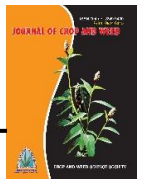
Journal of Crop and Weed, 20(1):1-11 (2024)

ISSN-O : 2349 9400; P : 0974 6315



https://www.cwss.in
www.cropandweed.com

Population structure, favourable alleles and grain quality traits of promising advanced breeding lines of lowland rice adaptable to acidic soils

[*1]I. GOPINATH, [2,5]M. RAI, [3,5] W. TYAGI,

[4]A. DEBNATH AND [6]K. L. BHUTI

[1] *Division of Genetics, ICAR-Indian Agricultural Research Insititute, New Delhi, India*
[2] *Post Graduate College of Agriculture, Dr. Rajendra Prasad Central Agricultural University, Bihar, India*
[3] *International Crop Research Institute for Semi-Arid Tropics, India*
[4] *Department of Genetics and Plant Breeding, Rajiv Gandhi University, Itanagar, Arunachal Pradesh, India*
[5] *School of Crop Improvement, College of Post Graduate Studies in Agricultural Sciences, Central Agricultural University, Imphal, Manipur*
[6] *Department of Agricultural Biotechnology and Molecular Biology, Dr. Rajendra Prasad Central Agricultural University, Bihar, India*

Received: 13.01.2024, Revised: 29.01.2024; Accepted: 31.01.2024

DOI: https://doi.org/10.22271/09746315.2024.v20.i1.1753

ABSTRACT

Advanced Breeding Lines (ABLs) developed from crosses between acidic soil adapted genotypes and elite varieties were evaluated phenotypically under phosphorus deficient acidic lowland soil for sixteen grain quality traits. The rationale was to identify ABLs adapted to various biotic and abiotic stresses with desired grain quality, which can be used in breeding pipeline to develop promising lines. Genetic relationship was ascertained using genome-wide SSRs and SNPs along with trait linked markers for blast resistance (Pi(20)t, Pi54 and Pita), chalkiness (Chalk5), etc. Amount of iron and zinc in unpolished dehulled grains were estimated using energy dispersive X- ray fluorescence spectrophotometer. Amylose content ranged from 18% to 27% among the breeding lines, while the protein content ranged from 5.5% to 11%. We were able to identify superior ABLs on the basis of higher yield and presence of favourable alleles for key traits including grain quality. Four diverse breeding line, (CAUS 103, CAUS 104, CAUS 105 and CAUS 107) with lesser disease score, higher yield and grain quality traits suitable to Northeast India consumer preference identified in the current study could be nominated for multi-locational trials or can be used as parents in further breeding programmes. The study also provides a broad framework of breeding strategy, prioritizing traits and parental selection for rice breeding in acidic region.

Keywords: Rice, ABLs, grain quality, acidic soil

The North Eastern Hill Region of India (NEHR) had experienced minimal impact of the Green Revolution with an insignificant change in varietal landscape of rice resulting in majority of farmers still growing indigenous heirloom genotypes adapted to acidic soils having low yield, but with preferred taste and other grain qualities. This is one of many reasons for less productivity in the region as compared to national average, along with food grain deficit of -2.51% in the region (Roy *et al.*, 2015). Crop improvement strategy in the region, therefore, should aim to utilize the genetic diversity of local landraces to increase productivity and induce tolerance to many biotic and abiotic stresses prevailing in the region such as rice blast disease and highly acidic infertile soils; while maintaining/enhancing the preferred grain quality characteristics. Molecular characterization of the breeding material is required for identification and genetic purity testing of the breeding lines as favourable, distinct and diverse high yielding lines are the future of rice improvement strategy.

**Email: gopiikkurti@gmail.com*

How to cite: Gopinath, I., Rai, M., Tyagi, W., Debnath, A. and Bhuti, K.L. 2024. Population structure, favourable alleles and grain quality traits of promising advanced breeding lines of lowland rice adaptable to acidic soils. *J. Crop and Weed*, 20(1): 1-11.

Thousands of microsatellite (SSR) markers with details pertaining to chromosomal location and polymorphism levels are available to rice researchers (McCouch *et al.,* 2002). SSRs and SNP based polymorphism has been extensively used in rice for assessing the level and structure of genetic diversity. Also, several gene-based/ linked markers are reported for major genes for agronomically important traits like blast resistance. Designated standard genome-wide DNA markers along with genic markers help in a reliable and efficient selection of lines for the earmarked breeding programmes aiming at high yield coupled with desirable grain quality. Grain quality preference of consumers must be kept in view while selecting high-yielding advanced generation promising lines for breeding. Rice grain quality is a combinatorial description of grain appearance, nutritional contents and palatability of cooked rice kernel in addition to conformation and colour of whole rice kernel (Yun *et al*., 2014). Grain quality traits of rice are complex physicochemical properties (Juliano, 1971) *viz.*, dimension, shape and weight, fragmentation, hardness, milling properties, the chemical composition of the endosperm, aroma and nutritional factors. Being mostly sold as polished grain, physical quality of rice is immediately obvious to consumers as decided by differences in chalkiness, grain dimensions and firmness on a visual-sensory basis. Grain preferences of Southeast Asia are unique. Lower amylose levels (10 to 20%) along with a more cohesive cooked short grain (Oslen and Purugganass, 2002) in rice are preferred in Southeast Asia including Northeast India. Whatever the preferences, lack of dietary diversification in developing countries has increased micronutrient deficiency of iron and zinc in rice consumers. There is approximately a four-fold difference reported in iron and zinc concentration in some varieties (Madhubabu *et al.,* 2017; Rao *et al.,* 2014). There is, therefore, a need for elite quality varieties biofortified with Fe and Zn that can be a stable and sustainable source of nutrition. A good breeding programme would, therefore, require characterization of genotypes or breeding lines concerning variability in grain quality as well. The present study reports molecular characterization and evaluation of advanced breeding lines of rice derived from different cross combinations for grain quality traits to select the best diverse high yielding lines superior to the checks for multiplication and future breeding programs.

MATERIALS AND METHODS

Plant materials

Twenty-two advanced breeding lines (ABLs) of rice in their F_7 generation used in the present study were derived from diverse parents consisting of elite varieties and acidic soil adapted rice genotypes (Table S1). The lines were tested against checks CAUR1 (LR19) and Shahsarang (LR11). The materials were sown in randomized block design (RBD) with three replications in 4 x 1 m^2 plots (5 lines per genotype) following a spacing of 20 cm x 20 cm in the experimental farm of College of Post Graduate Studies in Agricultural Sciences, CAU (I), Umiam, Meghalaya under rainfed lowland acidic conditions (for details regarding soil conditions kindly refer Yumnam *et al*., 2017). Per hectare yield of breeding lines was obtained from yield data of bulk harvested plants from the plot after multiplication with appropriate factor.

Genotyping

Leaf samples were collected from parents and each of the homozygous ABLs after 30 days of transplanting. Total genomic DNA was extracted using CTAB method (Doyle and Doyle, 1990). A set of 30 SSR markers (distributed over all the 12 rice chromosomes) (McCouch *et al*., 2002; http://www.gramene.org), five blast gene specific markers, 34 trait-based SNPs (available at International Rice Research Institute (IRRI)) were used for genotyping (Table S2). The temperature profile and PCR conditions for SSR markers were as per Gramene database (http://www.gramene.org). PCR components used for polymorphism survey were similar to Bhutia *et al.* (2021). The PCR products were loaded in 2% agarose gel along with 100 bp DNA ladder as size standard and alleles scored.

Grain quality characterization

Harvested dehulled grains were evaluated for sixteen quality traits including two computed variables based on physical, cooking and nutritional characteristics in completely randomized design (CRD) involving 3 replications. Seed length (SL), seed breadth (SB), seed length breadth ratio (SL:B), decorticated grain length (GL), decorticated grain breadth (GB), linear elongation after cooking (LE) and grain length breath ratio (GL:B) were measured using an android software SeedCounter v.1.9.6 (Komyshev *et al.* 2016). Chalkiness (CA) was evaluated based on the percent chalky area and scored following Rice Quality Reference Manual provided by International Rice Research Institute (IRRI). Grain length breadth ratio was interpreted as reported by Ramiah committee (1969). In addition, alkali spreading (ASV) test (Little *et al*., 1958; Jennings *et al.,* 1979), gelatinization temperature (GT) interpretation (Lanceras *et al.* 2000), protein content (PC) from unpolished grains (Lowry *et al.,* 1951) and amylose content (AC) (Juliano, 1971) were measured. Iron (Fe) and Zinc (Zn) content were estimated using unpolished dehulled grains using energy dispersive X- ray fluorescence spectrophotometer (ED-XRF).

Data Analysis

Clustering and parental ancestry estimation

The parental ancestry was estimated using Bayesian clustering method using the STRUCTUREv2.3.4 software (Pritchard *et al.,* 2000). Cluster likelihood estimation was calculated based on 10 independent runs for a

variable number of clusters, from *k*=2 to *k*=10 (length of burning period: 100,000 and number of Markov Chain Monte Carlo (MCMC) repeats: 50,000) with admixture model. The value *k*=4 was chosen due to the low variation of probability values (X) and repetitive clustering representing smallest stable *k* value as optimum value. Euclidean distance measures between individual genotypes were calculated according to modified Roger's distance (Wright, 1978). Aglomerative hierarchical clustering of genotypes was performed using Ward's method (Ward, 1963). Cluster analysis was performed using R platform (https://www.r-project.org).

Statistical analysis

Summary statistics such as mean, standard deviation (SD), minimum and maximum values, and coefficient of variation (CV) were determined for quality traits. The data was subjected to analysis of variance (ANOVA). The phenotypic and genotypic variances (Burton and De Vane, 1953), phenotypic and genotypic level of coefficient of variation (Johnson *et al.,* 1955) and categorization (Sivasubramanian and Madhavamenon, 1973), broad sense heritability (Hanson *et al.,* 1956) and categorization (Johnson *et al.,* 1955) were analysed. Genetic advance (GA) and genetic advance as percentage of mean (GAM) were calculated for each of the traits studied to test for additive effect and heritability.

RESULTS AND DISCUSSION

Marker profiling of breeding lines

A total of 52 alleles varying from 1 (RM154) to 4 (RM152) per marker were detected at 27 microsatellite loci across the 22 lines. Polymorphism was detected for 15 loci on chromosomes 1, 2, 4, 7, 8 and 9; whereas, the SSR markers present on chromosomes 3, 5, 6, 10, 11 and 12 were monomorphic. Genotype specific, unique bands produced at four loci in parent LR25 were not amplified in respective progeny lines. Relatively higher number of alleles from LR11, being the most common parent, were observed across all loci tested. Out of 34 SNPs tested, 21 were polymorphic among lines identifying 41 heterozygotes from 18 SNPs with a maximum of 14 heterozygotes found at locus snpOS037. Lines with high grain Zn content were clearly distinguished by snpOS0300. Three SSR markers linked to genes *Pi54*, *Pi20(t)* and *Pi5* as well as three SNPs specific for genes *Pita*, *Pi54* and *Chalk5* were found to be polymorphic among the breeding lines and parents. Favourable SSR and SNP alleles were observed for *Pi (20)t* (CAUS 101, CAUS 103, CAUS 104, CAUS 105, CAUS 110, CAUS 120 and CAUS 122), *Pi54* (CAUS 104, CAUS 107, CAUS 109, CAUS 116 and CAUS 117), *Pita* (CAUS 113, CAUS 115, CAUS 118 and CAUS 119) and *Chalk5* (CAUS 102, CAUS 103, CAUS 104, CAUS 114, CAUS 116, CAUS 118, CAUS 120 and CAUS 121). SNP data revealed no favourable alleles for genes *xa13*, *Pi9*, *BADH2*, *Sub1*, *Xa5*, *Xa21* and *Bph17* (Table 1). Lines found to be carrying multiple (>1) favourable alleles were CAUS 103, CAUS 104, CAUS 107, CAUS 116, CAUS 118 and CAUS 120.

Table 1: Favourable alleles observed in some of the ABLs

Advanced breeding line	Genes for which favourable alleles observed
CAUS 101	*Pi(20)t*
CAUS 102	*Chalk5*
CAUS 103	*Chalk5*, *Pi(20)t*
CAUS 104	*Chalk5*, *Pi54*, *Pi(20)t*
CAUS 105	*Pi(20)t*
CAUS 106	Nil
CAUS 107	*Pita*, *Pi54*
CAUS 108	Nil
CAUS 109	*Pi54*
CAUS 110	*Pi(20)t*
CAUS 111	Nil
CAUS 112	Nil
CAUS 113	*Pita*
CAUS 114	*Chalk5*
CAUS 115	*Pita*
CAUS 116	*Chalk5*, *Pi54*
CAUS 117	*Pi54*
CAUS 118	*Pita*, *Chalk5*
CAUS 119	*Pita*
CAUS 120	*Chalk5*, *Pi(20)t*
CAUS 121	*Chalk5*
CAUS 122	*Pi(20)t*

Clustering and STRUCTURE analysis

In order to understand the genotypic relationship among ABLs, STRUCTURE as well as Euclidean distance estimates were calculated. While six of the ABLs were classified into one of four major groups at *k*= 4, sixteen lines were admixtures (<85% of estimated derived ancestry) (Fig. 1a). Maximum divergence was found between cluster 1 and 3 as well as between cluster 2 and 3. Four breeding lines were categorized into same genomic group with parent LR11. The majority of admixtures occur within progeny group derived from cross between LR11 and LR15, with few notable exceptions such as CAUS 116. Lines CAUS 110, CAUS 111 and CAUS 122 show similar ancestry being half sib families. When the proportion of ancestry required to identify ancestry share was reduced to 70%, ABLs CAUS 103, CAUS 104, CAUS 106, CAUS 112, CAUS 113 and CAUS 115 still did not cluster clearly and were classified as admixed. LR15, as a parent was categorized as admixed based on the markers run in the current study and hence the breeding lines CAUS 117, CAUS 118 and CAUS 119 were also admixtures. The parents LR19 and LR23 belong in the same ancestry group.

The highest genetic distance was observed between CAUS 117 and CAUS 116 (7.34), and the lowest between CAUS 110 and CAUS 111 (1.03) and between CAUS 111 and CAUS 122 (1.05). These results correlate with observations from population structure. Dendrogram obtained from agglomerative hierarchical clustering was supported by AU/BP (approximately unbiased/bootstrapping probability) values. All the genotypes were grouped into two major clusters (Fig. 1b). Cluster I comprised of eight lines and two parents, and cluster II comprised of fourteen lines and five parents which was further sub-divided. LR18, LR13, LR25 and CAUS 116 were found to be outliers. From Gopinath *et al.* (2022), evaluation of association between the genetic distance observed for morphological-molecular profiles was not significant. However, on the basis of phylogeny, certain trends were observed. The lack of association could be attributed to several factors including sampling procedure and size, marker density and statistical approaches.

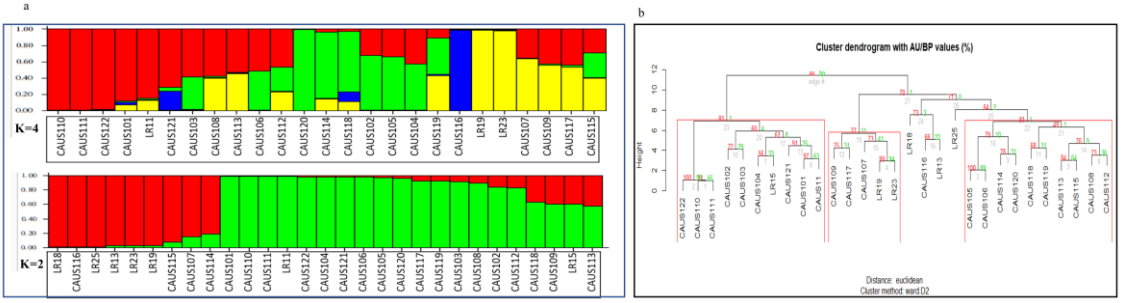


Fig. 1. Population structure for twenty-two breeding lines along with parents at *k*=4 (based on SSR markers only) and *k*=2 (based on SSRs and SNPs) using STRUCTURE (a) and cluster dendrogram based on polymorphic markers (b). Colours denote estimated proportions of shared ancestry in a genotype (a); AU (approximately unbiased) and BP (bootstrapping probability) values are given on the left and right, respectively over each grouping (b)

Quality characterization

Analysis of variance for grain quality data showed highly significant differences (P<0.01) among the breeding lines for all the sixteen traits. The descriptive statistics and mean values of all the traits under study are given in Table 2. PCV ranged from 21.73% for protein content, 12.41% for amylose content to 4.56% for seed length. Similarly, highest GCV was observed for protein content (21.52%) and lowest for seed length (4.14%). Higher difference between PCV and GCV was observed for linear elongation, test weight, protein content and computed grain length breadth ratio. The results were in conformity with earlier findings of Devi *et al.* (2016), Thongbam *et al.* (2012) and Madhubabu *et al.* (2017).

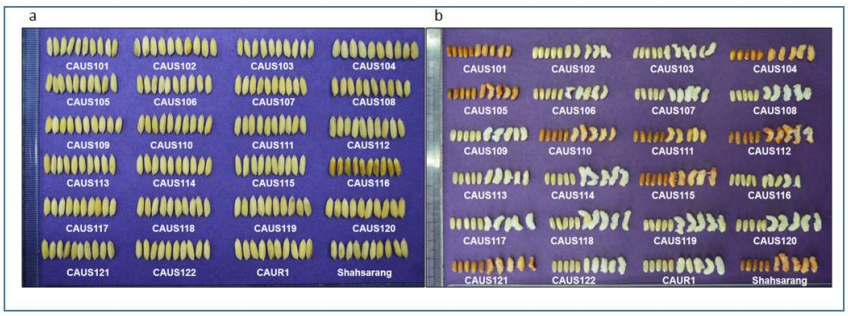


Fig. 2. Comparison between un-decorticated (a) and decorticated seeds (before and after cooking) (b) for twenty-two breeding lines with two checks

Table 2. Components of variance and genetic parameters for quality traits

Traits	Vp	Vg	PCV (%)	GCV (%)	ECV (%)	H	GA	GAM
TW	0.08	0.07	10.54	9.96	3.42	0.89	0.51	19.41
SL:B	0.03	0.03	6.56	5.81	3.05	0.78	0.30	10.59
SL	0.12	0.10	4.56	4.14	1.91	0.82	0.58	7.73
SB	0.02	0.02	5.13	4.74	1.96	0.85	0.24	9.02
GL	0.09	0.08	5.75	5.44	1.87	0.89	0.54	10.59
GB	0.02	0.01	6.49	6.15	2.07	0.90	0.24	12.01
GL:B	0.04	0.03	7.74	7.15	2.96	0.85	0.35	13.61
LE	0.02	0.02	7.70	6.67	3.84	0.75	0.23	11.91
AC	8.10	8.06	12.41	12.38	0.89	0.99	5.83	25.44
PC	2.65	2.60	21.73	21.52	2.99	0.98	3.29	43.91

Vp, Phenotypic variance; Vg, genotypic variance; PCV, phenotypic coefficient of variation; GCV, genotypic coefficient of variation; ECV, environmental coefficient of variation; H, heritability (broad sense); GA, genetic advance; GAM, genetic advance over mean.

Estimates of broad sense heritability were observed as highest for amylose content (0.99) followed by protein content (0.98) and lowest for linear elongation (0.75); however, heritability was high for all the traits evaluated. Estimated genetic advance (GA) was found to be the highest for amylose content (5.83) followed by protein content (3.29) and minimum for grain breadth (0.24). High heritability coupled with high expected genetic advance as percent of mean (GAM) was observed for protein content (43.91), amylose content (25.44) and test weight (19.41). Therefore, major role of additive gene action and scope for improvement through individual plant selection is suggested.

As per DUS characterization, lemma and palea colour did not differ among the breeding lines, except CAUS 116 which had brown furrows on straw of lemma and palea (Fig. 2a). The kernel colour differed primarily between red and brownish yellow for the lines under study (Fig. 2b). High linear elongation on cooking was observed ranging maximum for CAUS 118 (2.26 fold) to the least for CAUS 103 (1.63 fold). Besides, CAUS 104 (2.13), CAUS 105 (2.12), CAUS 108 (2.10) and CAUS 118 (2.26) showed more than two-fold linear expansion after cooking (Table S3). The highest seed length was recorded for CAUS 112 (8.38 mm) and the lowest for CAUS 116 (7.04 mm). However, all the ABLs were categorized as short in length as per the DUS guidelines. ABLs were characterised as medium for seed breadth except CAUS 101 (2.43), CAUS 116 (2.48), and CAUS 104 (3.03 mm), which were categorised as narrow and broad, respectively. Variation in computed decorticated grain length breadth ratio categorized 17 of the breeding lines and checks as short-slender for grain shape and 5 lines as short-bold (CAUS 102, CAUS 104, CAUS 108, CAUS 109 and CAUS 110). Besides, CAUS 104 (2.25) and CAUS 102 (2.22) were the broadest; whereas, CAUS 115 (5.54) had the longest grain length. High test weight (≥3 g) was recorded for CAUS 104, CAUS 112 and CAUS 120. The lines differing for percentage of chalkiness were scored as 1, 5 and 9. Six lines were given score 1 for least (<10%) chalky area (CAUS 101, CAUS 105, CAUS 108, CAUS 110, CAUS 113 and CAUS 116). The checks, CAUR1 (1) and Shahsarang (9), had extreme chalky scores, simplifying the comparison with breeding lines. The distribution of physical grain quality traits is represented in Fig. 3. Gelatinization temperature based on alkali spreading value was categorised as low (CAUS 111, CAUS 112, CAUS 114, CAUS 115, CAUS 119 and CAUS 120), intermediate (CAUS 101, CAUS 103, CAUS 107, CAUS 109, CAUS 110, CAUS 113, CAUS 117, CAUS 122 and CAUS 122) and high (CAUS 102, CAUS 104, CAUS 105, CAUS 106, CAUS 108, CAUS 116 and CAUS 118). The protein content was recorded as maximum in CAUS 102 (11.05%) and minimum in CAUS 117 (5.64%) with an average value of 7.49%. The checks CAUR1 and Shahsarang, included in the study had protein content of 5.79% and 7.34%, respectively; whereas, 21 lines recorded <10% protein content. The amylose content ranged from 27.58% (CAUS 115) to 18.02% (CAUS 107). The amylose content of checks was recorded as 20.91% and 27.18% for CAUR1 and Shahsarang, respectively (Fig. 4a). The highest amount of Zn and Fe was recorded in CAUS 122 (30.10 ppm and 12.90 ppm, respectively) and the lowest amount of both Zn and Fe were recorded in CAUS 116 (21 ppm and 9.40 ppm, respectively) (Fig. 4b).

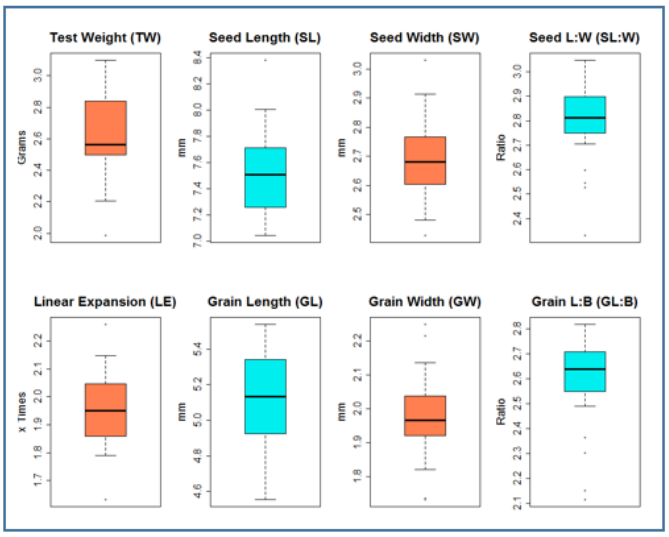


Fig. 3. Distribution of 8 physical quality traits: TW, SL, SW, SL:W, LE, GL, GW, GW. Bold line within boxplots represents the median, box edges represent upper and lower quantiles. The whiskers are 1.5 times the quantile of the data

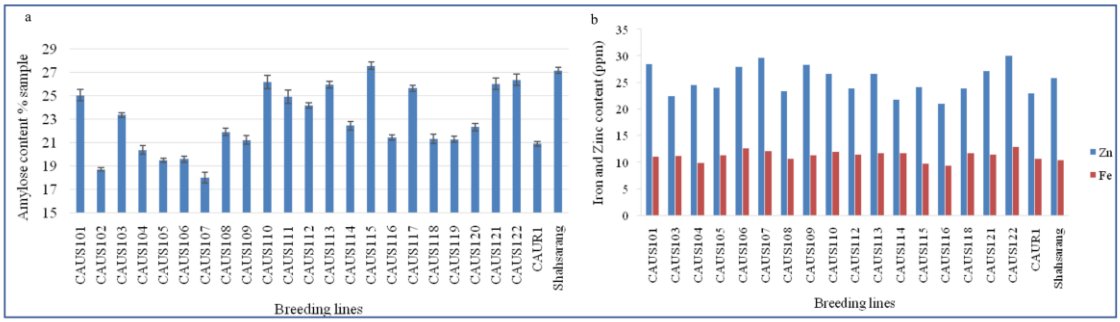


Fig. 4. Variation observed for amylose content (a) and micronutrient (Zn and Fe) content (in ppm) (b) among the breeding lines and checks

In the current study, a set of ABLs were tested in natural acidic soil conditions of farm without application of phosphatic fertilizers, using contrasting parents for combining grain yield, grain quality characteristics, and biotic and abiotic stress resistance, while ensuring sufficient diversity in the improved population with the help of molecular markers provides means of broadening the genetic base and improve traits in a breeding programme. Molecular markers lead to characterization of a genotype with unique fingerprint, thereby assisting breeder to select for parents from a collection of advanced breeding materials.

Several advanced breeding lines were identified carrying favourable alleles for multiple genes (*Pi54*, *Pita*, *Pi(20)t* and *Chalk5*) with a blast resistant score of 0-2 (unpublished data). Therefore, these promising blast resistant lines (CAUS 103, CAUS 104, CAUS 107 CAUS 116, CAUS 118 and CAUS 120) could be considered for multiplication or as parents in specific breeding programs for introgression of blast resistance genes.

Analysing ancestry share using STRUCTURE revealed population relatedness of half sib lines derived from different crosses. In the present study, five breeding lines were found to show more than 85% common ancestry, hence they could be characterised as belonging to a sub-group. The parents LR19 and LR23 show similar ancestry but are genetically distinct (Yumnam *et al.*, 2017). ABL CAUS 116 grouped into unique one suggesting that further study with more markers is needed to better understand and utilise this line along with the parent PAU201. At *k*=2, all the parents of the breeding lines were grouped separately from LR11. However, ABLs were found to carry maximum of LR11 ancestry signifying the skewness towards LR11 alleles after every selection. Euclidean distance showed considerable amount of divergence between CAUS 121 and CAUS 122, though derived from LR11. This is in consonance with the results obtained from STRUCTURE, where CAUS 122, CAUS 110 and CAUS 111 show similar ancestry though derived from different crosses with LR11

as common parent. CAUS 102 and CAUS 104 being the lines with bold grains clustered together.

Rice improvement programmes need to address grain quality parameters as well, so as to make rice growing remunerative (Fitzgerald, 2017). Genotypes adapted to acidic soils of the NE Indian region offer valuable gene pool and traits including grain colour for utilization in varietal improvement/development. Among the twelve red coloured lines identified, CAUS 104, CAUS 105 and CAUS 110 were >25% high yielding over the checks with acceptable levels of amylose content. Chalkiness determines the extent of grain breakage during milling. Our study identified six breeding lines without chalky regions and nine lines with acceptable minimal chalkiness suitable with high market value. Considering better chalky score, higher yield over checks, desirable grain quality, short slender grain shape and two-fold elongation after cooking, CAUS 103, CAUS 105, CAUS 108 and CAUS 110 can be targeted for multiplication. However, due to high amylose content in CAUS 110, it might not be acceptable by consumers in the Northeastern hilly states of India. Cooking and sensory properties of rice varieties influence seed demand. Sensory association of amylose content is such that it is positively correlated with hardness and negatively correlated with stickiness after cooking. Based on preference for sticky and soft rice with bold grains in Northeastern region of India, admixed lines CAUS 104, CAUS 105 and CAUS 107 with favorable alleles (*Pi(20)t*, *Pi54*, and *Pita*) and minimal chalk were identified as suitable for the location and target specific breeding programmes. Iron and zinc content estimated from unpolished grains was reported higher than checks in eight and twelve lines, respectively. These lines as per Indian Council of Agricultural Research guidelines (polished rice grains with >24 ppm Zn and >10 ppm Fe) can be labeled as micronutrient biofortified. Diverse ABLs with favourable alleles, higher yield, and better grain quality are important for broadening the genetic base of a breeding program. Any breeding line identified for good grain quality to meet consumer and ethnic perspectives should nevertheless give measurable yields in order to be recommended for varietal cultivation. As reported by Rai *et al.* (2020), based on agro-morphological characterization of the ABLs, the yield per hectare of breeding lines was highest for CAUS 110. Total of 14 lines were superior over the checks CAUR1 and Shahsarang by a maximum of 40% (CAUS 110) over the best check.

SUPPLEMENTARY MATERIAL

Supplementary Table S1. List of advanced breeding lines against their parentage used in the study

S. No.	Advanced breeding line	Parentage
1	CAUS 101	LR11 (Shahsarang) × LR18 (Paijong)
2	CAUS 102	LR11 (Shahsarang) × LR23 (Shahbhagi Dhan)
3	CAUS 103	LR11 (Shahsarang) × LR25 (Mynrii)
4	CAUS 104	LR11 (Shahsarang) × LR15 (Priya)
5	CAUS 105	LR11 (Shahsarang) × LR15 (Priya)
6	CAUS 106	LR11 (Shahsarang) × LR15 (Priya)
7	CAUS 107	LR11 (Shahsarang) × LR19 (CAUR1)
8	CAUS 108	LR11 (Shahsarang) × LR19 (CAUR1)
9	CAUS 109	LR11 (Shahsarang) × LR19 (CAUR1)
10	CAUS 110	LR11 (Shahsarang) × LR19 (CAUR1)
11	CAUS 111	LR11 (Shahsarang) × LR19 (CAUR1)
12	CAUS 112	LR11 (Shahsarang) × LR13 (IR64)
13	CAUS 113	LR11 (Shahsarang) × LR13 (IR64)
14	CAUS 114	LR11 (Shahsarang) × LR13 (IR64)
15	CAUS 115	LR11 (Shahsarang) × LR13 (IR64)
16	CAUS 116	LR11 (Shahsarang) × PAU201
17	CAUS 117	LR13 (IR64) × LR15 (Priya)
18	CAUS 118	LR13 (IR64) × LR15 (Priya)
19	CAUS 119	LR13 (IR64) × LR15 (Priya)
20	CAUS 120	LR13 (IR64) × LR15 (Priya)
21	CAUS 121	LR11 (Shahsarang) Backcross lines
22	CAUS 122	LR11 (Shahsarang) Backcross lines
23	CAUR1	Checks
24	Shahsarang	

Table S2. List of molecular markers used in the study

Panel of 30 diagnostic SSR markers			Random SNPs		
Sl. No.	**Marker**	**Chromosomal location**	**Sl. No.**	**Marker**	**Nucleotide polymorphism**
1	RM495	1	**1**	snpOS0296	A:G
2	RM1	1	**2**	snpOS0298	T:A
3	RM283	1	**3**	snpOS0307	G:T
4	RM5	1	**4**	snpOS0309	C:C
5	RM237	1	**5**	snpOS0312	G:A
6	RM431	1	**6**	snpOS0295	C:T
7	RM154	2	**7**	snpOS0297	T:C
8	RM452	2	**8**	snpOS0299	A:T
9	RM338	3	**9**	snpOS0300	T:C
10	OSR13	3	**10**	snpOS0301	T:G
11	RM514	3	**11**	snpOS0302	A:G
12	RM124	4	**12**	snpOS0303	A:T
13	RM507	5	**13**	snpOS0304	T:A
14	RM161	5	**14**	snpOS0305	G:T
15	RM133	6	**15**	snpOS0306	G:A
16	RM162	6	**16**	snpOS0308	A:C
17	RM125	7	**17**	snpOS0310	C:C
18	RM455	7	**18**	snpOS0311	G:T
19	RM118	7	**19**	snpOS0313	NA
20	RM408	8	**20**	snpOS0314	G:A
21	RM152	8	**21**	snpOS0315	C:G
22	RM44	8	**22**	snpOS0316	G:A
23	RM284	8	**23**	snpOS0317	C:A
24	RM433	8	**24**	snpOS0318	C:G
25	RM447	8	***Gene specific SNPs***		
26	RM215	9	**1**	snpOS0002 (*xa13*)	A:A
27	RM316	9	**2**	snpOS0006 (*Pita*)	A:C
28	RM271	10	**3**	snpOS0007b (*Pi9*)	C:C
29	RM536	11	**4**	snpOS0015 (*Pi54*)	C:T
30	RM277	12	**5**	snpOS0022 (*BADH2*)	C:A
Gene linked markers			**6**	snpOS0024 (*Chalk5*)	A:G
1	MRG4766 (*Pi1*)	2	**7**	snpOS0040 (*Sub1*)	C:C
2	JJ803 (*Pi5*)	9	**8**	snpOS0054 (*Xa5*)	C;T
3	RM224 (*Pi54*)	11	**9**	snpOS0061 (*Xa21*)	G:G
4	5083InDel (*qPbm11*)	11	**10**	snpOS0068 (*Bph17*)	G:G
5	RM7102 (*Pi(20)t*)	12			

Table S3. Quality trait values recorded for each breeding line under study

Breeding Lines	Traits															
	TW (g)	**SL:W**	**SL (mm)**	**SW (mm)**	**GL (mm)**	**GW (mm)**	**GL:W**	**LE**	**AC (%)**	**PC (%)**	**Zn (ppm)**	**Fe (ppm)**	**ASV***	**GT***	**Kernel Colour***	**CA Score[@]**
CAUS101	2.21	2.96	7.2	2.43	4.86	1.82	2.67	1.84	25.05	8.27	28.5	11.1	I	I	R	1
CAUS102	2.9	2.53	7.36	2.91	5.10	2.22	2.30	1.86	18.72	11.05	NA	NA	L	H	W	9
CAUS103	2.49	2.93	7.46	2.55	5.09	1.88	2.71	1.63	23.37	6.07	22.5	11.2	I	I	W	5
CAUS104	3.1	2.33	7.06	3.03	4.76	2.25	2.12	2.13	20.36	7.77	24.5	9.9	L	H	R	9
CAUS105	2.58	2.77	7.5	2.71	5.21	2.05	2.55	2.12	19.49	5.68	24	11.4	L	H	R	1
CAUS106	2.5	2.7	7.24	2.68	5.17	2.03	2.55	2.07	19.59	8.43	28	12.7	L	H	W	5
CAUS107	2.52	2.81	7.25	2.59	5.05	1.92	2.63	1.91	18.02	7.70	29.6	12.1	I	L	W	5
CAUS108	2.47	2.77	7.32	2.65	4.71	1.99	2.36	2.10	21.92	6.67	23.4	10.7	L	H	W	1
CAUS109	2.58	2.55	7.09	2.79	4.59	2.14	2.15	2.02	21.23	6.30	28.3	11.3	I	I	W	5
CAUS110	2.55	2.85	7.64	2.68	5.09	2.05	2.49	1.97	26.18	6.50	26.7	12	I	I	R	1
CAUS111	2.49	2.87	7.68	2.68	5.18	1.99	2.61	1.90	24.91	7.12	NA	NA	H	L	R	5
CAUS112	3.04	3.02	8.38	2.77	5.45	2.10	2.60	1.95	24.17	6.19	23.9	11.5	H	L	R	5
CAUS113	2.4	3.05	7.74	2.54	5.32	1.96	2.71	1.89	25.95	8.03	26.7	11.8	I	I	W	1
CAUS114	2.51	2.97	7.80	2.62	5.43	2.02	2.68	1.82	22.45	7.54	21.8	11.8	H	L	W	5
CAUS115	2.68	3.03	8.00	2.65	5.54	1.97	2.82	1.85	27.58	6.70	24.2	9.8	H	L	R	9
CAUS116	1.98	2.84	7.04	2.48	4.55	1.74	2.62	2.15	21.42	8.59	21	9.4	L	H	W	1
CAUS117	2.68	2.84	7.51	2.65	5.17	1.92	2.69	1.86	25.65	5.64	NA	NA	I	I	R	9
CAUS118	2.85	2.87	7.87	2.74	5.36	1.94	2.77	2.26	21.34	7.41	23.9	11.8	L	H	R	9
CAUS119	2.82	2.76	7.66	2.78	5.43	1.96	2.77	1.97	21.29	8.96	NA	NA	H	L	R	9
CAUS120	3.1	2.74	7.61	2.77	5.33	1.97	2.7	1.96	22.34	9.28	NA	NA	H	L	R	9
CAUS121	2.51	2.6	7.38	2.74	4.99	1.94	2.57	1.94	26.02	8.03	27.1	11.5	I	I	R	5
CAUS122	2.62	2.8	7.54	2.7	4.81	1.73	2.71	2.01	26.37	8.83	30.1	12.9	I	I	W	5
CAUR1	2.93	2.82	7.77	2.76	5.47	1.95	2.8	1.79	20.91	5.79	23.8	7.9	H	L	W	1
Shahsarang	2.55	2.8	7.26	2.59	5.07	1.91	2.65	1.99	27.18	7.34	25.9	10.4	H	L	R	9
CD (5%)	0.18	0.14	0.24	0.09	0.16	0.07	0.13	0.12	0.34	0.45						

* L, I and H for ASV and GT represent Low, Intermediate and High respectively. R and W represent red and white kernel colours. [@] CA score represents 1 (<10%), 5 (10-20%), 9 (>20%).

CONCLUSION

In summary, genotypes adapted to acidic soils of Northeast India with desirable grain quality but poor yield is a potential regional bioresource. They are also adapted to various biotic and abiotic stresses, and therefore can be used in breeding pipeline to develop promising lines involving elite varieties as parents. In this study, ABLs derived from crosses involving genotypes adapted to acidic soils were identified on the basis of superior yield, presence of favourable alleles for key traits including grain quality. Four diverse breeding lines (CAUS 103, CAUS 104, CAUS 105 and CAUS 107) with lesser disease score, higher yield and grain quality traits suitable to Northeast India consumer preference identified in the current study could be nominated for multi-locational trials or can be used as parents in further breeding programmes. The study also provides a broad framework of breeding strategy, prioritizing traits and parental selection for rice breeding in acidic region.

REFERENCES

Bhutia, K.L., Nongbri, E.L., Sharma, T.O., Rai, M. and Tyagi, W. 2021. A 1.84-Mb region on rice chromosome 2 carrying *SPL4*, *SPL5* and *MLO8* genes is associated with higher yield under phosphorus-deficient acidic soil. *J. Appl. Genet.*, 62: 207-222.

Burton, G.W. and DeVane, E.H. 1953. Estimating heritability in tall fescue (*Festuca arundinacea*) from replicated clonal material. *Agronomy J.*, 45(10): 478–481.

Devi, R.K., Parimala, K., Venkanna, V., Lingaiah, N., Hari, Y. and Chandra, S.B. 2016. Estimation of variability for grain yield and quality traits in rice (*Oryza sativa* L.). *Int. J. Pure Appl. Biosci.*, 4(2): 250-255.

Doyle, J.J. and Doyle, J.L. (1990). Isolation of plant DNA from fresh tissue. *Focus*, 12: 13-15.

Fitzgerald, M. (2017). Rice: grain quality characteristics and management of quality requirements. 10.1016/B978-0-08-100719-8.00012-7.

Gopinath, I., Rai. A. and Rai, M. 2022. Association between phenotypic and genotypic distance matrices of advanced breeding lines of lowland rice (*Oryza sativa* L.) from diverse parents. *Int. J. Environ. Climate. Change*. 12 (6): 33–38.

Hanson, G.H., Robinson, H.F. and Comstock, R.E. 1956. Biometrical studies of yield in segregating population of Korean lespedeza. *Agronomy J.* 48: 268- 272.

Jennings, P.R., Coffman, W.R. and Kauffman, H.E. 1979. Rice improvements. International Rice Research Institute. Los Banos, Philippines. p.186.

Johnson, H.W., Robinson, H.F. and Comstock, R.E. 1955. Estimates of genetic and environmental variability in soybean. *Agronomy J.*, 47: 314–318.

Juliano, B.O. 1971. A simplified assay for milled rice amylose. *Cereal Sci. Today.*, 16(10): 334–339.

Komyshev, E., Genaev, M. and Afonnikov, D. 2016. Evaluation of the SeedCounter, a mobile application for grain phenotyping. *Front. Plant. Sci.*, **7**:1990.

Lanceras, J., Huang, Z.L., Naivikul, O., Vanavichit, A., Ruanjaichon, V. and Tragoonrung, S. 2000. Mapping of genes for cooking and eating qualities in Thai jasmine rice (KDML105). DNA Res., 7: 93–101.

Little, R.R., Hilder, G.B., Dawson, E.H. and Elsie, H. 1958. Differential effect of dilute alkali on 25 varieties of milled white rice. *Cereal Chem.*, 35:111-126.

Lowry, O.H., Rosebrough, N.J., Farr, A.L. and Randall, J. 1951. Protein measurement with the folin phenol reagent. *J. Biol. Chem.*, 193: 265-275.

Madhubabu, P., Suman, K., Rathod, R., Fiyaz, A., Rao, S., Sudhakar, P. and Neeraja, C.N. 2017. Evaluation of grain yield, quality and nutrients content in four rice (oryza sativa l.) genotypes. *Cur. J. App. Sci. Techol.* 22(1): 1-12.

McCouch, S.R., Teytelman, L., Xu, Y., Lobos, K.B. and Clare, K. 2002. Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). *DNA Res*., 9:199–207.

Oslen, K.M. and Purugganass, M.D. 2002. Molecular evidence on the origin and evolution of glutinous rice. *Genet*., 162: 941-950.

Pritchard, J.K., Stephens, M., Rosenberg, N.A. and Donnelly, P. 2000. Association mapping in structured populations. *Am. J. Human. Genet.*, 67: 170–181.

Rai, A., Rai, M., Borpatragohain, B. and Kumar, S. 2020. Assessment of genetic variability, heritability and genetic advance for yield in advanced breeding line (*Oryza sativa* L.) of low land rice in Meghalaya. *Int. J. Cur. Microbiol. Appl. Sci*., 9(3): 706–717.

Rao, S.D., Babu, M.P., Kota, S.P., Bhadana, V.P., Surekha, G.S., Neeraja, C.N. and Babu, R.V. 2014. Assessment of grain zinc and iron variability in rice germplasm using energy dispersive X-ray fluorescence spectrophotometer (ED - XRF). *J. Rice Res.*, 7: 45-52.

Roy, A., Singh, N.U., Dkhar, D.S., Mohanty, A.K., Singh, S.B. and Tripathi, A.K. 2015. Food security in North-East region of India- a

state wise analysis. *Agric. Econ. Res. Rev*., 28: 259-266.

Sivasubramanian. S. and Madhavamenon, P. 1973. Genetic analysis of quantitative characters in rice through diallel crosses. *Madras Agric. J*., 60:1097-1102.

Thongbam, P.D., Raychaudhury, M., Durai, A., Das, S.K. and Ramya, K.T. 2012. Studies on grain and food quality traits of some indigenous rice cultivars of north-eastern hill region of India. *J. Agri. Sci*., 4(3): 259-270.

Ward, J.H. 1963. Hierarchical grouping to optimize an objective function. *J. American Stat. Assoc.*, 58: 236-244.

Wright, S. 1978. Evolution and genetics of populations, vol IV. University of Chicago Press, Chicago.

Yumnam, J.S., Rai, M. and Tyagi, W. 2017. Allele mining across two low-P tolerant genes *PSTOL1* and *PupK20-2* reveals novel haplotypes in rice genotypes adapted to acidic soils. *Plant Genet. Resour*., 15:221-229.

Yun, W.B., Kim, M.G., Handoyo, T. and Kim, M.K. 2014. Analysis of rice grain quality-associated quantitative trait loci by using genetic mapping. *American J. Pl. Sci.*, 5(9): 22-30.