drought tolerant checks	TNAU, Coimbatore
14	WAYRERAM		TNAU, Coimbatore
15	IR64	Drought susceptible checks	NBPGR, New Delhi
16	JAYA		NBPGR, New Delhi

Supplementary table 2. List of markers used in this study

S. No.	SSR markers	Sequence (Forward and Reverse) 5'→ 3'	Product Size (bp)	Chromosome number	Number of alleles	PIC value	Annealing Temperature (^{o}C)
1	RM102	AACTTTCCCACCACCACCGCGG	200	**12**	3	0.59	68
		AGCAGCAGCAAGCCAGCAAGCG					
2	RM164	TCTTGCCCGTCACTGCAGATATC	246	**5**	4	0.69	55
		GCAGCCCTAATGCTACAATTCTTC					
3	RM278	GTAGTGAGCCTAACAATAATC	141	**9**	4	0.7	55
		TCAACTCAGCATCTCTGTCC					
4	RM289	GTAGTGAGCCTAACAATAATC	108	**5**	5	0.77	56
		TCAACTCAGCATCTCTGTCC					
5	RM447	CCCTTGTGCTGTCTCCTCTC	111	**8**	6	0.82	55
		ACGGGCTTCTTCTCCTTCTC					
6	RM307	GTACTACCGACCTACCGTTCAC	174	**4**	5	0.61	55
		CTGCTATGCATGAACTGCTC					
7	RM413	GGCGATTCTTGGATGAAGAG	79	**5**	5	0.7	52
		TCCCCACCAATCTTGTCTTC					
8	RM11	TCTCCTCTTCCCCCGATC	140	**7**	4	0.7	53
		ATAGCGGGCGAGGCTTAG					
9	RM189	CGTCTTCCCCAACGCTAAAA	126	**9**	4	0.52	61
		CGCGGGGCTTCGCTTC					
10	RM252	TTCGCTGACGTGATAGGTTG	216	**4**	7	0.77	55
		ATGACTTGATCCCGAGAACG					
11	RM242	GGCCAACGTGTGTATGTCTC	225	**9**	4	0.73	55
		TATATGCCAAGACGGATGGG					
12	RM106	CGTCTTCATCATCGTCGCCCCG	297	**2**	3	0.57	55
		GGCCCATCCCGTCGTGGATCTC					

13	RM218	TGGTCAAACCAAGGTCCTTC GACATACATTCTACCCCCGG	148	**3**	4	0.55	55
14	RM219	CGTCGGATGATGTAAAGCCT CATATCGGCATTCGCCTG	202	**9**	7	0.81	55
15	RM404	CCAATCATTAACCCCTGAGC GCCTTCATGCTTCAGAAGAC	236	**8**	8	0.75	55
16	RM495	AATCCAAGGTGCAGAGATGG CAACGATGACGAACACAACC	159	**1**	4	0.48	55
17	RM302	TCATGTCATCTACCATCACAC ATGGAGAAGATGGAATACTTGC	156	**1**	4	0.66	55
18	RM541	TATAACCGACCTCAGTGCCC CCTTACTCCCATGCCATGAG	158	**6**	6	0.7	55
19	RM246	GAGCTCCATCAGCCATTCAG CTGAGTGCTGCTGCGACT	116	**1**	4	0.72	55
20	RM212	CCACTTTCAGCTACTACCAG CACCCATTTGTCTCTCATTATG	136	**1**	4	0.48	55
21	RM5752	TTGCAATTAATTCGATCTCC GCAGATCGATTCGTTAGTTC	138	**7**	5	0.68	55
22	RM336	CTTACAGAGAAACGGCATCG GCTGGTTTGTTTCAGGTTCG	154	**7**	5	0.77	55
23	RM408	CAACGAGCTAACTTCCGTCC ACTGCTACTTGGGTAGCTGACC	128	**8**	2	0.51	55
24	RM434	GCCTCATCCCTCTAACCCTC CAAGAAAGATCAGTGCGTGG	152	**9**	4	0.66	55
25	RM251	GAATGGCAATGGCGCTAG ATGCGGTTCAAGATTCGATC	147	**3**	6	0.73	55
26	RM263	CCCAGGCTAGCTCATGAACC GCTACGTTTGAGCTACCACG	199	**2**	6	0.73	55

27	RM1	GCGAAAACACAATGCAAAAA GCGTTGGTTGGACCTGAC	113	**1**	4	0.41	55
28	RM518	CTCTTCACTCACTCACCATGG ATCCATCTGGAGCAAGCAAC	171	**4**	3	0.46	55
29	RM250	GGTTCAAACCAAGCTGATCA GATGAAGGCCTTCCACGCAG	153	**2**	4	0.6	55
30	RM547	TAGGTTGGCAGACCTTTTCG GTCAAGATCATCCTCGTAGCG	235	**8**	4	0.63	55
31	RM276	CTCAACGTTGACACCTCGTG TCCTCCATCGAGCAGTATCA	149	**6**	5	0.78	56
32	RM418	TCGCGTATCGTCATGCATAG GAGCACATATGCCACGTACG	283	**7**	2	0.3	55
33	RM127	GTGGGATAGCTGCGTCGCGTCG AGGCCAGGGTGTTGGCATGCTG	223	**4**	4	0.61	55
34	RM243	GATCTGCAGACTGCAGTTGC AGCTGCAACGATGTTGTCC	116	**1**	4	0.7	55
35	RM262	CATTCCGTCTCGGCTCAACT CAGAGCAAGGTGGCTTGC	154	**2**	6	0.75	55
36	RM248	GTAGTGAGCCTAACAATAATC TCAACTCAGCATCTCTGTCC	175	**7**	5	0.57	60
37	RM544	TGTGAGCCTGAGCAATAACG GAAGCGTGTGATATCGCATG	248	**8**	4	0.56	55
38	RM202	CAGATTGGAGATGAAGTCCTCC CCAGCAAGCATGTCAATGTA	189	**11**	3	0.49	55
39	RM204	GTGACTGACTTGGTCATAGGG GCTAGCCATGCTCTCGTACC	169	**6**	6	0.76	55
40	RM272	AATTGGTAGAGAGGGGAGAG ACATGCCATTAGAGTCAGGC	119	**1**	1	0	55

41	RM231	CCAGATTATTTCCTGAGGTC CACTTGCATAGTTCTGCATTG	210	**3**	3	0.63	55
42	RM161	TGCAGATGAGAAGCGGCGCCTC TGTGTCATCAGACGGCGCTCCG	187	**5**	3	0.57	61
43	RM152	GAAACCACCACACCTCACCG CCGTAGACCTTCTTGAAGTAG	151	**8**	2	0.62	61
44	RM223	GAGTGAGCTTGGGCTGAAAC GAAGGCAAGTCTTGGCACTG	165	**8**	4	0.71	55
45	RM346	CGAGAGAGCCCATAACTACG ACAAGACGACGAGGAGGGAC	175	**7**	5	0.73	55
46	RM103	CTTCCAATTCAGGCCGGCTGGC CGCCACAGCTGACCATGCATGC	336	**6**	4	0.64	55
47	RM125	ATCAGCAGCCATGGCAGCGACC AGGGGATCATGTGCCGAAGGCC	127	**7**	4	0.68	55
48	RM510	AACCGGATTAGTTTCTCGCC TGAGGACGACGAGCAGATTC	122	**6**	5	0.73	55
49	RM171	AACGCGAGGACACGTACTTAC ACGAGATACGTACGCCTTTG	328	**10**	4	0.41	55
50	RM44	ACGGGCAATCCGAACAACC TCGGGAAAACCTACCCTACC	99	**8**	4	0.49	55
	Total Number of Alleles				216		

Supplementary Table. 3Dissimilarity matrix for 16 rice genotypes based on simple matching coefficient

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00															
2	0.49	0.00														
3	0.57	0.69	0.00													
4	0.67	0.69	0.41	0.00												
5	0.61	0.63	0.45	0.41	0.00											
6	0.73	0.82	0.51	0.57	0.59	0.00										
7	0.69	0.82	0.59	0.57	0.47	0.57	0.00									
8	0.69	0.63	0.55	0.57	0.49	0.55	0.59	0.00								
9	0.76	0.78	0.61	0.57	0.55	0.63	0.53	0.49	0.00							
10	0.76	0.73	0.61	0.61	0.53	0.51	0.55	0.47	0.49	0.00						
11	0.76	0.80	0.71	0.78	0.65	0.63	0.63	0.51	0.63	0.45	0.00					
12	0.69	0.78	0.69	0.78	0.69	0.67	0.65	0.63	0.67	0.69	0.73	0.00				
13	0.61	0.78	0.76	0.84	0.76	0.73	0.67	0.71	0.82	0.69	0.73	0.65	0.00			
14	0.86	0.73	0.88	0.84	0.78	0.84	0.86	0.69	0.78	0.82	0.82	0.73	0.78	0.00		
15	0.69	0.71	0.69	0.78	0.76	0.73	0.76	0.63	0.78	0.63	0.73	0.57	0.45	0.71	0.00	
16	0.80	0.82	0.86	0.84	0.78	0.88	0.78	0.76	0.71	0.78	0.80	0.76	0.71	0.63	0.73	0.00

1-IC458210, 2-IC465008, 3-IC67496, 4-IC248033, 5-IC115439, 6-IC457996, 7-IC464685, 8-IC208155, 9-IC458581, 10-IC378202, 11-IC206282, 12-IC115406, 13-APO, 14-IR64, 15-WAYRERAM, 16-JAYA.