The addition of submergence-tolerant Sub1 gene into high yielding MR219 rice variety and analysis of its BC$_2$F$_3$ population in terms of yield and yield contributing characters to select advance lines as a variety

Fahim Ahmed$^a$, Mohd. Y. Rafi $^{a,b}$, Mohd. Razi Ismail$^{a,b}$, Abdul Shukor Juraimi$^a$, Harun Abdul Rahim$^c$, Md. Abdul Latif$^d$, Mohammad Mahmudul Hasan$^{a,e}$ and Fatah Abro Tanweer$^{a,f}$

$^a$Department of Crop Science, Faculty of Agriculture, Universiti Putra Malaysia, Serdang, Malaysia; $^b$Institute of Tropical Agriculture, Universiti Putra Malaysia, Serdang, Malaysia; $^c$Bioscience and Agrotechnology Division, Malaysian Nuclear Agency, Kajang, Malaysia; $^d$Bangladesh Rice Research Institute (BRRI), Dhaka, Bangladesh; $^e$Bangladesh Institute of Nuclear Agriculture (BINA), Mymensingh, Bangladesh; $^f$Department of Plant Breeding and Genetics, Faculty of Crop Production, Sindh Agriculture University Tandojam, Sindh, Pakistan

**ABSTRACT**

A cross was made between MR219 (high yielding but submergence intolerant) and Swarna-Sub1 (submergence tolerant) to produce submergence-tolerant rice variety using the marker-assisted backcrossing (MABC) method to protect the farmers of low-lying land from flash floods during rain. Knowledge of yield and yield contributing factors plays a vital role in the selection process of a variety. This experim ent was designed to determine the genetic diversity among recently produced different lines of BC$_2$F$_3$ population and also to compare all the lines with MR219 to find the best one. Agronomical, yield and yield contributing data were taken, while genotypic and phenotypic coefficients, variance components and heritability were estimated. Introgression of the target gene, Sub1, was done using tightly linked marker, and also background recovery was measured using simple sequence repeat (SSR) markers in different generations. The observed recurrent parent genome (RPG) recovery of BC$_2$F$_2$ generation was 95.37%, which indicates high-level similarity between the recurrent parent (MR219) and the resulting lines. Thirty newly developed lines of BC$_2$F$_3$ population, resulting backcross of MR219 and Swarna-Sub1, were planted with four replications following randomised complete block design (RCBD). Newly developed lines were grouped into four clusters based on traits with UPGMA dendrogram and cluster analysis to select the 10 best plants. This study will help the future researchers to select the best plants of a breeding programme after introgression of a gene considering phenotype performances to develop new varieties.

**Introduction**

Flash floods can result in yield losses of rice up to 100%, depending upon different factors of submergence-prone ecosystems. An economic loss up to 1 billion US dollar has been estimated in South and South-East Asia due to loss of yield caused by flash floods.[1–3] In a survey of rice breeders’ concerns in south and south-east Asia,[4], 51% of respondents rated flash flooding as one of the three most important abiotic constraints on rice yields (the others being drought and salinity). Traditional varieties adapted to these submergence-prone environments are, however, low yielding due to their poor tillering ability, long droopy leaves, susceptibility to lodging and poor grain quality. Improved varieties that can combine high yield attributes with submergence tolerance are needed. Most rice cultivars cannot survive if the plants are completely submerged for more than 7 days.[5] The existing modern varieties are not well adapted to submergence.

Submergence tolerance is controlled by a single gene Sub1 that accounts almost 70% of phenotypic variation which is mapped[6–8] on a FR13A cultivar, and can be divided into three major sub-genes, Sub1A, Sub1B and Sub1C. Submergence tolerance was conferred by the overexpression of Sub1A and down-regulation of Sub1C in a Japonica cultivar to monitor the effect, and was found to be tolerant.[9]

Marker-assisted backcrossing (MABC) is the most appropriate method for the incorporation of a desired quantitative trait loci (QTL) into the genetic background of a mega variety.[10,11] The MABC approach represents a clear advantage over conventional breeding, because...
this strategy results in the development of the ideal genotype in a short period of time, which cannot be done through conventional breeding.[12]

It is necessary to determine the morphological characters of newly developed genotype to get a preliminary assessment of the whole population, which could determine its future use.[13] Variation among genotypes must be determined and grain physical or morphological properties should be considered as a vital factor for measuring it.[14] To determine the genetic diversity and variation among genotypes, researchers used their morphological and physiological characteristics as a general approach. [13] The overall success of a breeding programme can be determined by the yield and yield contributing characters of newly developed genotypes. Thirty newly developed homozygous submergence-tolerant lines were studied to determine their yield and yield contributing characters. The present study was designed to determine the genetic variations among newly developed lines by quantitative traits and also to estimate a broad sense of heritability and genetic advance of the studied lines.

Materials and methods

Plant materials

Malaysian rice variety MR219, which is intolerant to submergence, was crossed with a submergence tolerant variety, Swarna-Sub1, to produce a new variety which is tolerant and also high yielding. Swarna-Sub1 was used as a donor for the gene responsible for the tolerance Sub1. Newly developed 30 homozygous lines of UPM3-BC2F2-34 were tested for yield and yield contributing characters. The whole experiment was conducted in normal field (non-submerged) condition. Submergence tolerance test was conducted on BC2F2 population.

DNA extraction

Fresh leaves from 2 weeks old seedlings were collected and DNA was extracted using the modified cetyltrimethylammonium bromide (CTAB) method.[15] Quantification of DNA was done by running electrophoresis at a voltage of 80 V for 30 min on 1% agarose gel and visualization was done on Molecular Imager® Gel Doc™ XR+ System (BIO RAD, CA, USA). DNA samples were diluted with Tris-EDTA (TE) buffer to a final concentration of 50 ng and kept at −20 °C for further use.

Molecular marker analysis

Tightly linked markers, specific for Sub1 gene, were tested to find polymorphism between the parents. A total of 385 simple sequence repeat (SSR) markers without the foreground markers were used between two parents which is widely spread across the whole genome for background selection.

Foreground selection

Primer RM8300 [10] was used in foreground as tightly linked marker, because of its clear codominant nature and capability of producing easily scorables bands. Marker distance from Sub1, nature of codominance and capability of producing scorables bands were the main criteria for choosing markers.

Background selection

Molecular markers not linked to Sub1, polymorphic between two parents and covered all chromosomes including Chromosome 9 (carrier of Sub1) were used for the selection of background to determine the recovery of recipient genome. Evenly spaced SSR markers were selected for every chromosome. The microsatellite markers that revealed fixed (homozygous) alleles at non-target loci at one generation were not screened at the next backcrossing generation. In case of BC1F1 and BC2F1 generation, 66 polymorphic markers were used for background survey. Additional 19 SSR markers were tested in BC2F2 generation for estimating the amount of recipient genome (Supplementary material).

Experimental design and management practices

The experiment was designed in randomised complete block design (RCBD) with four replications. MR219 was used as control variety and one seedling per hill was planted. Complete N–P–K–S fertilizer was applied. Urea @ 100 kg/ha was applied both at 20 and 40 days after transplant (DAT). Moluscicide was used just after transplanting as an essential protective measure of the newly transplanted seedling against the attack of golden snails. Special care was taken for the protection from rodent and insects. Other cultural management practices were carried out as required.

Data collection

Phenotypic data were taken considering 11 attributes of the agronomical properties of grain quality and yield: plant height, number of tillers per hill, panicles per hill, panicle length, days to maturity, percentage of filled grain, grain length, grain width, days of 50% flowering, yield per hill and 1000 seed weight. The method of measurement used in this experiment is shown in Table 1.
Statistical analysis

The analysis of variance was conducted with the help of SAS 9.2. The mean differences were judged with Tukey’s test at a 5% level of significance.

Genetic parameters were estimated to determine genetic variation among genotypes and to assess genetic and environmental effects on various traits. Genetic parameters were calculated using necessary equations.[16–18]

Brief descriptions of those parameters are given below:

**Genotypic variance**:

\[ s^2_g = \frac{(\text{MSG} - \text{MSE})}{r} \]

where \( \text{MSG} \) is the mean square of genotypes, \( \text{MSE} \) is the mean square of the error, and \( r \) is the number of replications.

**Phenotypic variance**:

\[ s^2_p = s^2_g + s^2_e \]

where \( s^2_g \) is the genotypic variance and \( s^2_e \) is the mean square of the error.

**Error variance**:

\[ s^2_e = \text{MSE} \]

where \( \text{MSE} \) is the mean square of the error.

**Phenotypic and genotypic coefficient of variation**:

The estimates of phenotypic (PCV) and genotypic coefficient of variation (GCV) were obtained as explained by Choudhary and Singh [19] as follows:

\[ \text{PCV} = \frac{\sqrt{s^2_p}}{-X} \times 100 \]

\[ \text{GCV} = \frac{\sqrt{s^2_g}}{-X} \times 100 \]

where \( s^2_p \) is the phenotypic variance, \( s^2_g \) is the genotypic variance and \(-X\) is the mean of trait.

**Heritability estimate**:

\[ h^2_p = \frac{s^2_g}{s^2_p} \]

where \( s^2_g \) is the genotypic variance and \( s^2_p \) is the phenotypic variance.

**Expected genetic advance (GA)**:

\[ \text{GA} = K \times \sqrt{\frac{s^2_p}{-X}} \times h^2_p \]

where \( K \) is a constant, \( \sqrt{s^2_p/-X} \) is the phenotypic standard deviation, \( h^2_p \) is the heritability and \(-X\) is the mean of traits.

**Cluster and principal component analysis**

In this study, the analysis was done using NTSYS-PC software (Version 2.1) based on Jaccard’s similarity coefficient. The UPGMA algorithm and SAHN clustering were applied to determine genetic relationships among the genotypes. Principle component analysis (PCA) analysis of 30 lines and MR219 were calculated by EIGEN and PROJ modules of NTSYS-pc and Minitab software (version 15).

**Results and discussion**

**Foreground selection**

Out of 340 plants, 156 were found to be heterozygous (H) for marker RM8300 linked with Sub1 and 184 of them were found to be fixed with the recipient allele (A), where no plant was shown as fixed with donor allele (B) in BC1F1 generation. In BC2F1 generation, out of 368 plants, 172 plants were shown as heterozygous (H) and remaining 196 plants were shown as recipient allele (A) (Figure 1). Out of 287 plants, maximum numbers of plants, i.e. 137, were found to be heterozygous (H) (scored as H), 81 were found to be fixed with the recipient allele (scored as A) and 69 were fixed with the resistant/donor alleles and
scored as ‘B.’ The results fitted the expected 1:2:1 ratio for this generation ($\chi^2 = 1.5, p > 0.05$) in BC$_2$F$_2$.

Positive BC$_2$F$_2$ lines were tested with a CAPS marker named GnS2, specific to Sub1A locus and showed a banding pattern similar to resistant allele Sub1 of the donor Swarna-Sub1. Hence, it was possible to confirm that all the selected plants contained Sub1.

**Background selection**

Initially, in BC$_1$F$_1$ generation, background selection was carried out with 66 primers. The percentage of recovery of background markers which are heterozygous for the parents ranges from 65.55% to 77.8% in the selected plants. Double recombinant was found in BC$_2$F$_1$ generation from the population of P12-2. Double recombinants were confirmed from the best plants of BC$_1$F$_1$ population and they were also tested with remaining background primers. Plants 2–49 had the recipient allele recovery for all the chromosomes except chromosomes 3 and 4, without considering foreground markers. The maximum percentage of recurrent parental recovery was 95.4% on plant numbers 2–49. A graphical map was constructed using all polymorphic markers shown in Figure 2.

**Figure 1.** Gel picture of BC$_2$F$_1$ generation using foreground marker RM8300 as closely linked Sub1 marker.

**Figure 2.** Graphical genotype of BC$_2$F$_2$ population (P12-2-49) using background markers including Sub1-linked marker. The distances were represented in cM based on published map of Neeraja et al. [10].
Table 2. ANOVA table for 11 morphological characters.

| Sources of variation | DF | Plant height (cm) | Days to maturity (days) | No. of tiller/hill | Panicle length (cm) | Panicle/hill | % of filled grain | Grain length | Grain width | Yield/hill (g) | Days to 50% flowering | 1000 filled grain wt. (g) |
|----------------------|----|-------------------|------------------------|-------------------|--------------------|--------------|-------------------|--------------|-------------|---------------|------------------------|-----------------------------|
|                      |    |                   |                        |                   |                    |              |                   |              |             |               |                        |                             |
| Replications         | 3  | 8.78              | 5.01                   | 48.10             | 1.43               | 86.65        | 12.26             |              |             |               |                        |                             |
| Lines                | 30 | 99.18**           | 60.40**               | 76.1**            | 48.55**           | 67.74**      | 128.43**          | 0.0001**     | 0.00005**   | 40.04         | 22.17                  | 0.006                       |
| Error                | 120| 13.90             | 9.57                  | 12.54             | 8.14               | 11.77        | 17.81             |              |             |               |                        |                             |
| Corrected total      | 123|                   |                        |                   |                    |              |                   |              |             |               |                        |                             |
| **,** NS significant at p ≤ 0.01 and p ≤ 0.05, respectively, and NS denotes non-significant.

Recorded data analysis

The ANOVA table for morphological traits (Table 2) showed significant differences among lines for all of the traits under study, except grain length, grain width and 1000 filled grain weights. Means are presented in Table 3; because of the non-significant interaction effects, only the main effect has been explained.

Plant height

Significant difference was shown in plant height, where most of the lines were different, but not significantly different to the value of MR219. L1 was found to be the longest (108.28 cm) and L23 and L3 (107.62 cm) were the closest to L1. L22 was the shortest plant with a height of 89.1 cm (Table 3).

Days to maturity

A significant difference was found among maturity periods of the lines, but not significantly different to the value of MR219. L25 showed the longer period (124.4 days) to maturity, which is followed by L12 (119.6 days) and L13 (119.4 days) and the shortest period of maturity was shown by L28 (112 days). The rest of the lines showed different values than those of MR219 (Table 3).

Number of tillers per hill

In case of number of tillers per hill, significant difference was observed between the lines, where L28 showed significant difference from the values of MR219. L7 (26.4) showed the highest value followed by L24 (25.8). A minimum number of tillers were observed in L28 (15). Remaining lines had different values, which were statistically similar to that of MR219 (Table 3).

Table 3. Yield and yield contributing traits of 30 selected advance lines derived from UPM3-BC2F2-34.

| Lines     | Plant height (cm) | Days to maturity | Tiller no./hill | Panicle length (cm) | Panicle/hill | Filled grain % | GrL | GrW | Yield/hill (g) | 50% FL | 1000 gw (g) |
|-----------|-------------------|------------------|----------------|--------------------|--------------|----------------|-----|-----|---------------|--------|-------------|
| MR219     |                   |                  |                |                    |              |                |     |     |               |        |             |
| L1        |                   |                  |                |                    |              |                |     |     |               |        |             |
| L2        |                   |                  |                |                    |              |                |     |     |               |        |             |
| L3        |                   |                  |                |                    |              |                |     |     |               |        |             |
| L4        |                   |                  |                |                    |              |                |     |     |               |        |             |
| L5        |                   |                  |                |                    |              |                |     |     |               |        |             |
| L6        |                   |                  |                |                    |              |                |     |     |               |        |             |
| L7        |                   |                  |                |                    |              |                |     |     |               |        |             |
| L8        |                   |                  |                |                    |              |                |     |     |               |        |             |
| L9        |                   |                  |                |                    |              |                |     |     |               |        |             |
| L10       |                   |                  |                |                    |              |                |     |     |               |        |             |
| L11       |                   |                  |                |                    |              |                |     |     |               |        |             |
| L12       |                   |                  |                |                    |              |                |     |     |               |        |             |
| L13       |                   |                  |                |                    |              |                |     |     |               |        |             |
| L14       |                   |                  |                |                    |              |                |     |     |               |        |             |
| L15       |                   |                  |                |                    |              |                |     |     |               |        |             |
| L16       |                   |                  |                |                    |              |                |     |     |               |        |             |
| L17       |                   |                  |                |                    |              |                |     |     |               |        |             |
| L18       |                   |                  |                |                    |              |                |     |     |               |        |             |
| L19       |                   |                  |                |                    |              |                |     |     |               |        |             |
| L20       |                   |                  |                |                    |              |                |     |     |               |        |             |
| L21       |                   |                  |                |                    |              |                |     |     |               |        |             |
| L22       |                   |                  |                |                    |              |                |     |     |               |        |             |
| L23       |                   |                  |                |                    |              |                |     |     |               |        |             |
| L24       |                   |                  |                |                    |              |                |     |     |               |        |             |
| L25       |                   |                  |                |                    |              |                |     |     |               |        |             |
| L26       |                   |                  |                |                    |              |                |     |     |               |        |             |
| L27       |                   |                  |                |                    |              |                |     |     |               |        |             |
| L28       |                   |                  |                |                    |              |                |     |     |               |        |             |
| L29       |                   |                  |                |                    |              |                |     |     |               |        |             |
| L30       |                   |                  |                |                    |              |                |     |     |               |        |             |

Means with different lower case letters are significantly different at p < 0.05 with Tukey test. GrL, grain length; GrW, grain weight; FL, 50% flowering; gw, grain weight.
Panicle length
Panicle length showed significant difference among the individuals, but not significantly different from the value of MR219. L23 (29.79 cm) was the tallest line, which is close to the value of L30 (29.26 cm). On the other hand, L11 (18.47 cm) showed the smallest value (Table 3).

Panicles per hill
L28 (20.8) showed a significantly different value than MR219. L7 (35.2) showed the highest value, which is followed by L4 (33.8) and L14 (33). The lowest value was observed in L28. Remaining lines had different values, but most of them were not significantly different with the recipient parent MR219 (Table 3).

Percentage of filled grains
Filled grains percentage is an important criterion for the selection of a variety. Although there were significant differences between the values, most of them were found to be statistically similar to the value of MR219. There were significant differences observed in the population as L29 (78.6%) was the highest value followed by L20 (76.2%) and L2 (74.8%). The lowest value was observed in L28. Remaining lines had different values, but most of them were not significantly different with the recipient parent MR219 (Table 3).

Grain length
There were no statistical differences between the values observed in lines with the value of MR219, which indicates that this character is controlled by genetic factors and it can be said that the genetic properties of the newly developed lines were similar to MR219 (Table 3).

Grain width
Statistical differences between the values were not observed in line with the value of MR219. This character is controlled genetically and it can also be concluded that the genetic properties of the newly developed lines were similar to MR219 (Table 3).

Grain yield per hill
L28 (59.02 g) showed the lowest value, which is significantly different from the value of MR219. Even though the remaining new lines showed different values but they were not significantly different. Most of the values observed were found to be statistically similar to the value of MR219. L7 (72.78 g) had the highest value which is close to the value of L16 (68.56 g).

50% flowering
Significance difference was observed between the lines but the values of most of the lines were statistically similar to the value of MR219. L21 and L27 (95.8 days) both had the maximum value. The minimum value was observed in L6 (76.2 days).

1000 seed weight (g)
There was no significant difference observed between the lines.

In this study, significant variations of traits among the lines and also with MR219 have been observed. Ullah et al. [20] noticed significant difference among the traits, while studying 10 Boro rice variety of Bangladesh, considering 10 morphological traits. Highly significant variation of traits among 40 rice accessions were also observed in another study [21]. In a similar kind of study, researchers observed highly significant variation between the morphological traits of rice [22,23]. Abar-shahr et al. [24] report highly significant variation on 30 genotypes of rice and the same kind of result was observed by Chandra et al. [25] while studying 19 and 14 quantitative traits of 57 accessions of upland rice.

Tiller number is an important feature for the selection of lines as decrease of tiller number will eventually result in decrease on total yield. The highest tiller number was recorded with L7 (26.40) and the lowest with L28 (15) also suggest this findings. In case of yield/hill, the highest value was recorded on L7 (72.78) and the lowest value was on L28 (59.02). The highest and lowest values were observed because of difference in tiller number per hill and also rate of filled grain percentage. A similar kind of research was conducted by the previous researchers with similar conclusions [26,27].

Phenotypic coefficient of variation (PCV), genetic coefficient of variation (GCV) and the estimation of genotypic heritability
Genotypic variation, phenotypic variation and heritability estimation were calculated and are shown in Table 4. Different types of variation were observed among different quantitative traits. In case of PCV, the highest value was shown by number of tillers/hill (22.20%) which is close to the value of panicle/hill (17.56%) and panicle length (12.84%). The lowest value was shown by days to maturity (2.87%). GCV showed variation in results in this case: number of tillers/hill (19.45%) is the highest value followed by panicle/hill (14.84%) and panicle length (11.05%). The lowest value was shown by days to maturity (2.45%).

The extent of variability for all traits under study needs to be measured for the development of a crop through a breeding programme as all the traits showed highly significant variation between all lines. Whether a trait is going to transfer from parent to offspring depends on its heritability, which plays a vital role in the
selection process.[28] A plant breeder designed a breeding programme based on the observation found on GCV, PCV, GA (genetic advance) and heritability.

It was observed that some traits, i.e. plant height, days to maturity, number of panicles per hill, total number of tillers per hill, number of filled grains per hill, yield per hill and days to 50% flowering showed higher PCV than GCV, while variability of traits were studied to the range of PCV and GCV. It suggests the influence of environment on the expression of such traits. Iftekharuddaula et al. [29] reported higher PCV value than GCV for panicle/square meter and yield/plant. Kavitha and Reddy [30] suggested higher degree of PCV than GCV for the traits, because of a higher magnitude of interaction of genotypes within the environment.

Some traits, such as plant height, days to maturity, number of panicles per hill, total number of tillers per hill, number of filled grains per hill, yield per hill and days to 50% flowering showed higher PCV than GCV, which is similar to the study of Pandey and Anurag [31] and Habib et al. [32], where they observed higher PCV and GCV values in case of number of filled grains per panicle and yield per hill. Akhtar et al. [33] and Zahid et al. [34] observed high GCV and PCV values in the case of tiller per hill. In aromatic rice genotype, high PCV and GCV were also recorded.[35] High values of PCV and GCV were also noticed with some traits like number of unfilled grain per panicle and number of spikelet.[36]

In the case of heritability, several traits, i.e. plant height, days to maturity, tiller numbers per hill, panicles per hill, percentage of filled grains and days to 50% maturity showed more than 70% heritability (Table 4). Yield per hill showed 82%. High heritability indicates the greater chance of genetic transfer of the traits of interest to the next generation. It can be concluded that these traits inherited easily in the studied population as they contain high heritability and those traits were little affected by environmental factors.

Iftekharuddaula et al. [29] reported high heritability for days to maturity, number of filled grains per panicle and 1000 grain weight, as in this study, we observed high heritability in the case of yield per hill, total number of tillers per hill, panicles per hill, panicle length, number of filled grains per panicle, and days to maturity. Akhtar et al. [33] and Habib et al. [32] also reported similar kind of results for days to maturity and the number of filled grains per panicle. Higher heritability value in the case of yield per hill was also reported by Pandey and Anurag [31]. Ghosh and Sharma [36] reported high heritability for total number of spikelet per panicle, yield per hill, days to 50% flowering, flag leaf length, and 100 grain weight. Plant height was noticed as a moderately heritable trait.

The distinction between the mean genotypic values of the selected population with the original population, from which those were selected, was measured by genetic advance. In this study, number of panicles per hill, number of filled grains, total number of tillers per hill, number of and yield per hill showed high genetic advance. Pandey and Anurag [31] also reported high genetic advance for the total number of spikelet.

PCV and GCV, together with heritability, provide clues of the expected genetic advance from the selection.[16] High heritability estimates based on phenotypic performance have proven to be useful in superior genotypes selection. However, genetic advance, together with estimation of heritability, would be more effective in predicting the effect for the selection of the most superior individual.[18] In this study, yield per hill, total number of spikelet per panicle, and number of filled grains per panicle had high heritability along with high genetic advance. Chaudhury and Das [37] reported similar findings for number of filled grains per panicle and yield per hill.

Heritability and genetic advance showed high values in the number of filled grain per panicle, number of spikelet per panicle and yield per hill, which indicates the prevalence of other genes for the expression of the traits. In segregating generation, selection through these characters would be effective.[38] Some traits, i.e. number of filled grain per panicle, number of spikelet per panicle and yield per hill, showed higher GCV, PCV, GA (%) and heritability, which indicates very low environmental influence over those traits and are transmissible through breeding programmes, which in turn indicates that selecting on the basis of those characters would be effective.

**Table 4. Estimation of genetic variables of 11 morphological characteristics of selected lines derived from UPM3-BC2F2-34.**

| Traits                        | Mean  | MSG   | MSE   | $\sigma_h^2$ | $\sigma_p^2$ | PCV   | GCV   | $h_g^2$ | GA    |
|-------------------------------|-------|-------|-------|-------------|-------------|-------|-------|--------|-------|
| Plant height (cm)             | 96.81 | 97.18 | 7.18  | 22.50       | 29.68       | 5.63  | 4.90  | 0.76   | 8.79  |
| Days to maturity (days)       | 116.1 | 35.41 | 3.02  | 8.10        | 11.11       | 2.87  | 2.45  | 0.73   | 4.31  |
| Number of tiller/hill         | 20.9  | 71.10 | 5.01  | 16.52       | 21.53       | 22.20 | 19.45 | 0.77   | 35.10 |
| Panicle length (cm)           | 23.3  | 28.84 | 2.32  | 6.63        | 8.95        | 12.84 | 11.05 | 0.74   | 19.59 |
| Panicle/hill                  | 28.4  | 78.14 | 7.11  | 17.76       | 24.87       | 17.56 | 14.84 | 0.71   | 25.83 |
| Percentage of filled grain    | 68.8  | 163.12| 15.11 | 37.00       | 52.11       | 10.49 | 8.84  | 0.71   | 15.35 |
| Yield/hill (g)                | 64.53 | 158.89| 8.32  | 37.64       | 45.96       | 10.51 | 9.51  | 0.82   | 17.72 |
| Days to 50% flowering         | 87.4  | 117.21| 12.01 | 26.30       | 38.31       | 7.08  | 5.87  | 0.69   | 10.02 |

MSG, mean square of genotypes; MSE, mean square of the error; $\sigma_p^2$, genotypic variance; $\sigma_h^2$, phenotypic variance; $h_g^2$, heritability estimate.
Cluster analysis

Euclidean distance was calculated between the selected homozygous individuals and also the recurrent parent MR219 from the standardized tabulated data. Using these values, a standard UPGMA dendrogram was constructed (Figure 3). Thirty lines with MR219 were grouped into four clusters based on 11 traits at 1.59 dissimilarity coefficients. The cut-off point was 1.48 and it was set for the convenience of discussion.

In the cluster analysis, it can be observed that maximum lines were grouped with the recurrent parent MR219 and as a result of higher dissimilarities, the rest of the lines were organized into different groups (Table 4).

Based on morphological traits, all lines and MR219 were grouped into four clusters at a dissimilarity coefficient of 1.59 (approx.), which indicates the level of diversity among the traits. This result shows the effectiveness of quantitative traits on grouping of rice genotypes. Ahmadikhah et al. [39] grouped 58 rice varieties into four cluster groups considering 18 morphological traits where the genetic distance was 0.75. The four groups are namely A, B, C, and D, where group A consisted of one member, groups B consisted of 14, group C consisted of 20 and D consisted of 23 members.

Veasey et al. [40] classified 20 rice populations into 10 clusters considering 20 traits, where the largest group was the last (seven members) and the smallest were groups number one, two and seven with only one member.

Principal component analysis

PCA helps to verify the cluster analysis results about different groups found in the dendrogram. Three-dimensional (3D) plots found from PCA are often useful for the proper understanding of the studied population. In case of multivariate analysis, like cluster analysis and PCA, one result suggesting the other makes the result more accurate. In PCA analysis, similar genotypes were found together in the same group (Figure 4). PCA analysis suggested the same type of result observed in cluster analysis (Table 5) and also in 3D plots shown in Figure 4. 65.3% variation in all of the measured traits showed a high correlation in the PCA study (Table 6). Eigen vectors analysis suggested that 21.8%, 18.5%, 13.7% and 11.3% of the total variation of the studied traits could be explained by the first four principal components.

According to the results, PC1 depicted a total variation of 21.8% (Table 6). Seven traits among 11 morphological traits under study were negatively and remaining
four were positively correlated to PC1. Plant height (0.085), 50% flowering (0.212), filled grain (0.242) and grain width (0.108) were positively correlated traits, and days to maturity (−0.351), panicle length (−0.203), tiller number/hill (−0.509), panicles per hill (−0.484), grain length (−0.111), yield/hill (−0.510) and 1000 seed weight (−0.030) were negatively associated. With the decrease of positively correlated traits and the increase of negatively correlated traits, first PC increases. This indicated that, if one trait among the negatively correlated traits increases, it will influence the others, so that they would increase as well. PC1 could be used as a standard for the measurement of the quality of the positively correlated traits. Tiller number (−0.509) was the most strongly correlated traits of PC1.

18.5% of the total variation was explained by PC2. Seven traits were found to be correlated negatively and remaining traits were correlated positively among the 11 traits in the study. The negatively correlated traits were days to maturity (−0.187), plant height (−0.183), panicles per hill (−0.027), filled grain (−0.511), grain length (−0.019), grain width (−0.12) and 1000 seed weight (−0.0527). Positively correlated traits were tiller number per hill (0.113), panicle length (0.105), yield per hill (0.187) and days of 50% flowering (0.561). In PC2, we can conclude that, its value decreases with the increase in the values of traits which are positively correlated, but increases with the negatively correlated traits. From the correlation value, it can be concluded that filled grain% (−0.51) was the most strongly correlated trait of PC2 (Table 6).

PC3 explained 13.7% of the variation and four traits were found positively correlated. Those include panicle length (0.111), grain length (0.588), filled grain% (0.199) and days to maturity (0.129). Negatively correlated traits were plant height (−0.273), tiller number per hill (−0.1844), grain width (−0.607), panicle number per hill (−0.260), yield per hill (−0.034), 1000 seed weight (−0.159), days of 50% flowering (−0.115) and filled grain (−0.018). With the increase of negatively correlated traits and the decrease of positively correlated traits, PC3 increases. Panicle length was most strongly correlated to this PC3 with the value of 0.11 (Table 6).

11.3% of the total variation was expressed by the fourth PC which consisted of six negatively and five positively correlated traits. Yield per hill (0.099), panicles per hill (0.159), grain length (0.102), grain width (0.127) and 1000 seed weight (0.20) were the positively correlated traits.

Table 5. Advance line with MR219 based on 11 yield contributing factors according to cluster.

| Group   | Lines                      |
|---------|----------------------------|
| Cluster I | MR219, L2, L4, L5, L6, L7, L8, L9, L10, L12, L13, L15, L14, L16, L18, L20, L22, L25, L24, L26, L29 |
| Cluster II | L11, L17, L19, L21, L27     |
| Cluster III | L23, L30                   |
| Cluster IV | L1, L3, L28                |

Table 6. Eigen vectors and Eigen values of first four components of PCA.

| Variables                  | PC1       | PC2       | PC3       | PC4       |
|---------------------------|-----------|-----------|-----------|-----------|
| Eigenvalue                | 2.3940    | 2.0339    | 1.5077    | 1.2432    |
| Proportion (%)            | 21.8      | 18.5      | 13.7      | 11.3      |
| Cumulative (%)            | 21.8      | 40.3      | 54.0      | 65.3      |
| Days to maturity          | −0.35418  | −0.18731  | 0.129027  | −0.42328  |
| Plant height              | 0.085631  | −0.18307  | −0.27305  | 0.70621   |
| Tiller number             | −0.50993  | 0.113602  | −0.18442  | 0.00253   |
| Number of panicles/hill   | −0.4873   | 0.02714   | −0.26059  | 0.159828  |
| Panicle length            | −0.20306  | 0.105512  | 0.111631  | −0.41548  |
| Filled grain (%)          | 0.024222  | −0.51147  | 0.199435  | −0.10226  |
| Days of 50% flowering     | 0.212306  | 0.561237  | −0.11529  | −0.18992  |
| Grain length              | −0.11194  | 0.01957   | 0.588174  | 0.102998  |
| Grain width               | 0.108164  | −0.12696  | −0.60736  | 0.127083  |
| 1000 grain weight         | −0.03061  | −0.52748  | −0.15908  | 0.201727  |
| Yield/hill (g)            | −0.51007  | 0.187889  | −0.03471  | 0.099738  |

Figure 4. Three-dimensional plots of PCA indicating the 30 advance lines with MR219 based on 11 morphological characters.
traits. Days to maturity (−0.423), plant height (−0.706), tiller number per hill (−0.002), panicle length (−0.415), filled grain (−0.102) and days of 50% flowering (−0.189) were negatively correlated traits. In PC4, the number of panicle per hill was the most strongly correlated (0.159) trait (Table 6).

PCA was performed to clarify the morphological differences in the results of the few Eigen vectors, which can explain overall diversity among the genotypes. 65.3% of the total variation was explained by the first four principal components, 21.8% of the variation was explained by PC1, 18.5% of the variation was explained by PC2, 13.7% of the variations were explained by PC3 and 11.3% of the variation was explained by PC4. The first 10 principal components accounted for 67% of the total variation. Caldo et al. [41] implied a strong correlation among the traits. Lasalita-Zapico et al. [42] reported a total variation of approximately 82.7% among 32 upland rice varieties, where PC1 explained 66.9% of the variation and PC2 explained 15.87% of the variation.

Conclusions

Submergence tolerant-resistant lines were developed by crossing MR219 with a submergence-tolerant donor Swarna-Sub1 through MABC using MR219 as a recurrent. Sub1, the gene responsible for submergence tolerance, was successfully inserted into the genetic background of the recipient parent and checked for confirmation. In case of morphological characters, yield per hill, total number of tillers/hill and the number of filled grains had high heritability and genetic advance. All of the above were considered as general criteria for the selection. It is very useful for the researchers to select a crop considering economically important traits with very little environmental influence. All the studied lines showed similar homozygosity with the standard variety. It was noticed that the values of grain length, grain width, yield per hill, grain weight, filled grains, and panicle length were higher or equal to MR219. The newly developed lines that performed lower than MR219 were few, which indicates a tendency towards MR219. Some better plants were identified based on the measured traits through cluster analysis and could be selected for future research. On the basis of the assessed traits and cluster analysis, the following plants L7, L13, L24, L12, L26, L10, L30, L14, L27 and L9 could potentially be used as parents that produce higher yields, submergence tolerance and good morphological traits.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The authors would like to acknowledge long-term research grant scheme (LRGS), Food Security Project [no. number 5525001], Ministry of Higher Education, Malaysia, for the financial support on rice breeding.

ORCID

Mohd. Y. Rafii http://orcid.org/0000-0003-4763-6367

References

[1] Herdt RW. Research priorities for rice biotechnology. Rice Biotechnol. 1991;6:19−54.
[2] Dey M, Upadhyaya H. Yield loss due to drought, cold and submergence in Asia. Rice research in Asia: progress and priorities. Wallingford: CAB International; 1996. p. 291−303.
[3] Mackill DJ, Coffman WR, Garrity DP. Rainfed lowland rice improvement. Manila: International Rice Research Institute; 1996. 242 p.
[4] Mackill D. Varietal improvement for rainfed lowland rice in south and southeast Asia: results of a survey. Progress in rainfed lowland rice. Los Baños (Philippines): International Rice Research Institute; 1986. p. 115−144.
[5] Adkins S, Shiraishi T, McComb J. Submergence tolerance of rice—a new glasshouse method for the experimental submergence of plants. Physiol Plant. 1990;80(4): 642−646.
[6] Xu K, Mackill DJ. A major locus for submergence tolerance mapped on rice chromosome 9. Mol Breed. 1996;2(3): 219−224.
[7] Nandi S, Subudhi P, Senadhira D, et al. Mapping QTLs for submergence tolerance in rice by AFLP analysis and selective genotyping. Mol Gen Genet. 1997;255(1):1−8.
[8] Xu K, Xu X, Ronald P, et al. A high-resolution linkage map of the vicinity of the rice submergence tolerance locus Sub1. Mol Gen Genet. 2000;263(4):681−689.
[9] Xu K, Xu X, Fukao T, et al. Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. Nature. 2006;442(7103):705−708.
[10] Neeraja C, Maghirang-Rodriguez R, Pamplona A, et al. A marker-assisted backcross approach for developing submergence-tolerant rice cultivars. Theor Appl Genet. 2007;115(6):767−776.
[11] Iftekharuddaula K, Newaz M, Salam M, et al. Rapid and high-precision marker assisted backcrossing to introgress the SUB1 QTL into BR11, the rainfed lowland rice mega variety of Bangladesh. Euphytica. 2011;178(1):83−97.
[12] Mackill D, Septiningsih E, Pamplona A, et al. Marker assisted selection for submergence tolerance in rice. Mol Plant Breed. 2007;5(2):207−208.
[13] Hien NL, Sarhadi WA, Oikawa Y, et al. Genetic diversity of morphological responses and the relationships among Asia aromatic rice (Oryza sativa L.) cultivars. Tropics. 2007;16(4):343−355.
[14] Mathure S, Shaikh A, Renuka N, et al. Characterisation of aromatic rice (Oryza sativa L) germplasm and correlation between their agronomic and quality traits. Euphytica. 2011;179(2):237−246.
[15] Doyle JJ. Isolation of plant DNA from fresh tissue. Focus. 1990;12:13–15.

[16] Burton GW. Quantitative inheritance in grasses. Proceedings of 6th International Grassland Congress; State College (PA); 1952. p. 277–283.

[17] Burton GW, Devane E. Estimating heritability in tall fescue (Festuca arundinacea) from replicated clonal material. Agron J. 1953;45(10):478–481.

[18] Johnson HW, Robinson H, Comstock R. Genotypic and phenotypic correlations in soybeans and their implications in selection1. Agron J. 1955;47(10):477–483.

[19] Choudhary M, Singh R. A comprehensive study of genic variation in natural populations of Drosophila melanogaster. III. Variations in genetic structure and their causes between Drosophila melanogaster and its sibling species Drosophila simulans. Genetics (USA). 1987;117:697–710.

[20] Ullah M, Bashar M, Bhuiyan M, et al. Interrelationship and cause-effect analysis among morpho-physiological traits in Biroin rice of Bangladesh. Int J Plant Breed Genet. 2011;5(3):246–254.

[21] Pandey P, Anurag PJ, Tiwari D, et al. Genetic variability, diversity and association of quantitative traits with grain yield in rice (Oryza sativa L.). J BioSci. 2009;17:77–82.

[22] Umadevi M, Veerabadhiran P, Manonmani S. Genetic variability, heritability, genetic advance and correlation for morphological traits in rice genotypes. Madras Agric J. 2009;96(7–12):316–319.

[23] Pandey P, Avasthi L. Studies on genetic variability for yield contributing traits in aromatic rice. Crop Res (Hisar). 2002;23(2):214–218.

[24] Abarshahr M, Rabiei B, Lahigi HS. Assessing genetic diversity of rice varieties under drought stress conditions. Notul Sci Biol. 2011;3(1):114–123.

[25] Chandra S, Ziemke J, Schoeberl M, et al. Effects of the 2004 El Nino on tropospheric ozone and water vapor. Geophys Res Lett. 2007;34(6):1–5.

[26] Dutta R, Baset M, Khanam S. Plant architecture and growth characteristics of fine grain and aromatic rices and their relation with grain yield. International Rice Commission Newsletter (FAO) Bulletin de la Commission Internationale du Riz (FAO) Noticiario de la Comision Internacional del Arroz (FAO); 2002.

[27] Mondal M, Islam A, Siddique M. Performance of 17 modern transplant aman cultivar in the northern region of Bangladesh. Bangladesh J Crop Sci. 2005;16:23–29.

[28] Lush JL. Heritability of quantitative characters in farm animals. 1949;35(S1):356–375.

[29] Iftekharuddaula K, Badshah M, Hassan M, et al. Genetic variability, character association and path analysis of yield components in irrigated rice (Oryza sativa L.). Bangladesh J Plant Breed Genet. 2001;14(2):43–49.

[30] Kavitha S, Reddy S. Variability, heritability and genetic advance of some important traits in rice (Oryza sativa L.). Andhra Agric J. 2002;49(3–4):222–224.

[31] Pandey P, Anurag PJ. Estimation of genetic parameters in indigenous rice. Adv Agric Bot. 2010;2(1):79–84.

[32] Habib S, Bashar M, Khalequzzaman M, et al. Genetic analysis and morpho-physiological selection criteria for traditional Biroin Bangladesh rice germplasms. J Biol Sci. 2005;5(3):315–318.

[33] Akhtar N, Nazir M, Rabbawaz A, et al. Estimation of heritability, correlation and path coefficient analysis in fine grain rice (Oryza sativa L.). J Anim Plant Sci. 2011;21(4):660–664.

[34] Zahid MA, Akhtar M, Sabir M, et al. Correlation and path analysis studies of yield and economic traits in Basmati rice (Oryza sativa L.). Asian J Plant Sci. 2006;5(4):643–645.

[35] Shahidullah S, Hanafi M, Ashrafuzzaman M, et al. Biomass accumulation and energy conversion efficiency in aromatic rice genotypes. C R Biol. 2010;333(1):61–67.

[36] Ghosh S, Sharma D. Research note genetic parameters of agr-o-morpho-physiological traits in rice (Oryza sativa L.). Electron J Plant Breed. 2012;3(1):71–714.

[37] Chaudhury P, Das P. Genetic variability, correlation and path coefficient analysis in deep water rice. Ann Agric Res. 1998;19(2):120–124.

[38] Paramasivam K, Giridharan S, Kirubakaran Soundararaj A, et al. Heterosis and combining ability for grain characters in rice. Madras Agric J. 1996;83:110–114.

[39] Ahmadikah A, Nasrollaniejad S, Alishah O. Quantitative genetic analyses and morpho-anatomical genetic diversity in American wild rice species. Philos Trans R Soc Lond B Biol Sci. 2008;363(1538):297–308.

[40] Veasey EA, Silva EFD, Schammass EA, et al. Morphoagronomic genetic diversity in ancestral lines of Philippine rice cultivars. Philippine J Crop Sci. 1996;21(3):86–92.

[41] Caldo R, Sebastian L, Hernandez J. Morphology-based genetic diversity analysis of ancestral lines of Philippine rice cultivars. Philippine J Crop Sci. 1996;21(3):86–92.

[42] Lasalita-Zapico FC, Namocatcat JA, Carino-Turner JL. Genetic diversity analysis of traditional upland rice cultivars in Kihan, Malapatan, Sarangani Province, Philippines using morphometric markers. Philippine J Sci. 2010;139(2):177–180.