

## Arranging a Set of Accessions with Whole Genome Sequence Amenable for GWAS Studies in Rice

B. MAHADEVA PRABHU, H. V. RAME GOWDA, R. ABHISHREE AND M. S. SHESHSHAYEE

Department of Crop Physiology, College of Agriculture, UAS, GKVK, Bengaluru - 560 065

e-Mail : mahadevaprabhu.8@gmail.com

### AUTHORS CONTRIBUTION

B. MAHADEVA PRABHU :

Carried out experiment, data analysis and draft the manuscript ;

H. V. RAME GOWDA :

Conceptualization and design ;

R. ABHISHREE :

Provided technical inputs ;

M. S. SHESHSHAYEE :

Conceptualization, design, drafting the manuscript

### Corresponding Author :

B. MAHADEVA PRABHU

Department of Crop Physiology, College of Agriculture, UAS, GKVK, Bengaluru

Received : August 2022

Accepted : March 2023

### ABSTRACT

Water scarcity due to climate change is emerging as a major threat for agriculture in recent years. Growing rice, a water intensive crop in the puddled condition is becoming a major challenge as it consumes around 50 per cent of available fresh water for agriculture. It's critical to improve yield potential of rice under water limited conditions to meet the increasing demands. Though, adapting changes in agronomic practices like growing rice under semi-aerobic conditions could reduce water usage, a concomitant yield loss has occurred. Developing a rice cultivar with minimal water use with increased yield or without reducing the actual yield potential is a great challenge. Addressing this, by evaluating genetic diversity at the DNA/molecular level has become an interested approach with the availability of diversified germplasm resources and genetic information across species generated through high-throughput technologies and innovations. Evaluating the accessions with large phenotypic and genotypic diversity and looking for a desirable qualitative trait through techniques *viz.*, GWAS would lead to a superior cultivar with improved yield potential. Considering Water Use efficiency (WUE) as a qualitative physiological trait and dissecting its subcomponent traits and looking for the variations present in the genes governing the latter would harness the advantages of WUE in crop improvement programs. With this background, a set of 853 diversified accessions with Whole Genome Sequence (WGS) was procured. Further, a subset was assembled based on various criteria like geographical, physiological and molecular diversity that will be amenable for GWAS.

**Keywords :** Water use efficiency, GWAS, Water scarcity, Whole genome sequence

RICE, the most important staple cereal in India is an extremely water intensive crop consuming more than 50 per cent of all the fresh water used in agriculture. With the erratic rainfall leading to decreased water availability growing rice under puddled conditions has emerged as the major constraint to achieve potential productivity. So, it's highly important to focus on water conservation strategy without compromising on the yield, while improving the yield of local varieties (Karthika and Shanker, 2022). A few water saving agronomic practices have been developed, among them, the semi-irrigated aerobic cultivation is known to save up to 40 per cent of irrigation water, with a significant

yield loss of up to 50 per cent (Bouman *et al.*, 2005). The yield reduction could be associated with low spikelet fertility under water limited condition. To overcome this, breeding for rice under water limited conditions is crucial. Genetic enhancement of crops to evolve superior cultivars through improving physiological traits among several drought adaptive traits have been identified and introduction of double haploids technique (Chaitanya and Raju, 2022) are being used for crop improvement (Reynolds and Tuberosa, 2008 and Araus *et al.*, 2008).

Evaluation of genetic diversity is considered as a basic platform for crop improvement. Plant scientists

always evaluated the populations over years and locations to understand genetic diversity present in the species of their interest. Insights on useful trait inheritability, environmental influence, value of different parents for breeding and obtaining desirable offsprings were collectively performed by using replication and sophisticated experiments. But, the introduction of evaluation of genetic diversity at the DNA level has become an interested approach.

Datasets associated with genotyping and phenotyping have been generated to address both basic and applied questions. This interests have been shown in the nature and origin of mutations and functional significance of the genes to obtain qualitative and quantitative traits. Tremendous improvement and innovations in the genomics technology over the last 20 years, through multidimensional approaches collaborating biologists, bioinformaticians and chemists have been driven with the efficient support of the available diversified germplasm resources and genetic information across species through high-throughput genotyping and next generation sequencing (NGS) (Tung *et al.*, 2010 and Elshire *et al.*, 2011).

Whole-genome sequence, emerging as an unifying tool in biology has lead to development of diversified panels and large mapping populations in many crop species to facilitate trait dissection and gene discovery. With the availability of huge genetic information, adapting Genome Wide Association Studies (GWAS) technique would lead to the discovery of QTLs associated with the phenotypic trait of interest.

Among various physiological traits contributing to yield, water use efficiency (WUE) and water use are considered as important traits under water limited conditions. Physiologically WUE is defined as  $\mu\text{mol CO}_2$  fixed per mole of water transpired or is the ratio of biomass to water transpired. Improving crops for WUE most often was not associated with higher yields due to the occurrence of trade-off between photosynthesis and transpiration. However, it is evident that at a given water use, plants with higher WUE would have higher growth rates,

which can be predominantly due to chloroplast carbon assimilatory capacity. Therefore, developing a greater understanding of the various chloroplast and photosynthetic mechanisms governing variability in WUE has high relevance in crop improvement programs.

To harness the advantages of WUE, it is critical to dissect this complex trait into its subcomponents. With this background, a program was conceived to identify the genetic variability in WUE and its sub-component traits using the accurate large-scale phenotyping facility. To accomplish this study, a set of accessions with diversity in geographical origin, genotypic and phenotypic variability was studied with a set of 150 germplasm accessions with whole genome sequence is assembled considering various criteria like population group, days to flowering and allelic variations found in the genes for future phenotyping experiments that will be amenable for GWAS.

## MATERIAL AND METHODS

### Seed Material and Growth Condition

A diversified panel of 853 germplasm accessions with whole genome sequence from 3K Rice genome panel were procured from the IRRI Centre located at ICRISAT, Hyderabad. The selected accessions were grown in the field at ZARS Mandya, during *kharif* 2019 under aerobic condition. Around 30 plants were maintained for each accession and the seeds were dibbled at a depth of 3-4 cm while maintaining 20×10 cm spacing between rows and plants respectively. The recommended agronomic practices were followed. The accessions were grown till the crop reaches its physiological maturity and seeds were collected for the further experiments.

### Selection of a Diversified Subset

The passport data containing various informations on the accessions like, country of origin, sub population group, days to 50 per cent flowering etc., are available at the IRRI website (SNP Seek). These data were used to select a subset of accessions belonging to subpopulation group of Indica. Further, observations on days to 50 per cent flowering recorded at ZARS

Mandya and the data available with passport information were compared to narrow down the accessions with 85 to 95 days to 50 per cent flowering.

### Selection of Genes

Water Use efficiency is contributed by various sub-component traits *viz.*, CO<sub>2</sub> diffusion, carboxy lation, photochemical processes, epidermal patterning, stomatal mechanisms, VPD responses, etc. These subcomponent traits are governed by various genes and transcription factors, so a through literature search was done in selecting a set of genes governing different subcomponent traits of WUE and based on the importance and functions.

### Sequencing Data Analysis

Single nucleotide polymorphism data for the set of selected accessions with reference to Nipponbare are available at IRRI website (SNP Seek) and were downloaded for the experimental purpose. Around 2178 SNP markers were detected for 217 accessions and these SNPs were used to perform population structure using the STRUCTURE HARVESTER software with maximum adhoc measure ( $\Delta K$ ) of five. Further Cluster analysis was performed for 217 accessions based on random 2198 SNP markers by Neighbor-joining method using DARwin.

ICRISAT, Hyderabad. Hence, considering phenotypic and molecular diversity was crucial.

### Selection of a Set of Accessions Based on Subpopulation Group

The 853 accessions obtained from the IRRI center located at ICRISAT, Hyderabad, represented 48 countries Fig. 1, indicating a diversification in their origin and the accessions belong to 12 subpopulation groups (Table 1). Since the goal of our study was to improve crop productivity in Indian conditions, only Indica subpopulation group was selected and the 853 lines were narrowed down to 522 lines.

TABLE 1  
Representing 853 lines from different sub population groups

Subpopulation	No of genotypes
Admixture	27
Aromatic	27
Australian	78
Indica 1 A	41
Indica 1 B	41
Indica 2	176
Indica 3	87
Indicax	178
Japonicax	15
Subtropical	70
Temperate	10
Tropical	103

### RESULTS AND DISCUSSION

The main aim of this study was to select a subset from the 853 lines obtained from the IRRI center located at

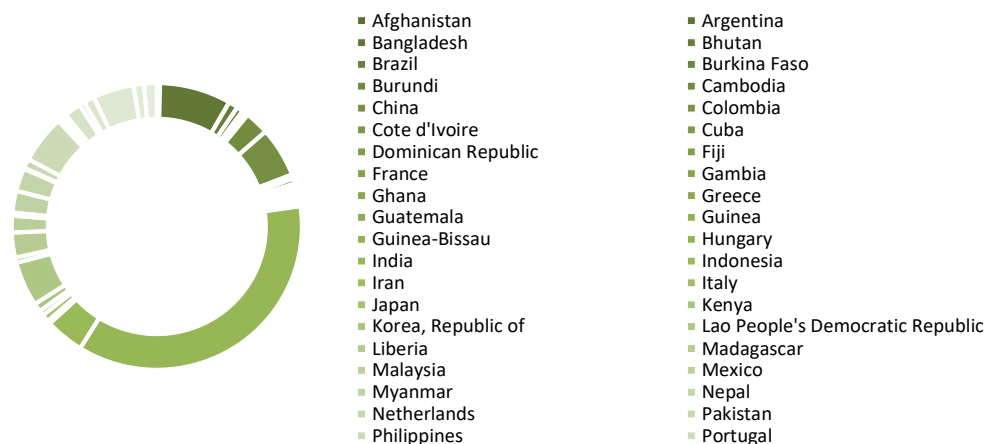


Fig. 1 : Collection of 853 lines from different geographical area



Fig. 2 : Phenotyping of 853 rice accessions at Zonal Agricultural Research Station, V C Farm, Mandya

### Narrowing Down of the Accessions Based on Days to Flowering

Every germplasm will have passport data on various traits including days to flowering. We observed a large variation in days to flowering between 50 and 184 days after sowing from a passport data (<https://snp-seek.irri.org/>). To verify these data, the entire set of 853 accessions was grown in ZARS, VC Farm Mandya during *kharif* 2019 (Fig. 2). Days to Flowering recorded at Mandya did not show any association with passport data while 70 accessions remained unflowered due to photosensitivity (Fig. 3 and Table 2). This indicates the possibility of variations in photosensitivity of germplasm collected

TABLE 2

Comparison of flowering duration between passport data and measured data

Flowering Duration (days)	Passport data	<i>Kharif</i> 2019
51-60	5	3
61-70	47	0
71-80	110	130
81-90	177	125
91-100	142	246
101-110	134	138
111-120	106	78
121-130	68	12
131-140	31	51
141-150	13	-
151-160	11	-
161-170	8	-
Photosensitive	-	70

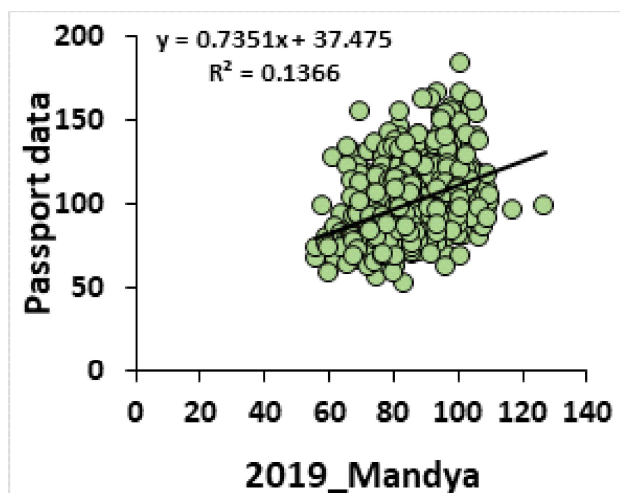


Fig. 3 : Correlation for days to fifty percent flowering taken from passport data and recorded data from VC farm Mandya

from various locations. Flowering time is controlled by many genes, which are expressed or suppressed in close interaction with environmental factors such as day length and temperature (Tsuji *et al.*, 2008). Since the environmental conditions in Phillipines and India are comparatively different we observed a larger variation in days to flowering ( $R^2 = 0.13$ ). Further comparing both the data, around 217 accessions were selected having a range between 85 and 95 days to flowering. Development of early-flowering or photoperiod-insensitive cultivars has been an important objective of rice breeding (Tsuji *et al.*, 2008). Hence the selection was made to meet this objective.

### Dissection of Water use Efficiency in to Subcomponent traits

With the goal of improving crop productivity through WUE, dissecting the subcomponents of

WUE plays a major role. Considering this, around 35 putative genes governing various subcomponents traits of WUE *viz.*, CO<sub>2</sub> diffusion, carboxylation, photochemical processes, epidermal patterning, stomatal mechanisms, VPD responses *etc.* were identified through literature search and genetic diversity analysis has been done to know the variations in sequences of the candidate genes leading to functional differences. Recently (Sheshshayee *et al.*, 2013) provided some experimental evidences, that clearly establish the relevance of WUE as an important physiological trait that can be used for crop improvement. It was further proved that introgression of WUE with other traits such as total water use (roots) and improved acquired tolerance mechanism significantly enhanced growth rate under both well-watered and water limited condition (Raju *et al.*, 2014 and Sheshshayee *et al.*, 2018). Furthermore, after identifying a robust marker associated with WUE and root traits (Raju *et al.*, 2016), a marker assisted back-cross breeding strategy was adopted to introgress WUE and water use characters on to the background of a late variety IR64. The trait introgressed lines on an average resulted in a 23 per cent increased yield under water limited conditions while saving more than 50 per cent of irrigation water (Dharmappa *et al.*, 2019) These clearly emphasize the possibility of improving productivity through an improvement in WUE.

### Structure Analysis

A total of 2591 SNPs present among the 217 accessions were retrieved. After filtering the missing data, a total of 2278 SNPs (Table 3) were selected

TABLE 3  
Number of SNPs on each chromosome used to run cluster analysis

Chromosome	Number of SNPs
CHR 1	251
CHR 2	244
CHR 3	110
CHR 4	403
CHR 5	132
CHR 6	244
CHR 7	158
CHR 8	108
CHR 9	153
CHR 10	113
CHR 11	152
CHR 12	130

for further analysis. Population structure of 217 accessions was analyzed using STRUCTURE with different K values ranging between 1 and 10. STRUCTURE harvester showed the maximum adhoc measure ( $\Delta K$ ) at five (Fig. 4). The optimum K value was fixed at so as to divide the 217 accessions into five sub-populations (Fig. 5).

Neighbor joining analysis using DARwin clustered the accessions also into five clusters, with two distinctly distant clusters (C1 and C2) as illustrated in Fig. 6. A core set of germplasm was constructed using heuristic search available with Power Core and a core collection containing 150 rice accessions was constituted. The selected accessions represented 29 countries of origin (Fig. 6) and five subpopulation

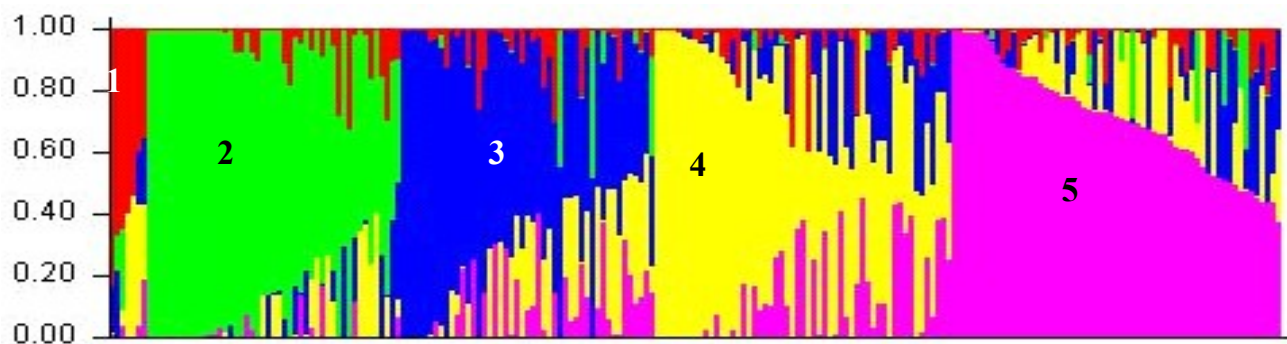


Fig 4. Population structure of 217 accessions

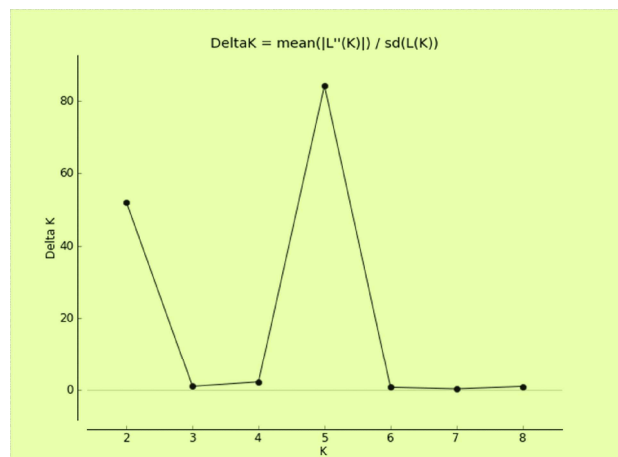


Fig. 5 : Maximum adhoc measure ( $\Delta K$ ) of two observed from STRUCTURE harvester

TABLE 4  
150 lines representing different subpopulation groups

Subpopulation	No of genotypes
Indica1A	11
Indica1B	6
Indica2	55
Indica3	16
Indicax	62

group (Table 4) indicating large geographical diversity maintainance and these accessions will be used for all phenotyping and GWAS activities. Similarly, Nachimuthu *et al.* (2015) assessed the molecular and genetic diversity and relatedness by evaluating the set of 192 diverse rice germplasm lines with 61 genome wide SSR markers leading to identification of 205 alleles revealing two sub groups.

So this indicates that understanding the population structure and assessing the kinship relatedness between superior alleles and traits which is necessary for successful association mapping program.

### Haplotype Analysis

The sequences of the selected genes were downloaded from the IRRI database and aligned to screen for haplotype diversity. Haplotype analysis revealed the existence of large sequence variations in 35 candidate genes (2-15 haplotypes) except Rca and BP-73 which had no polymorphism (Table 5). Diversity analysis and population structure analysis were done for the selected 217 lines using nucleotide sequence polymorphisms prevailing in the 35 genes. The existing SNPs and InDels among the 35 genes were retrieved from the SNPseek database (<https://snpseek.irri.org/>) and used. Haplotype assembly is one

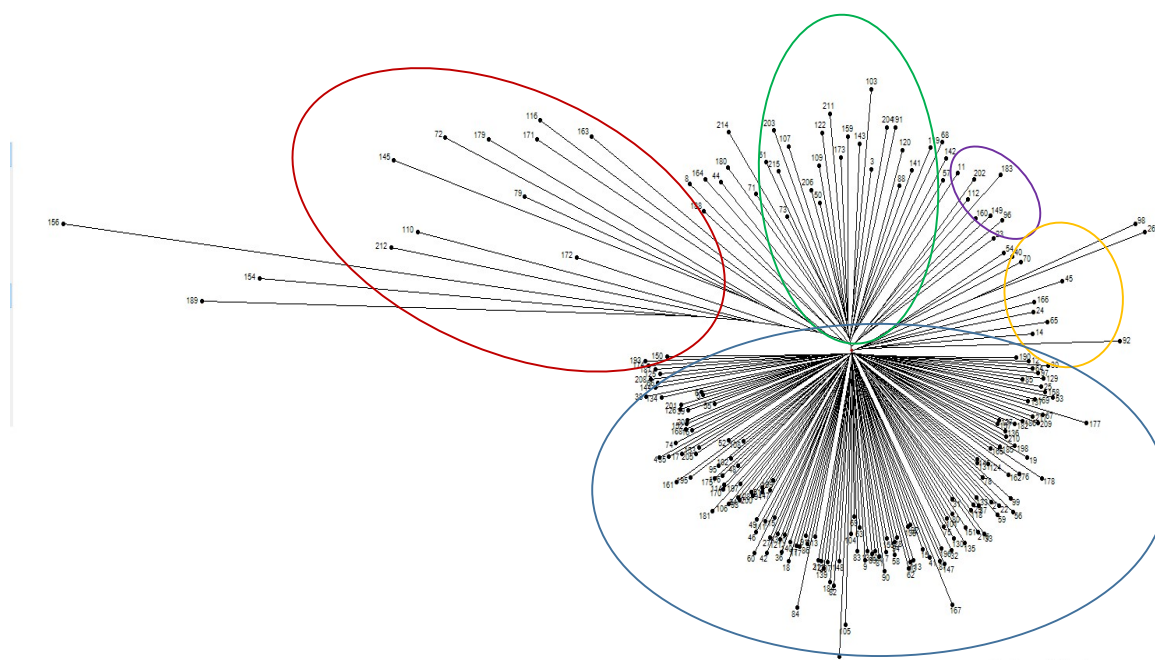


Fig. 6 : Cluster analysis of 217 accessions based on 2278 SNP markers for 35 candidate genes using DARwin

TABLE 5  
Haplotype diversity of selected candidate genes 3K  
panel rice accessions.

Gene	Locus ID	Haplotypes
SDD1	Os03g0143100	12
ERECTA	Os02g0777400	15
SLAC1	Os04g0574700	15
GORK	Os06g0250600	2
OST2	Os03g0689300	4
rbcL	Os12g0277500	3
RbcS	Os12g0274700	15
<b>Rca</b>	Os11g0707000	-
PGK	Os06g0668200	5
GPS	Os04g0615000	15
OsCKX2	Os01g0197700	15
<b>BP-73</b>	Os03g0183100	-
SAP16	Os07g0569700	15
HAP3	Os01g0834400	15
SPS	Os01g0919400	9
SBP	Os01g0916400	15
TMM	Os01g0623000	15
FAMA	Os05g0586300	15
COP1	Os02g0771100	2
DST	Os10g0456800	8
EPFL9	Os01g0914400	12
SAMDC	Os02g0611200	2
P5CS	Os01g0848200	2
NAC	Os01g0191300	3
AP37	Os01g0797600	15
CIPK	Os03g0319400	15
PIP	Os02g0823100	15
SnRK2	Os03g0250000	15
OsCDPK	Os03g0128700	15
SBP	OSNPB_040234600	15
TPI	OSNPB_010147900	2
RPI	OSNPB_030781400	15
FBA	OSNPB_060664200	2
RPE	OSNPB_030169100	15
SPS	OSNPB_020184400	2

of the promising approaches used in breeding program for crop improvement. In recent years, haplo-pheno analysis is extensively used for identification of superior haplotypes. In 3K rice genome panel, haplotypes of 21 genes governing grain yield and quality has been identified (Abbai *et al.*, 2019). Moreover, haplotypes of deep water adaptation in rice and for direct seeded rice (DSR) have been identified (Kuroha *et al.*, 2018 and Chen *et al.*, 2019). Hence in this study more number of haplotypes is a indirect indicator of vast genetic and phenotypic variability in the selected population.

*Acknowledgements* : Authors profusely thank late Prof. M. Udaya Kumar for having initiated the work and guiding the entire team throughout. Thanks to the Department of Science and Technology (DST), Govt. of India for the DST-INSPIRE Fellowship. The research is supported by the National Agriculture Science Fund (NASF-2581). Dr. Raveendran, Director of Research (TNAU) and his team who helped in analysis of the data are heartily acknowledged. The authors wish to thank Dr. Deepak and his team at ZARS, V.C. Farm, Mandya for their support in field experiments. The immense help rendered by Dr. Prathibha, M. D., Dr. Preethi Vijayaraghavareddy and Dr. Priyanka Basavaraddi is highly appreciated by the authors and will be thankful for their contribution. The authors also thank the members of Molecular Marker Lab and Phenomics team, Department of Crop Physiology for providing technical inputs during the experiment.

#### REFERENCES

- ABBAI, R., SINGH, V. K., NACHIMUTHU, V. V., SINHA, P., SELVARAJ, R., VIPPARLA, A. K., SINGH, A. K., SINGH, U. M., VARSHNEY, R. K. AND KUMAR, A., 2019, Haplotype analysis of key genes governing grain yield and quality traits across 3K RG panel reveals scope for the development of tailor made rice with enhanced genetic gains. *J. Plant Biotechnol.*, **17** (8) : 1612 - 1622.
- ARAUS, J. L., SLAFER, G. A., ROYO, C. AND SERRET, M. D., 2008, Breeding for yield potential and stress adaptation in cereals. *Crit. Rev. Plant Sci.*, **27** (6) : 377 - 412.

- BOUMAN, B. A. M., PENG, S., CASTANEDA, A. R. AND VISPERAS, R. M., 2005, Yield and water use of irrigated tropical aerobic rice systems. *Agric. Water Manag.*, **74** (2) : 87 - 105.
- CHEN, K., ZHANG, Q., WANG, C. C., LIU, Z. X., JIANG, Y. J., ZHAI, L. Y., ZHENG, T. Q., XU, J. L. AND LI, Z. K., 2019, Genetic dissection of seedling vigour in a diverse panel from the 3,000 Rice (*Oryza sativa* L.) Genome Project. *Sci. Rep.*, **9** (1) : 1 - 15.
- CHAITANYA, G. AND MOHAN RAJU, B., 2022, Optimization of *in-vitro* and rogenic protocol for development of haploids and double haploids in Indica rice hybrid, KRH 4. *Mysore J. Agric. Sci.*, **56** (1) : 39 - 48.
- DHARMAPPA, P. M., DODDARAJU, P., MALAGONDANAHALLI, M. V., RANGAPPA, R. B., MALLIKARJUNA, N. M., RAJENDRAREDDY, S. H., RAMANJINAPPA, R., MAVINAHALLI, R. P., PRASAD, T. G., UDAYAKUMAR, M. AND SHESHSHAYEE, S. M., 2019, Introgression of root and water use efficiency traits enhances water productivity: an evidence for physiological breeding in rice (*Oryza sativa* L.). *Rice*, **12** (1) : 1 - 14.
- ELSHIRE, R. J., GLAUBITZ, J. C., SUN, Q., POLAND, J. A., KAWAMOTO, K., BUCKLER, E. S. AND MITCHELL, S. E., 2011, A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS one*, **6** (5) : 19379.
- KARTHIKA, V. AND SHANKER, A. G., 2022, Efficient callus induction and regeneration protocol for the rice Indica genotype Sonali. *Mysore J. Agric. Sci.*, **56** (1) : 171 - 179.
- KUROHA, T., NAGAI, K., GAMUYAO, R., WANG, D. R., FURUTA, T., NAKAMORI, M., KITAOKA, T., ADACHI, K., MINAMI, A., MORI, Y. AND MASHIGUCHI, K., 2018, Ethylene-gibberellin signaling underlies adaptation of rice to periodic flooding. *Sci.*, **361** (6398) : 181 - 186.
- NACHIMUTHU, V. V., MUTHURAJAN, R., DURAIALAGURAJA, S., SIVAKAMI, R., PANDIAN, B. A., PONNIAH, G., GUNASEKARAN, K., SWAMINATHAN, M., K. K. S. AND SABARIAPPAN, R., 2015, Analysis of population structure and genetic diversity in rice germplasm using SSR markers: an initiative towards association mapping of agronomic traits in *Oryza sativa*. *Rice*, **8** (1) : 1 - 25.
- RAJU, B. R., NARAYANASWAMY, B. R., MOHANKUMAR, M. V., SUMANTH, K. K., RAJANNA, M. P., MOHANRAJU, B., UDAYAKUMAR, M. AND SHESHSHAYEE, M. S., 2014, Root traits and cellular level tolerance hold the key in maintaining higher spikelet fertility of rice under water limited conditions. *Funct. Plant Biol.*, **41** (9) : 930 - 939.
- RAJU, B. R., MOHANKUMAR, M. V., SUMANTH, K. K., RAJANNA, M. P., UDAYAKUMAR, M., PRASAD, T. G. AND SHESHSHAYEE, M. S., 2016, Discovery of QTLs for water mining and water use efficiency traits in rice under water-limited condition through association mapping. *Mol. Breed.*, **36** (3) : 1 - 16.
- REYNOLDS, M. AND TUBEROSA, R., 2008, Translational research impacting on crop productivity in drought-prone environments. *Curr. Opin. Plant Biol.*, **11** (2) : 171 - 179.
- SHESHSHAYEE, M. S., MOHANKUMAR, M. V., RAJU, B. R., PRATIBHA M. D., RAJANNA, M. P., MOHANRAJU, B. AND UDAYAKUMAR, M., 2013, Enhancing water use efficiency besides effective use of water is a potential strategy in developing rice cultivars suitable for semi-irrigated aerobic cultivation. International Dialogue on perception Prospects of Designer Rice Society for Advancement of Rice Research, Directorate of Rice Research, Hyderabad, India, pp. : 261 - 272.
- SHESHSHAYEE, M. S., VIJAYARAGHAVAREDDY, P., SREEVATHSA, R., RAJENDRAREDDY, S., ARAKESH, S., BHARTI, P., DHARMAPPA, P. AND SOOLANAYAKANAHALLY, R., 2018, Introgression of physiological traits for a comprehensive improvement of drought adaptation in crop plants. *Front. Chem.*, **6** : 92.
- TSUJI, H., TAMAKI, S., KOMIYA, R. AND SHIMAMOTO, K., 2008, Florigen and the photoperiodic control of flowering in rice. *Rice*, **1** (1) : 25 - 35.
- TUNG, C. W., ZHAO, K., WRIGHT, M. H., ALI, M. L., JUNG, J., KIMBALL, J., TYAGI, W., THOMSON, M. J., MCNALLY, K., LEUNG, H. AND KIM, H., 2010, Development of a research platform for dissecting phenotype-genotype associations in rice (*Oryza* spp.). *Rice*, **3** (4) : 205 - 217.