Acta Biologica Szegediensis

Volume 66(1):37-46, 2022
Acta Biologica Szegediensis
http://abs.bibl.u-szeged.hu/index.php/abs

DOI:10.14232/abs.2022.1.37-46

ARTICLE

Rice performance prediction to deficit irrigation using microsatellite alleles and artificial intelligence

Bahareh Ghasemi[1], Hossein Sabouri[2,*], Hossein Hosseini Moghaddam[2], Abbas Biabani[3], Mohamad Javad Sheikhzadeh[3]

[1]Agricultural Biotechnology, Department of Plant Production, Collage of Agriculture Science and Natural Resource, Gonbad Kavous University, Gonbad Kavous, Golestan, Iran
[2]Crop Physiology, or, Department of Plant Production, Collage of Agriculture Science and Natural Resource, Gonbad Kavous University, Gonbad Kavous, Golestan, Iran
[3]Computer Science, Collage of Basic Sciences & Engineering, Gonbad Kavous University, Gonbad Kavous, Golestan, Iran

ABSTRACT Rice germplasm investigated as completely randomized design under flooding and deficit irrigation conditions. The results of the association analysis indicated that RM29, RM63, and RM53 could be used for rice breeding programs to improve yields under deficit irrigation. The highest accuracy of rice performance prediction was 98.36 for the RFA (RFA) for panicle length, flag leaf length, and width, and the number of primary branches, after that, the MLP algorithm had better prediction power than other algorithms. When a genotypes code was considered as a criterion to classify the genotypes under the drought stress at the reproductive stage, the random forest algorithm (RFA) was the best algorithm based on the predictive accuracy (67.93), kappa value (0.514) and root mean square error (0.293). Based on the artificial intelligence methods, the RFA presented the best results to predict the response of genotypes to deficit irrigation using the microsatellite molecular data.

Acta Biol Szeged 66(1):37-46 (2022)

KEY WORDS
association analysis
microsatellite marker
random forest algorithm
rice germplasm

ARTICLE INFORMATION
Submitted
22 November 2021
Accepted
06 June 2022.

***Corresponding author**
E-mail: hossein.sabouri@gonbad.ac.ir

Introduction

Rice is an important agricultural food for more than half of the world's population (Todaka et al. 2012). It plays an important role in the world's food cycle, by supplying the main food for more than 50% of the world's population. Since over 90% of rice is produced and consumed in Asia, it is an important crop in this continent (Wang et al. 2013).

The high genetic diversity among physiological and morphological traits in different populations, their correlation with grain yield, and high heritability can be used as selection criteria along with new molecular methods to select the superior genotypes (Pierre et al. 2010). Today, molecular markers are used as an important tool for crop breeding (Shamsabadi et al. 2021; Shirmohammadli et al. 2018). There are several reports on the use of SSR molecular markers to study the genetic diversity of drought-tolerant rice cultivars (Giasi Oskoei 2014; Sarayloo 2015; Sabouri et al. 2022).

Chuang et al. (2011) studied 32 microsatellite markers and 36 rice cultivars from different countries for the polymorphic analysis, and observed 306 alleles at 32 loci. The number of alleles per locus varied from 3 to 21. The genetic diversity of 106 rice genotypes was analyzed by microsatellite markers. SSR markers showed the highest polymorphic levels of 80% to 100%. The microsatellite detected 181 alleles for 37 SSR primers. The number of alleles per gene locus varied from 2 to 13. The PIC value ranged from 0.45 to 0.81 (Sarma and Rathi 2012). The association analysis of 128 rice genotypes and 11 agronomic traits was studied over two years. The population was analyzed using 125 SSR markers covering the entire genome. In total, 16 markers showed significant associations with different traits, and the researchers suggested that the association analysis was efficient for investigating different rice genotypes in breeding programs (Zhou et al. 2012). In a study on the genetic diversity among different rice genotypes using morphological characteristics of SSR markers, Basmati cultivars were differentiated from non-Basmati cultivars (Das et al. 2012).

Segmentation algorithms and decision tree (Bramer 2007; Kantardzic 2003) were practical classifications that often occur in everyday life and include the division of objects, so that a unique number, called "class", is assigned to each of these objects. The uniqueness means that each object should be assigned exactly to a class. The decision tree is an approach widely used for classification and pre-

Table 1. The number and name of evaluated genotypes

No.	Genotype	No.	Genotype
1	IRRI 133	52	IR13L137
2	IR 09L324	53	CT 18614-4-1-2-3-2
3	IR 11A506	54	IR 64
4	IRRI 153	55	IR12L356
5	BP 11820-5F-KN-10-2	56	HHZ 15-SAL13-Y1
6	HHZ 2-SUB2-DT1-DT1	57	HHZ 26-SAL12-Y1-Y1
7	IR14L248	58	IR12L369
8	IR14L110	59	HHZ 3-SAL6-Y1-Y1
9	IR13L400	60	HHZ 1-DT3-Y1-Y1
10	IRBLKH-K3	61	HHZ 4-DT6-LI2-LI1
11	IR 10A199	62	IR14L121
12	HHZ 4-SAL12-LI1-LI1	63	IRBLT-K59
13	HHZ 22-Y3-DT1-Y1	64	IR 11N121
14	IR13L406	65	HHZ 6-DT1-LI1-LI1
15	HHZ 1-DT7-LI2-LI1	66	IR 11C123
16	IR 10A237	67	IR08L217
17	IR 11A581	68	IR10A121
18	IR 10A314	69	IR14L256
19	HHZ 4-DT3-Y1-Y1	70	IR12L357
20	IR 11A501	71	HHZ 18-Y3-Y1-Y1
21	IR 10F221	72	HHZ 24-DT11-LI1-LI1
22	HHZ 21-SAL13-Y1-Y1	73	IR14L238
23	IR 09N251	74	IR 11N137
24	HHZ 23-DT16-DT1-DT1	75	HHZ 3-SAL13-Y2-DT1
25	IRRI 104	76	IR13L268
26	IR14L101	77	IR 09N127
27	IR12L353	78	IR14L260
28	HHZ 3-SAL4-Y1-Y1	79	IR 09L204
29	SAKHA 105	80	IR06A145
30	HHZ 21-Y4-Y2-Y1	81	IR 10A227
31	IR 11A410	82	IRBLK-KU
32	IR14L240	83	HHZ 4-SAL5-LI1-LI1
33	IR14L247	84	IR 05A272
34	HHZ 15-SAL13-Y3	85	IRRI 103
35	IR10L139	86	IRRI 146
36	B 40	87	IRRI 154
37	IR13L397	88	IRBLZT-IR56
38	IR12L201	89	HHZ 10-DT8-DT1-DT1
39	IR13F589	90	HHZ 14-SAL19-Y1
40	IR13F228	91	HHZ 15-DT7-SAL2
41	IRRI 132	92	IR 04A216
42	IR14L160	93	IRBLZ5-CA[CO]
43	HHZ 10-DT5-LI1-LI1	94	HHZ 16-SAL13-LI1-LI1
44	IRBLSH-S	95	IR 09L324
45	IR12L159	96	IRBLSH-IS
46	B11598C-TB-2-1-B-7	97	HHZ 1-DT4-LI1-LI1
47	IR13L382	98	IRBLKS-CO
48	IR09N516	99	IR08L216
49	HHZ 4-SAL5-Y2-Y1	100	IRBLTA2-IR64
50	IR14L262	101	IRBLTA-ME
51	IR14L235	102	IRBLKM-TS

diction. It is an effective way of generating classes from high-speed datasets. The decision tree works by building a tree of rules. New items can be started by simple steps from each decision node, starting at the root node and ending at the leaf node. The training and testing method is used to construct a class that divides the available dataset into training and test sets. Frate et al. (2003) used artificial neural networks with multilayer perceptron structure with two hidden layers and sigmoid operator function to estimate soil moisture and found that neural networks were suitable tools to estimate the soil moisture. Pachepsky et al. (1996) reported neural network models could estimate soil water storage capacity with better accuracy and lower error by investigating neural network and regression models in 230 soil samples.

The present study aimed to investigate the genetic diversity between rice genotypes and effective markers on rice traits and predict the response of the genotypes to deficit irrigation according to molecular data via the artificial intelligence.

Material and Methods

Phenotypic evaluations

In this section, 102 rice genotypes (Table 1) were studied in 3-kg pots under flooding and deficit irrigation at Gonbad Kavous University Greenhouse in three replications in 2017-2018. The deficit irrigation was applied as intermittent irrigation for 10 days from the maximum tillering stage. The plant height (PH), plant weight (PLW), number of fertile panicles (NFT), number of infertile panicles (NIFT), total panicle weight (PAW), main panicle length (PL), panicle extrusion (PE), flag leaf length (FLL), flag leaf width (FLW), number of primary panicles (NPB), number of filled grains (NFG), number of unfilled grains (NUFG), filled grain weight (FGW), leaf weight (LW), and shoot diameter (SD) were recorded. The genotypes' responses to stress were evaluated after applying the stress, and then genotypes' scores were determined for drought tolerance based on a method by Loresto and Chang (1981) according to Table 2. for each genotype.

The two selection indices including STI (Fernandez 1992) and TOL (Rosielle and Hambline 1981) were calculated based on grain yield under control and water stress conditions according to the following formulae:

$$STI = (FGW_s \times FGW_p)/(FGW_{mp})^2$$

$$TOL = FGW_p - FGW_s$$

Table 2. Scoring system for drought resistance screening at reproductive 12-15 days after start of stress IRRI, 1981.

Decimal score	Heading	Panicle exertion	Panicle size	Spikelet fertility (%)	Grain filling	Leaf rolling
1	No delay	Full	Normal	91-100	Mostly well filled	Slight folding
3	Delayed by less than 1 week	Full	Normal	76-90	Mostly well filled	Half rolling
5	Delayed by more than 1 week	Partial	Slightly reduced	51-75	Mostly half filled	Full to tight
7	Delayed by less than 2 weeks	Half-exerted	Reduced by half	11-50	half-filled to empty	Tight
9	No heading until soil moisture is replenished	Half-exerted	Reduced by half	0-10	Mostly empty	Tight

where FGWs and FGWp are the yield of the ith genotype under stress and normal conditions, respectively. FGWmp is the mean yield of all genotypes in control condition.

Genotypic assessments

The DNA extraction was performed in line with a method by Saghi Maroof et al. (1994). The agarose gel electrophoresis was utilized to determine the concentration and quality of extracted DNA. The PCR reaction was performed for 21 SSR markers (Table 3).

Statistical analysis

The analysis of variance and mean comparison were performed by SAS 9.2 software. Cluster and correlation analysis were conducted between traits via SPSS 23. The regression method and SPSS 23 were used to investigate the relationships of the traits and SSR alleles. To this end, the phenotypic traits were considered as dependent variables; and amplified alleles as independent variables. The response of the genotypes to drought stress based on artificial the intelligence-based methods was predicted through Weka 3.8 software by Hall et al. (2009).

Results and Discussion

Analysis of variance (ANOVA) and mean comparison

The differences between genotypes were significant for all traits under flooding and deficit irrigation conditions. The significance of the traits indicated the genetic diversity among genotypes in terms of studied traits. Different responses of rice genotypes between flooding and deficit irrigation were investigated by a other researchers, including Lanceras et al. (2004). The mean of all traits, except for a NIFT, was higher at the flooding stage than the deficit irrigation stage. In flooding, IRRI133 and B40 genotypes were among the top 10% genotypes in terms of PH, PLW, PAW, NUFG. In the deficit irrigation condition, B40 and BP 11820-5F-KN-10-2 genotypes were among the top 10% genotypes in terms of PH, PLW, NFT, PAW, FGW, and SD. In general, flooding and deficit irrigation conditions of B40 were better than other genotypes.

Relationships of traits

Under flooding condition, the highest correlations belonged to PLW and LW (0.793**), and also PLW and NFT

Table 3. Characteristics of the studied SSR markers

Primer	Chr.	Forward	Reverse	Reference
RM530	2	GCACTGACCACGACTGTTTG	ACCGTAACCCGGATCTATCC	Vikram et al. 2011; Donde et al. 2019
RM127	4	GTGGGATAGCTGCGTCGCGTCG	AGGCCAGGGTGTTGGCATGCTG	Lanceras et al. 2004
RM129	1	TCTCTCCGGAGCCAAGGCGAGG	CGAGCCACGACGCGATGTACCC	Prasad et al. 2016
RM216	10	GCATGGCCGATGGTAAAG	TGTATAAAACCACACGGCCA	Dixit et al. 2014; Vikram et al. 2011
RM231	3	CCAGATTATTTCCTGAGGTC	CACTTGCATAGTTCTGCATTG	Diwan et al. 2013
RM236	2	GCGCTGGTGGAAAATGAG	GGCATCCCTCTTTGATTCCTC	Kumar et al. 2014
RM22	3	GGTTTGGGAGCCCATAATCT	CTGGGCTTCTTTCACTCGTC	Donde et al. 2019; Vikram et al., 2011
RM60	3	AGTCCCATGTTCCACTTCCG	ATGGCTACTGCCTGTACTAC	Donde et al., 2019; Vikram et al. 2011
RM12091	1	CTGCAAATGCACAGGAATCAGG	TCCTCTCGCCTTTCTTTCTCTCC	Vikram et al. 2011
RM263		CCCAGGCTAGCTCATGAACC	GCTACGTTTGAGCTACCACG	Vikram et al. 2011
RM520	3	AGGAGCAAGAAAAGTTCCCC	GCCAATGTGTGACGCAATAG	Venuprasad et al. 2009
RM511	12	CTTCGATCCGGTGACGAC	AACGAAAGCGAAGCTGTCTC	Bernier et al. 2007
RM157	3	CCTCCTCCTCACGAATCCCGCC	GGGCTTCTTCTCCGCCGGCTTC	Prasad et al. 2016
RM1029	1	GATTTCCTGCGAATGAGAGAAGG	GACTTCAGGGACAAGCAGTTCC	kumar et al., 2017
RM304	10	TCAAACCGGCACATATAAGAC	GATAGGGAGCTGAAGGAGATG	Swamy et al. 2017; Kumar et al. 2017

Table 4. Correlations matrix in flooding and deficit irrigation condition (above diameter: correlation at deficit irrigation and below diameter: correlation at flooding condition)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	0.414**	0.205	0.130	0.155	0.312**	-0.058	0.118	0.122	0.129	0.074	0.244*	0.126	0.078	0.063
2	0.265*	1	0.560**	0.039	0.632**	0.214*	-0.104	0.163	0.082	0.195	0.061	0.095	0.167	0.564**	0.091
3	0.133	0.710**	1	-0.136	0.668**	0.113	0.125	-0.029	0.187	0.056	0.446**	-0.233*	0.510**	0.008	-0.155
4	-0.052	-0.004	-0.161	1	0.179	0.171	0.150	0.012	0.166	0.150	0.061	0.181	0.020	0.104	0.097
5	0.148	0.609**	0.507**	0.030	1	0.337**	0.138	0.151	0.246*	0.177	0.271**	-0.087	0.416**	0.226*	-0.016
6	0.242*	0.085	-0.048	-0.186	0.050	1	0.168	0.548**	0.383**	0.483**	0.094	0.290**	0.201	0.054	0.090
7	0.074	-0.123	0.144	0.039	-0.064	-0.104	1	0.017	0.023	-0.147	0.309**	-0.252*	0.323**	-0.183	0.092
8	0.337**	0.157	0.094	-0.092	0.155	0.134	-0.124	1	0.421**	0.563**	-0.020	0.233*	0-.016	0.369**	0.050
9	0.151	0.251*	-0.005	-0.088	0.159	-0.017	-0.277**	0.197	1	0.533**	0.210*	0.233*	0.168	0.049	0.040
10	0.146	0.189	-0.109	-0.155	0.189	0.245*	-0.257*	0.305**	0.270**	1	0.119	0.442**	0.094	0.373**	0.047
11	0.169	0.238*	0.081	-0.084	0.329**	-0.020	-0.221*	0.106	0.306**	0.307**	1	-0.227*	0.664**	-0.086	-0.026
12	0.225*	0.153	0.002	-0.075	0.179	0.087	-0.190	0.182	0.198	0.221*	0.045	1	-0.327**	0.173	0.170
13	0.176	0.106	0.003	-0.142	0.327**	-0.001	-0.099	0.194	0.257*	0.298**	0.572**	-0.104	1	-0.152	0.046
14	0.163	0.793**	0.452**	0.064	0.374**	0.065	-0.313**	0.324**	0.274**	0.282**	0.198	0.161	0.103	1	0.105
15	0.112	0.312**	0.146	-0.059	0.188	-0.031	-0.182	-0.045	0.309**	0.003	0.073	0.100	0.007	0.234*	1

* and * significant in 0.05 and 0.01 level of probability respectively.
1. Plant height, 2. Plant weight, 3. Number of fertile panicles, 4. Number of infertile panicles, 5. Panicle weight, 6. main panicle length, 7. Panicle extrusion, 8. Flag leaf length, 9. Flag leaf width, 10. Number of primary branches, 11. Number of filled grains, 12. Number of unfilled grains, 13. Filled grain weight, 14. Leaf weight, 15.Shoot diameter.

(0.710**). The correlations between traits can be due to the pleiotropic effect of genes, gene interconnections, or epistatic effects. In deficit irrigation conditions, the highest correlations belonged to NFT and PAW (0.668**), and also, NFG with FGW (0.664**). Researchers found that under optimal irrigation conditions, there was a high genetic correlation between biological yield (BIO), harvest index (HI), panicle fertility percentage (PAF), NFT, PH, and grain yield (YID). However, BIO and HI had significant correlation with YID under deficit irrigation conditions (Lanceras et al. 2004). In a study regarding the drought stress conditions, there was a significant correlation (0.75**) between root length (ROL) and stem length (STL) in rice (Lum et al. 2014). In a study on rice under drought stress conditions, the highest positive and significant correlation coefficient (0.67**) were reported between the ROL and STL (Srividhya et al. 2011). The stepwise regression was used to select the traits that were essential in explaination growth variation of the genotypes. In flooding conditions, the FGW was considered as dependent variable and other traits as independent variables. The NFG explained maximum variation of 32.7% (F = 43.696 and regression coefficient of 0.320). When the number of genotypes in deficit irrigation was considered as dependent, and other traits as independent variable, the NFG, PAW and NUFG respectively was explained as the highest percentage (50.01%) of variations (F = 44.653 and regression coefficients of 0.010 and 0.052). In cereals, the varieties which can produce high biomass prior to flowering and increase loading in the stem, are among drought-tolerant varieties (Winkel 1989). Higher total biomass production, resulting in more root biomass, is probably a desirable trait for arid environments (Boogard et al. 1996). The correlations between traits presented in Table 4.

Association analysis

Molecular data were fitted to each of the traits to determine the association between morphological traits and amplified alleles (Table 5 and 6). Most alleles containing information in flooded conditions belonged to the NFG. 7 alleles were associated with this trait, of which 4 alleles had a negative effect and 3 alleles had a positive effect on the number of full grains. The highest number of positive alleles belonged to PAW (RM29a, RM75c, RM53e, RM45e, and RM56g). 6 alleles were attached to this trait, of which 5 alleles had a positive and increasing effect. An allele was negatively associated with pH. This allele can be used to lower the pH in selector-assisted selection programs. For PLW, NFT, PE, FLL, FLW, NPB, NUFG, FGW, LW and SD, 1, 4, 4, 4, 6, 1, 5, 3,5, 2 and 2 alleles containing information were identified, respectively. RM53e was associated with PLW, NPT, PAW, and LW. RM45e alleles had the highest association with the traits by correlating with PH and PAW; and RM63e alleles by correlating with FGW, and NFG.

Table 5. Forward regression between traits as dependent variables and amplified alleles as independent variable (association analysis) at flooding conditions

Traits	Intercept	Allele	B†	STD††	F†††	R^2
PH	77.701	RM45e	-2.754**	0.933	8.718**	0.088
PLW	11.073	RM53e	1.553**	0.553	7.885**	0.081
NFT	2.866	RM25d	0.905**	0.287	7.834**	0.080
		RM53e	0.916**	0.126	8.379**	0.158
		RM25b	-0.678*	0.285	7.875**	0.212
		RM41e	0.222*	0.106	7.245**	0.250
NIFT	0.152	RM45f	0.196**	0.042	15.232**	0.145
		RM37b	0.044**	0.013	13.407**	0.232
		RM46d	0.107*	0.042	11.666**	0.285
PAW	2.547	RM29a	2.626**	0.932	21.554**	0.193
		RM75c	0.695**	0.181	15.087**	0.253
		RM53e	0.552**	0.148	13.070**	0.308
		RM45e	0.679**	0.211	12.646**	0.368
		RM56g	0.196**	0.058	13.438**	0.439
		RM75g	-0.497*	0.206	12.789**	0.474
PE	2.454	RM46d	1.038**	0.226	21.272**	0.191
		RM29a	-2.528**	0.978	13.845**	0.237
		RM65i	1.852*	0.803	11.501**	0.282
		RM65f	-1.725*	0.804	10.129**	0.318
FLL	25.115	RM43b	2.099**	0.450	13.419**	0.130
		RM29d	4.832**	1.295	10.822**	0.196
		RM40b	-2.599**	0.820	9.114**	0.237
		RM40d	1.963*	0.808	8.377**	0.278
		RM43c	-1.330*	0.595	7.933**	0.316
		RM63f	-0.564*	0.270	7.599**	0.349
FLW	7.510	RM29e	-0.780**	0.0268	8.432**	0.086
NPB	10.444	RM25c	-0.519**	0.139	15.915**	0.150
		RM43d	0.692**	0.181	12.528**	0.220
		RM25f	0.275**	0.137	10.219**	0.258
		RM46d	-1.324**	0.355	9.223**	0.298
		RM 46e	1.076**	0.364	9.786**	0.363
NFG	68.344	RM45g	7.686**	1.998	7.363**	0.076
		RM63e	-3.366**	1.044	7.132**	0.138
		RM33j	-4.565**	1.585	7.380**	0.20
		RM45f	-7.173**	2.673	7.406**	0.254
		RM41a	4.569**	1.587	7.464**	0.30
		RM40d	2.923*	1.144	7.571**	0.348
		RM29f	-11.457*	5.752	7.283**	0.378
NUFG	15.645	RM45g	4.435**	1.287	4.840**	0.051
		RM45h	-4.479*	1.802	5.547**	0.111
		RM41e	1.583*	0.784	5.184**	0.150
FGW	1.077	RM75c	0.079**	0.024	6.900**	0.071
		RM56f	0.044**	0.014	6.701**	0.131
		RM65g	-0.028*	0.011	6.327**	0.177
		RM63e	-0.062**	0.022	6.674**	0.235
		RM33g	-0.062*	0.028	6.606**	0.278
LW	1.225	RM53e	0.146*	0.058	5.805*	0.061
		RM25a	-0.060*	0.030	5.029*	0.102
SD	4.015	RM37h	0.029**	0.007	15.693**	0.148
		RM29a	0.382**	0.097	16.926**	0.276

PH: Plant height, Plant weight:PLW, Number of fertile tiller: NFT, Number of infertile tiller: NIFT, Panicle weight: PAW, Panicle length: PL, Panicle exertion: PE, Flag leaf length: FLL, Flag leaf width: FLW, Number of primary branches: NPB, Number of filled grains: NFG, Number of unfilled grains: NUFG, Filled grain weight: FGW, Leaf weight; LW, Shoot diameter: SD

† Regression coefiecient, †† Standard errore of regression coefiecient, and ††† F test for regression analysis

* and * significant in 0.05 and 0.01 level of probability respectively.

Table 6. Forward regression between traits as dependent variables and amplified alleles as independent variable (association analysis) at deficit irrigation

Traits	Intercept	Allele	B[†]	STD[††]	F[†††]	R^2
PLW	4.962	RM63g	2.301**	0.263	13.529**	0.131
		RM63c	-2.171**	0.293	38.222**	0.462
NFT	0.966	RM3g	0.663**	0.079	7.176**	0.074
		RM63b	-0.313**	0.131	28.657**	0.392
		RM63c	-0.380**	0.129	22.314**	0.432
		RM37h	0.057*	0.026	18.633**	0.461
NIFT	1.901	RM43g	0.449**	0.096	9.350**	0.094
		RM45a	0.360**	0.097	10.486**	0.191
		RM25e	0.224**	0.059	11.637**	0.284
		RM29b	1.048*	0.416	10.829**	0.332
		RM43f	0.355*-	0.161	10.023**	0.368
PAW	0.954	RM63g	0.563**	0.086	6.501*	0.067
		RM63c	-0.539**	0.096	20.791**	0.318
		RM41d	0.448*	0.186	16.538**	0.361
PL	15.672	RM29e	1.855*-	0.910	4.153*	0.044
PE	1.424	RM43a	0.679**	0.182	4.551*	0.048
		RM43e	-0.800**	0.258	7.309**	0.141
FLL	19.965	RM40b	-1.087**	0.346	7.881**	0.081
		RM43b	1.217**	0.464	7.567**	0.145
		RM29e	-3.137*	1.396	6.958**	0.192
FLW	6.110	RM29e	-0.846**	0.270	7.486**	0.077
		RM68h	0.827**	0.270	5.943**	0.118
		RM68g	-0.709**	0.270	6.533	0.182
NPB	8.102	RM29e	-2.079**	0.453	10.684**	0.106
		RM41c	0.510**	0.150	10.472**	0.190
		RM25c	-0.597**	0.158	9.104**	0.237
		RM25f	0.466**	0.157	9.224**	0.298
		RM43d	0.431**	0.207	8.527	0.331
NUFG	52.268	RM45a	10.441**	2.749	6.216**	0.065
		RM75a	4.047**	1.480	6.377**	0.125
		RM45h	-9.425**	3.860	6.473**	0.181
		RM41b	18.658**	8.740	6.190	0.222
LW	0.739	RM41c	0.087*	0.041	4.415*	0.047
SD	3.969	RM37h	-0.010*	0.005	4.718*	0.050

PH: Plant height, Plant weight:PLW, Number of fertile tiller: NFT, Number of infertile tiller: NIFT, Panicle weight: PAW, Panicle length: PL, Panicle exertion: PE, Flag leaf length: FLL, Flag leaf width: FLW, Number of primary branches: NPB, Number of filled grains: NFG, Number of unfilled grains: NUFG, Filled grain weight: FGW, Leaf weight; LW, Shoot diameter: SD

† Regression coefiecient, †† Standard errore of Regression coefiecient, and ††† F test for regression analysis

* and * significant in 0.05 and 0.01 level of probability respectively

Under deficit irrigation conditions, RM63g, RM63b, RM63c, RM37h had a positive and significant association with number of NFT. Also, RM29e, RM41c, RM25c, RM25f and RM43d linked to NPB. RM29e allele was associated with PL, FFF, FLW, NPB; RM63g allele was associated with PLW, and PAW. RM37h allele was associated with the SD under both flooding and deficit irrigation. The results of the analysis indicated that SSR loci of RM29, RM63, and RM53 had the highest explaination for variations in traits and could be useful in the subsequent studies of breeding programs.

The association analyses of 128 rice genotypes and 11 agronomic traits were investigated over two years. The population was analyzed by 125 microsatellite markers covering the entire genome. In total, 16 markers showed significant associations with different traits. Also, the researchers suggested the efficiency of association analysis was useful in investigating different rice genotypes in

Table 7. Classification accuracy for classification based on TOL and STI system

	Classification system	**TOL**							**STI**						
	Classification algorithm	RT	RF	SMO	IBK	G48	NB	MLP	RT	RF	SMO	IBK	G48	NB	MLP
Traits	PH	19.96	97.82	97.28	95.10	95.65	97.82	97.82	96.73	97.82	97.82	95.96	96.19	97.82	97.82
	PLW	82.60	90.21	87.5	80.43	79.89	85.86	86.41	80.97	85.86	85.86	75.54	74.45	80.43	84.23
	NFT	95.65	97.28	97.82	94.56	94.02	97.28	97.28	83.15	86.5	83.69	70.65	71.19	77.71	82.06
	NIFT	91.84	95.10	92.93	90.21	91.84	93.47	91.84	76.63	79.34	74.45	67.39	68.47	68.47	75.54
	PAW	83.69	89.13	85.86	78.26	75	85.32	86.41	70.65	79.89	71.73	59.23	71.19	66.84	75
	PL	84.23	91.30	85.32	81.52	79.89	85.32	86.41	96.19	98.36	98.36	96.73	96.73	98.36	98.36
	PE	73.91	80.97	75.54	55.43	73.91	75	72.82	80.43	80.97	81.52	66.84	73.91	73.91	80.97
	FLL	83.15	89.67	83.69	78.80	77.71	83.15	84.78	96.73	98.36	98.36	96.73	97.82	98.36	98.36
	FLW	83.15	88.04	84.23	73.91	82.06	79.89	85.86	97.28	98.36	98.36	97.73	97.82	98.36	98.36
	NPB	82.60	88.58	84.23	79.89	81.52	80.43	82.06	97.28	98.36	98.36	96.73	97.82	98.36	98.36
	NFG	93.47	97.28	97.28	94.56	94.56	97.28	97.28	83.15	85.32	83.15	67.93	74.45	79.89	80.97
	NUFG	91.84	94.02	91.30	88.04	88.04	89.67	90.76	82.60	89.13	85.86	78.26	78.26	85.86	83.69
	FGW	77.71	86.95	80.43	77.28	73.91	78.80	82.06	74.45	77.17	73.91	61.95	71.19	71.73	75.54
	LW	76.08	72.28	70.10	54.34	70.65	66.84	70.65	77.71	85.86	78.80	73.91	75	77.17	79.89
	SD	81.52	89.67	83.69	78.26	76.08	84.23	84.23	92.39	93.47	91.84	86.95	87.5	92.39	92.93

RT: Random Tree; RF: Random Forest; SMO: Sequential Minimal Optimization; KNS: K-Nearest Neighbors; J48; NB: Nave Bayes; MP: Multiplayer Perceptron. PH: Plant height, Plant weight:PLW, Number of fertile tiller: NFT, Number of infertile tiller: NIFT, Panicle weight: PAW, Panicle length: PL, Panicle exertion: PE, Flag leaf length: FLL, Flag leaf width: FLW, Number of primary branches: NPB, Number of filled grains: NFG, Number of unfilled grains: NUFG, Filled grain weight: FGW, Leaf weight; LW, Shoot diameter: SD

breeding programs (Zhou et al. 2012). Rabiei et al. (2013) studied the genetic diversity and used 8 rice species from 46 SSR markers and 245 RFLP markers. 20 out of 46 SSR markers, showed polymorphism. Number of alleles per locus ranged from 2 alleles (RM215, RM133, RM433) to 6 alleles (RM271).

Predicting the response of genotypes using molecular data

The 10-fold cross validation was used to validate the classification algorithms. Dataset was divided to k equal parts. Each time one part was used for validation and other parts were utilized as a training set. To determine the types of genotype responses to deficit irrigation, first, the phenotypic traits were grouped via TOL and STI classification systems. To this end, five percent of lower-value genotypes were conventionally classified as susceptible, and five percent of higher-value genotypes as tolerant, and other genotypes as a semi-susceptible group. The classification algorithms were trained by the software by a zero-one matrix from the genetic data and grouping results based on the TOL and STI system. The training algorithm (based on genetic data and without using the phenotypic data) will predict genotype reaction (susceptible, semi-susceptible and resistant) based on molecular data.

Prediction accuracy, *Cohen's kappa* and RMSE (Fig. 1, 2 and 3) were used to evaluate and select the best algorithm (Smeeton 1985). The *Cohen's kappa* coefficient is a statistical criterion of agreement within the evaluator between two measurements to classify similar cases. Therefore, the higher the kappa coefficient of an algorithm, the higher the accuracy of the algorithm. RMSE is a criterion to measure the difference between values (sample and population value) predicted by a model or an estimator and the observed values. Thus the lower error, the higher the accuracy.

At the reproductive stage and based on the prediction

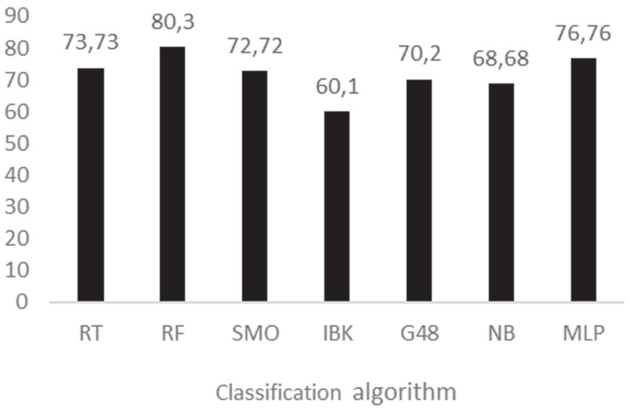

Figure 1. Accuracy of categorization for scoring system. RT: Random Tree; RF: Random Forest; SMO: Sequential Minimal Optimization; KNS: K-Nearest Neighbors; J48; NB: Nave Bayes; MP: Multiplayer Perceptron

Table 8. Kappa statistics values for classification based on TOL and STI system

	Classification system	TOL							STI						
	Classification algorithm	RT	RF	SMO	IBK	G48	NB	MLP	RT	RF	SMO	IBK	G48	NB	MLP
Traits	PH	0.483	0.658	0.604	-0.008	0.000	0.658	0.658	0.557	0.658	0.658	0	0.350	0.658	0.658
	PLW	0.442	0.619	0.593	0.000	0.298	0.490	0.548	0.515	0.583	0.642	0.312	0.281	0.511	0.593
	NFT	0.536	0.656	0.741	0.000	-0.009	0.656	0.656	0.538	0.589	0.543	0.199	0.012	0.423	0.156
	NIFT	0.518	0.645	0.546	0.000	0.320	0.569	0.505	0.513	0.550	0.468	0.230	0.342	0.327	0.490
	PAW	0.522	0.613	0.581	0.054	0.054	0.561	0.567	0.410	0.574	0.432	0.165	0.402	0.329	0.495
	PL	0.438	0.625	0.500	-0.020	-0.047	0.474	0.513	0.518	0.660	0.660	0.000	0.386	0.660	0.660
	PE	0.436	0.559	0.484	0.118	0.403	0.465	0.425	0.595	0.586	0.622	0.233	0.470	0.449	0.601
	FLL	0.485	0.616	0.526	0.101	0.242	0.453	0.549	0.487	0.660	0.660	0.000	0.493	0.660	0.660
	FLW	0.511	0.609	0.565	0.196	0.475	0.440	0.600	0.535	0.660	0.660	0.000	0.493	0.660	0.660
	NPB	0.508	0.589	0.532	0.137	0.418	0.409	0.468	0.534	0.660	0.660	0.000	0.493	0.660	0.660
	NFG	0.425	0.656	0.656	0.000	0.000	0.656	0.656	0.583	0.580	0.581	0.219	0.255	0.522	0.514
	NUFG	0.595	0.642	0.567	0.000	0.262	0.464	0.520	0.480	0.606	0.562	0.165	0.188	0.533	0.515
	FGW	0.444	0.578	0.472	0.019	0.067	0.399	0.492	0.494	0.539	0.480	0.258	0.423	0.439	0.511
	LW	0.529	0.454	0.411	0.097	0.419	0.347	0.422	0.503	0.638	0.521	0.293	0.408	0.482	0.526
	SD	0.441	0.616	0.507	0.009	0.075	0.477	0.509	0.657	0.643	0.617	0.000	0.467	0.619	0.647

RT: Random Tree; RF: Random Forest; SMO: Sequential Minimal Optimization; KNS: K-Nearest Neighbors; J48; NB: Nave Bayes; MP: Multiplayer Perceptron. PH: Plant height, Plant weight:PLW, Number of fertile tiller: NFT, Number of infertile tiller: NIFT, Panicle weight: PAW, Panicle length: PL, Panicle exertion: PE, Flag leaf length: FLL, Flag leaf width: FLW, Number of primary branches: NPB, Number of filled grains: NFG, Number of unfilled grains: NUFG, Filled grain weight: FGW, Leaf weight; LW, Shoot diameter: SD

Table 9. Mean square root values of the error for classification based on TOL and STI system

	Classification system	TOL							STI						
	Classification algorithm	RT	RF	SMO	IBK	G48	NB	MLP	RT	RF	SMO	IBK	G48	NB	MLP
Traits	PH	0.147	0.125	0.287	0.171	0.167	0.120	0.120	0.154	0.128	0.283	0.166	0.157	0.121	0.121
	PLW	0.324	0.25	0.321	0.319	0.327	0.295	0.279	0.352	0.273	0.326	0.356	0.374	0.341	0.294
	NFT	0.149	0.139	0.283	0.175	0.195	0.134	0.128	0.317	0.278	0.333	0.362	0.386	0.357	0.339
	NIFT	0.221	0.191	0.301	0.254	0.220	0.201	0.222	0.368	0.319	0.363	0.394	0.391	0.421	0.389
	PAW	0.316	0.258	0.326	0.336	0.372	0.287	0.268	0.408	0.315	0.371	0.400	0.397	0.430	0.389
	PL	0.316	0.245	0.328	0.322	0.344	0.297	0.288	0.147	0.110	0.280	0.150	0.146	0.104	0.101
	PE	0.395	0.318	0.36	0.422	0.369	0.382	0.393	0.347	0.302	0.341	0.378	0.372	0.377	0.317
	FLL	0.321	0.260	0.333	0.330	0.343	0.312	0.290	0.135	0.109	0.280	0.141	0.119	0.104	0.100
	FLW	0.325	0.271	0.333	0.358	0.310	0.342	0.293	0.122	0.110	0.280	0.148	0.119	0.104	0.104
	NPB	0.332	0.255	0.332	0.323	0.320	0.337	0.310	0.116	0.110	0.280	0.152	0.119	0.104	0.103
	NFG	0.177	0.147	0.285	0.195	0.186	0.134	0.137	0.322	0.281	0.335	0.364	0.367	0.343	0.331
	NUFG	0.216	0.206	0.309	0.268	0.273	0.252	0.235	0.317	0.265	0.332	0.339	0.351	0.286	0.316
	FGW	0.363	0.278	0.344	0.359	0.377	0.353	0.326	0.399	0.318	0.365	0.401	0.388	0.404	0.388
	LW	0.378	0.328	0.376	0.431	0.407	0.434	0.413	0.373	0.286	0.349	0.352	0.355	0.360	0.337
	SD	0.324	0.262	0.333	0.336	0.360	0.294	0.304	0.226	0.195	0.314	0.254	0.270	0.219	0.198

RT: Random Tree; RF: Random Forest; SMO: Sequential Minimal Optimization; KNS: K-Nearest Neighbors; J48; NB: Nave Bayes; MP: Multiplayer Perceptron. PH: Plant height, Plant weight:PLW, Number of fertile tiller: NFT, Number of infertile tiller: NIFT, Panicle weight: PAW, Panicle length: PL, Panicle exertion: PE, Flag leaf length: FLL, Flag leaf width: FLW, Number of primary branches: NPB, Number of filled grains: NFG, Number of unfilled grains: NUFG, Filled grain weight: FGW, Leaf weight; LW, Shoot diameter: SD

accuracy, the Kappa and RMSE in traits, namely total plant weight, number of fertile panicles, number of infertile panicles, length of main panicles, number of filled grains, and number of unfilled grains, could be better predicted by the RFA in the TOL classification system. In other traits, the RFA in the STI classification system

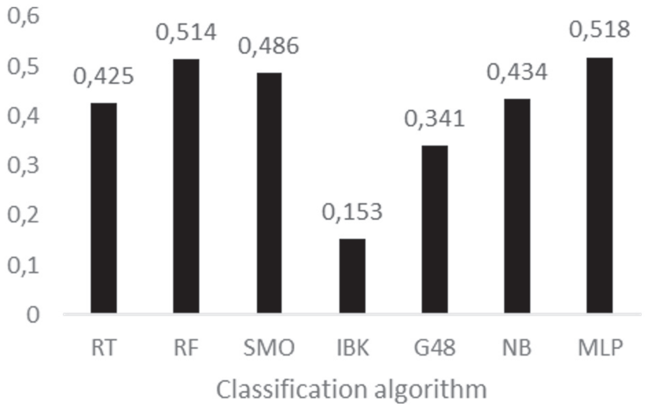

Figure 2. Mean kappa statistics for scoring system. RT: Random Tree; RF: Random Forest; SMO: Sequential Minimal Optimization; KNS: K-Nearest Neighbors; J48; NB: Nave Bayes; MP: Multiplayer Perceptron

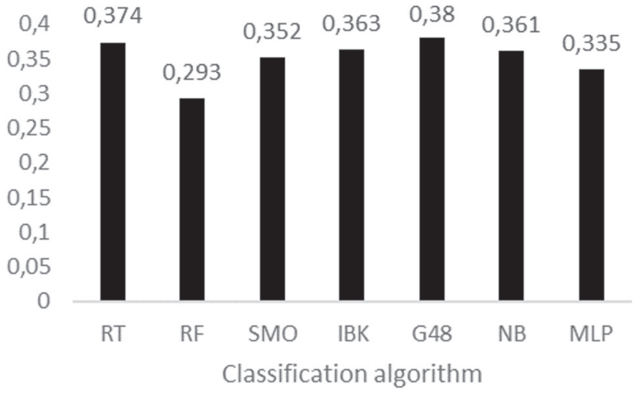

Figure 3. Mean square root error statistics for scoring system. RT: Random Tree; RF: Random Forest; SMO: Sequential Minimal Optimization; KNS: K-Nearest Neighbors; J48; NB: Nave Bayes; MP: Multiplayer Perceptron

made a better prediction. The highest accuracy was 98.36 for main panicle length, flag leaf length, flag leaf width, and number of primary branches. After the RFA, the MLP algorithm had better prediction power than other algorithms in both systems. The SMO algorithm was put in the next rank in terms of the prediction accuracy.

When a genotype code was considered as a criterion for the classification of genotypes under drought stress at the reproductive stage, the RFA with a prediction accuracy of 67.93, kappa value of 0.514, and RMSE of 0.293 was the best algorithm followed by the MLP algorithm as the second rank (Table 7, 8 and 9). All the algorithms made the prediction in one second; however, the MLP algorithm took about 50 times more time to make the prediction.

Conclusion

According to the results, if we predict the response of genotypes based on zero-one data of the markers, we can classify them based on the introduced algorithms in susceptible, tolerant and semi-susceptible groups. According to the results of the artificial intelligence, the maximum accuracy of 98.36 for the RFA was obtained in traits, namely main panicle length, flag leaf length, flag leaf width, and number of primary branches. After the RFA, the MLP algorithm had better predictive power in both systems than other algorithms.

Acknowledgments

The authors acknowledge the Research Deputy of Gonbad Kavous University for providing funding to complete this work

References

Bernier J, Altin GN, Serraj R, Kumar A, Spaner D (2007) Breeding upland rice for drought resistance. J Sci Food Agric 88(6):927-939.

Bramer M (2007) Principle of Data Mining. Springer-Verlag London. pp 341.

Boogard R, Veneklaas E, Lambers H (1996) The association of biomass allocation with growth and water use efficiency of two *T. aestivum* cultivars. Aust J Plant Physiol 23(6):751-761.

Chuang HY, Lur SH, Hwu KK, Chang MC (2011) Authentication of domestic Taiwan rice varieties based on fingerprinting analysis of microsatellite DNA markers. Bot Stud 52: 393-405.

Das B, Sengupta S, Ghosh M, Ghose TK (2012) Assessment of diversity among a set of aromatic rice genotypes from India. Int J Biodivers Conserv 4(5):206-218.

Diwan JM, Channbyregowda V, Salimath PS, Bhat R (2013) Molecular mapping of early vigour related QTLs in rice. Res J Bio 1:24-30.

Dixit S, Singh A, Kumar A (2014) Rice breeding for high grain yield under drought: a strategic solution to a complex problem. Int J Agron 1-16.

Donde R, Kumar J, Gouda G, Kumar MG, Mukherjee M, Baksh SY, Mahadani P, Sahoo K, Behera L, Kumar SD (2019) Assessment of genetic diversity of drought tolerant and susceptible rice genotypes using microsatellite markers. Rice Sci 26(4):239-247.

Fernandez GCJ (1992) Effective selection criteria for assessing plant stress tolerance. In Kuo CC, Ed., Proceedings of an International Symposium on Adaptation of Food Crops to Temperature and Water Stress. AVRDC Publication, Tainan, 257-270.

Frate FD, Ferrazoli P, Schiavon G (2003) Retrieving soil moisture and agricultural variables by microwaver adiometry using neural network. Remote Sens Environ 84:174-183.

Giasi Oskoei M., Farahbakhsh H, Sabouri H, Mohamadinejad G (2014) Evaluation of rice cultivars in drought and normal conditions based on sensitive and tolerance indices. J Crop Prod 6(4):55-75.

Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA data mining software. an update. ACM SIGKDD Explor Newslett 11(1):10-18.

Kantardzic M (2003) Data Mining: Concepts, Models, Methods, and Algorithms. Wiley-Interscience.

Kumar A, Basu S, Ramegowda V, Pereira A (2017) Mechanisms of drought tolerance in rice. In Sasaki T, Ed., Achieving Sustainable Cultivation of Rice. Vol 1. Burleigh Dodds Science Publishing Ltd.

Kumar A, Dixit S, Ram T, Yadaw RB, Mishra KK, Mandal NP (2014) Breeding highyielding drought-tolerant rice: genetic variations and conventional and molecular approaches. J Exp Bot 65(21):6265-6278.

Lanceras JC, Pantuwan G, Jongdee B, Toojinda T (2004) Quantitative trait loci associated with drought tolerance at reproductive stage in rice. Plant Physiol 135(1):384-399.

Loresto GC, Chang TT (1981) Decimal scoring system for drought reactions and recovery ability in screening nurseries of rice. Int Rice Res Newsl 6(2):9-10.

Lum MS, Hanafi MM, Rafii YM, Akmar ASN (2014) Effect of drought stress on growth proline and antioxidant enzyme activities of upland rice. J Animal and Plant Sci 24(5):1487-1493.

Pachepsky YA, Timlin D, Varallyay G (1996) Artificial neural networks to estimate soil water retention from easily measurable data. Soil Sci 60:727-733.

Pierre CS, Crossa J, Manes Y, Reynolds MP (2010) Gene action of canopy temperature in bread wheat under diverse environments. Theor Appl Genet 120(6):1107-1117.

Rabiei B, Zarbafi S, Allah Gulipur M (2013) Genetic analysis of traits related to appearanceand baking in different rice cultivars. Iran J Crop Sci 44:597-612.

Rosielle AA, Hambline J (1981) Theoretical aspects of selection for yield in stress and non-stress environments. Crop Sci 21:943-946

Sabouri A, Dadras AR, Azari M, Saberi Kouchesfahani A, Taslimi M, Jalalifar R (2022) Screening of rice drought-tolerant lines by introducing a new composite selection index and competitive with multivariate methods. Sci Rep 12:2163.

Saghi Maroof MA, Biyaoshev RM, Yang GP, Zhang Q, Allard RW (1994) Extraordinarily polymorphic microsatellites DNA in barly species diversity, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 91(12):4566-5570.

Sarayloo M, Sabouri H, Dadras A (2015) Assessing genetic diversity of rice genotypes using microsatellite markers and their relationship with morphological characteristics of seedling stage under non- and drought-stress conditions. Cereal Res 5(1):1-15.

Sarma RN, Rathi S (2012) Microsatellite diversity in indigenous glutinous rice landraces of Assam. Indian J Biotech 11:23-29.

Shamsabadi EE, Sabour H, Soughi H, Sajadi SJ, Dadras AR (2021) Using of GGE biplot in determination of genetic structure and heterotic groups in wheat (*Triticum aestivum* L.). Acta Biol Szeged 65(1):17-27.

Shirmohammadli S, Sabouri H, Ahangar L, Ebadi A, Sajjadi S (2018) Genetic diversity and association analysis of rice genotypes for grain physical quality using iPBS, IRAP, and ISSR markers. J Genet Res 4(2):122-129.

Smeeton NC (1985) Early History of the Kappa Statistic. Biometrics 41(3), International Biometric Society, 795.

Srividhya A, Vemireddy LR, Ramanarao PV, Sridhar S, Jayaprada M, Anuradha G, Srilakshmi B, Reddy HK, Hariprasad AS, Siddiq EA (2011) Molecular mapping of QTLs for drought related traits at seedling stage under PEG induced stress conditions inrice. Amer J Plant Sci 2:190-201.

Swamy BPM, Shamsudin NAA, Rahman SNA, Mauleon R, Ratnam W, Teressa MS, Kumar C, Kumar A (2017) Association mapping of yield and yield related traits nnder reproductive stage drought stress in rice (*Oryza sativa* L.). Rice 10:21.

Todaka D, Nakashima K, Shinozaki K, Yamaguchi-Shinozaki K (2012) Toward understanding transcriptional regulatory networks in abiotic stress responses and tolerance in rice. Rice 5(1):6.

Venuprasad R, Dalid CO, Del Valle M, Zhao D, Espiritu M, Sta Cruz MT, Amante M, Kumar A, Atlin GN (2009) Identification and characterization of large-effect quantitative trait loci for grain yield under lowland drought stress in rice using bulk-segregant analysis. Theor Appl Genet 120:177-190.

Vikram P, Swamy MBP, Dixit SH, Ahmed HU, Cruz MTS, Singh AK, Kumar A (2011) qDTY1.1, a major QTL for rice grain yield under reproductive-stage drought stress with a consistent effect in multiple elite genetic backgrounds. BMC Genetics 12:89.

Wang H, Xu X, Zhan X, Zhai R, Wu W, Shen X, Dai G, Cao L, Cheng SH (2013) Identification of qRL7, a major quantitative trait locus associated with rice root length in hydroponic conditions. Breed Sci 63(3):267-274.

Winkel A (1989) Breeding for drought tolerance in cereals. Vertr Pflanzenzucht 16:357-368.

Zhou J, You A, Ma Z, Zhu L, He G (2012) Association analysis of important agronomic traits in japonica rice germplasm. African J Biotech 11(12):2957-2970.