The Journal of Agricultural Science

cambridge.org/ags

Development of recombinant high yielding lines with improved protein content in rice (*Oryza sativa* L.)

K. Chattopadhyay[1], S. G. Sharma[1], T. B. Bagchi[1], K. A. Molla[1], S. Sarkar[1], B. C. Marndi[1], A. Sarkar[2], S. K. Dash[1] and O. N. Singh[1]

[1]ICAR-National Rice Research Institute, Cuttack, Odisha 753 006, India and [2]ICAR- Central Institute for Women in Agriculture, Bhubaneswar, Odisha 751003, India

Crops and Soils Research Paper

Cite this article: Chattopadhyay K, Sharma SG, Bagchi TB, Molla KA, Sarkar S, Marndi BC, Sarkar A, Dash SK, Singh ON (2018). Development of recombinant high yielding lines with improved protein content in rice (*Oryza sativa* L.). *The Journal of Agricultural Science* **156**, 241–257. https://doi.org/10.1017/S0021859618000230

Received: 16 July 2017
Revised: 21 February 2018
Accepted: 7 March 2018

Key words:
Bulk-pedigree breeding; grain quality; high protein rice; NIR spectroscopy; response to selection

Author for correspondence:
K. Chattopadhyay, E-mail: krishnenducrri@gmail.com

Abstract

Rice has the lowest grain protein content (GPC) among cereals. Efforts have been made to improve GPC through the modified bulk-pedigree method of selection. A total of 1780 F_8 recombinant lines were derived in the year 2013 from five different cross combinations involving two high-GPC landraces, namely ARC10075 and ARC10063, three high-yielding parents, namely Swarna, Naveen and IR64, and one parent, namely Sharbati, known for superior grain quality with high micronutrient content. Near-infrared spectroscopy was used to facilitate high-throughput selection for GPC. Significant selection differential, response to selection and non-significant differences between the predicted and observed response to selection for GPC and protein yield indicated the effectiveness of this selection process. This resulted in lines with high GPC, protein yield and desirable levels of amylose content. Further, based on high mean and stability for GPC and protein yield over the environments in the wet seasons of 2013, 2014 and the dry season of 2014, 12 elite lines were identified. Higher accumulation of glutelin fraction and non-significant change in prolamin/glutelin ratio in the grain suggested safe guarding of the nutritional value of rice grain protein of most of these identified lines. Since rice is the staple food of millions, the output of breeding for high GPC could have a significant role in alleviating protein malnutrition, especially in the developing world.

Introduction

Rice is a staple food for more than half of the world's population. Thus, improvement of grain yield is the primary target of plant breeders. Further, rice is also the main source of nutrition for a large section of those people. Therefore, improvement of the nutritional quality of rice grain is also highly important. Rice is generally deficient in protein. However, due to its balanced amino acid profile and high content (0.80) of highly nutritive and digestible glutelins, rice protein is a significant contributor to meet the increasing demand of high-quality cereal as a source of protein (Fitzgerald *et al.* 2009). Therefore, high-protein rice has the potential to enhance nutrition for the poor in rural areas who depend mainly on rice (Li *et al.* 2004). Grain protein content (GPC) is a significant factor in the nutritional quality of rice (Ufaz & Galili 2008). Among available rice germplasm, a wide range of variability in GPC (5–18% with an average of 9.5%) was observed by scientists at the International Rice Research Institute (IRRI), Manila. These data suggest the possibility of improving GPC of rice grain. Earlier studies reported that the high GPC trait has been transferred to a high-yielding background in many other cereals such as wheat and rye (Vasal 2002). In rice, many quantitative trait loci (QTLs) along with associated markers have also been identified for ensuring transfer of GPC to high yielding background (Yoshida *et al.* 2002; Aluko *et al.* 2004; Zheng *et al.* 2011; Yang *et al.* 2015). But due to low heritability and significant influence of crop nutrient management practices, improvement of rice cultivars for this quantitative trait through a simple breeding scheme is a real challenge (Khush & Juliano 1984). Screening of rice germplasm for high GPC is also a tedious job, which further aggravates the challenge. Overcoming these constraints requires high-throughput screening for GPC – a task that may facilitate identification of the most effective selection scheme to make rice a more nutritious cereal. The bulk-pedigree method of selection has proved quite effective for traits with low heritability (Wynne & Gregory 1981). Near-infrared (NIR) spectroscopy has also proved effective in predicting GPC and is a reliable tool for genetic analysis and high-throughput selection (Shao *et al.* 2011). In the present study, modified bulk-pedigree selection and high-throughput selection with NIR spectroscopy were combined to find high-yielding rice cultivars with high GPC. Rice protein is made of four components, glutelins, albumins, globulins and prolamins. Among these, glutelins form the major share (nearly 0.70–0.80).It largely determines protein

© Cambridge University Press 2018

CAMBRIDGE UNIVERSITY PRESS

quality due to the higher content of lysine and other essential amino acids. Protein molecules accumulate within two types of protein bodies (PB). It is known that PB-II in the endosperm is more digestible in humans than PB-I and therefore is nutritionally more important. As PB-I contains mostly prolamins, which are nutritionally inferior, and PB-II contains mostly glutelins, nutritionally superior (Ogawa *et al.* 1987), the fractionation of protein and quantitative assessment of glutelins and prolamins through sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) can extrapolate the nutritional quality of protein (Jiang *et al.* 2014). A higher prolamin content increases the hardness of cooked rice. Therefore, maintaining the prolamin/glutelin ratio ensures a high cooking quality of rice. The current investigation focused on the extent of variability for GPC among derived lines and the magnitude and direction of its qualitative and quantitative improvement by studying the response to selection in advanced generations.

Materials and methods

Seed materials (Oryza sativa L.) derived through bulk-pedigree selection

Three high-yielding *Indica* rice cultivars (IR64, Swarna and Naveen) and another *Indica* cultivar (Sharbati) with good grain quality and high iron (Fe) and zinc (Zn) contents were crossed with two *Indica* rice donors for high GPC (ARC10075 and ARC10063). Around 2500 F_2 plants per population were raised. From 200 F_2 plant progenies, 42 were selected based on desirable plant type and higher yield potential from five different cross combinations (IR64 × ARC10075, IR64 × ARC10063, Swarna × ARC10063, Naveen × ARC10063 and Sharbati × ARC10063) (Table 1). They were bulked up to the F_4 generation, from which the pedigree method of selection was followed to develop 1710 F_8 selection lines in total from five breeding populations.

Field experiment

Seedlings generated from seeds of F_4 bulk progenies including their parents were transplanted into replicated plots in the wet season of 2011 at the experimental farm of the National Rice Research Institute, Cuttack, Odisha, India (20°5′N, 85°83′E and 29 m asl). The crop was raised with row–row distance of 20 cm and plant–plant distance of 15 cm. Nitrogen (N), phosphorus (P) and potassium (K) were supplied at 80, 60 and 40 kg/ha, respectively. Phosphorus (as single super phosphate) was applied as a basal dose, while N (as urea) and K (as muriate of potash) were applied in two equal doses (at 30 days after transplanting and at 50% flowering). The following observations were taken for ten randomly selected plants of each genotype: plant height (cm), length of panicle (cm), number of panicles and grain yield (g). F_8 lines including their parents were grown in replicated plots of 10 m^2 size under the same set of agronomic practices as above in the wet season of 2013 (Environment-1). The average yield (g/m^2) based on three sampling locations (1 m^2 area) within each plot was taken. In F_8, all plants within each line were found to be phenotypically similar and no segregation was detected. Seeds harvested in bulk from the F_8 generation of each line were raised (as the F_9 generation) for evaluation in the wet and dry seasons of 2014 (Environment-2 and Environment-3, respectively).

Estimation of grain protein and amylose content

Micro-Kjeldahl method

Grain protein content was determined by the standard micro-Kjeldahl method (Yoshida *et al.* 1976) by taking ten grains (12–14% moisture content) of brown rice (grains devoid of husk, but with the brown bran layer intact, which is normally removed during polishing). The GPC was calculated by multiplying the N content (%) by 5.95.

Using near-infrared spectroscopy

The near-infrared spectroscopy (NIR) (model NIRS DS2500, FOSS Analytical, Sweden) was calibrated for GPC and amylose content using brown rice of 170 genotypes including the germplasm and breeding lines that varied in GPC from 6.7 to 15.8% as measured by the micro-Kjeldahlmethod. The software package WinISI-III Project Manager v.1.50 e was used to calibrate the instrument. Prediction equations for these parameters were developed by modified partial least squares (mPLS) regression and evaluated by external validation. The 1,6.6,1 mathematical model for amylase content and 1,4,4,1 model for protein content were identified best and were fitted to NIRS (Bagchi *et al.* 2015) for the prediction of amylose and protein content in grains of F_8 lines. The model for GPC in brown rice proved fairly accurate. It was observed that the data points specifying predicted GPC were distributed randomly around the lines taken for validation (Fig. 1), shown by their high R^2 (0.96) and R values (0.97).

Extraction and quantification of storage proteins

Extraction of rice proteins was performed as described by Ju *et al.* (2001) with slight modification (Krishnan & Okita 1986). Rice flour (6–7 g) was defatted with n-hexane. The defatted flour (5 g) was extracted by stirring in 20 ml of distilled water at room temperature (RT) for 2 h to obtain the albumin fraction and centrifuged at 3000 ***g*** for 30 min. The residue was re-extracted with 20 ml of 5% sodium chloride (NaCl) for 2 h at room temperature to obtain the globulin fraction followed by centrifugation at 3000 ***g*** for 30 min. The residue was extracted for prolamins with 15 ml of 70% ethanol for 2 h followed by glutelin extraction with 37.5 ml of 0.2 M sodium borate buffer (pH 10) containing 0.5% sodium dodecyl sulphate (SDS) and 0.6% β-mercaptoethanol at RT for 2 h (Juliano 1980; Sugimoto *et al.* 1986). Each extraction was performed twice. The glutelins and prolamines from rice flour were extracted and partially purified, essentially as described by Krishnan & Okita (1986). The extracted proteins were freeze-dried and stored at −70 °C. The protein content of each fraction was measured according to Lowry *et al.* (1951).

Sodium dodecyl sulphate polyacrylamide gel electrophoresis analysis

The partially purified glutelins and prolamines were fractionated on preparative SDS-PAGE gels. Glutelin solutions dissolved in extraction buffer were mixed with the same volume of 2 × SDS sample buffer (100 mMTris–HCl buffer (pH 6.8) 4% (w/v) SDS, 20% (v/v) glycerol, 10% (v/v) 2-mercaptoethanol and 0.02% bromophenol blue). The SDS-PAGE was carried out according to standard protocol (Sambrook & Russell 2001) with 5% stacking and 12% resolving Polyacrylamide gel. Electrophoresis was carried out using a Bio Rad Mini-PROTEAN 3 Electrophoresis Cell (Bio

https://doi.org/10.1017/S0021859618000230 Published online by Cambridge University Press

https://doi.org/10.1017/S0021859618000230 Published online by Cambridge University Press

Table 1. Yield, yield contributing traits, grain protein content (GPC) and protein yield of parents and F_4 bulk population

Genotype no.	Designation	Type of materials/parentage	Maturity duration (days)	Plant height (cm)	Panicle length (cm)	Number of panicles/plant	GPC (%)[a]	Plant yield (g)	Protein yield/ plant (g)
1	ARC10075	Germplasm line with high GPC	150	174	30.6	10	11	20	2.2
2	ARC10063	Germplasm line with high GPC	130	164	29.3	8	12.02	18	2.16
3	Swarna	High yielding cultivar for rainfed ecosystem	140	96	27.3	8	8.29	28	2.32
4	IR64	High yielding cultivar for irrigated ecosystem	132	120	24.3	9	8.21	26	2.13
5	Naveen	High yielding cultivar for rainfed and irrigated ecosystem	130	122	25	10	8.3	26	2.16
6	Sharbati	Quality rice with high Fe content	144	128	26	10	7.75	20	1.55
7	CPL-A-F_4-1	IR64/ARC10075	150	127	29.3	16	8.98	28	2.51
8	CPL-A-F_4-2		150	136	28	12	8.33	21	1.75
9	CPL-A-F_4-3		138	153	36	14	11.72	22	2.58
10	CPL-A-F_4-4		138	142	25.6	25	7.31	30	2.19
11	CPL-A-F_4-5		138	133	31	18	9.85	29	2.86
12	CPL-A-F_4-6		137	128	23.3	9	10.66	15	1.6
13	CPL-A-F_4-7		138	141	31	20	11.54	28	3.23
14	CPL-A-F_4-8		150	126	28.3	17	11.02	27	2.98
15	CPL-A-F_4-9		144	149	26.3	19	9.52	28	2.67
16	CPL-B-F_4-1	IR64/ARC10063	143	114	28.3	22	10.6	30	3.18
17	CPL-B-F_4-2		120	120	28.3	21	11.7	31	3.63
18	CPL-B-F_4-3		150	86	27	26	11.45	20	2.29
19	CPL-B-F_4-4		141	88	25.3	14	11.7	15	1.76
20	CPL-B-F_4-5		120	107	25.3	21	11.08	22	2.44
21	CPL-B-F_4-6		147	84	24	16	12.5	16	2
22	CPL-B-F_4-7		144	95	22.3	9	12.39	23	2.85
23	CPL-B-F_4-8		150	132	25.6	9	10.31	15	1.55
24	CPL-B-F_4-9		149	95	26	28	9.6	24	2.3
25	CPL-B-F_4-10		127	112	28	15	9.31	18	1.68
26	CPL-B-F_4-11		127	126	24.3	13	8.68	17	1.48
27	CPL-C-F_4-1	Swarna/ARC10063	150	106	27	15	12.27	22	2.7
28	CPL-C-F_4-2		150	106	27	11	10.52	24	2.52
29	CPL-C-F_4-3		122	118	31	9	10.87	23	2.5
30	CPL-C-F_4-4		150	109	29	13	8.89	35	3.11
31	CPL-C-F_4-5		150	114	27.3	11	9.62	18	1.73

(*Continued*)

https://doi.org/10.1017/S0021859618000230 Published online by Cambridge University Press

Table 1. (*Continued.*)

Genotype no.	Designation	Type of materials/parentage	Maturity duration (days)	Plant height (cm)	Panicle length (cm)	Number of panicles/plant	GPC (%)[a]	Plant yield (g)	Protein yield/ plant (g)
32	CPL-C-F_4-6		150	109	29.3	9	11.5	27	3.11
33	CPL-D-F_4-1	Naveen/ARC10063	150	100	26.5	19	10.87	28	3.04
34	CPL-D-F_4-2		150	108	28	14	11.54	25	2.89
35	CPL-D-F_4-3		150	95	26.3	16	12.37	20	2.47
36	CPL-D-F_4-4		150	111	27	19	12.04	28	3.37
37	CPL-D-F_4-5		150	110	28.3	28	10.52	32	3.37
38	CPL-D-F_4-6		120	113	26	28	8.85	31	2.74
39	CPL-D-F_4-7		150	116	29.3	23	11.62	30	3.49
40	CPL-D-F_4-8		150	123	26.3	8	12.43	19	2.36
41	CPL-D-F_4-9		150	116	29.3	23	12.5	31	3.88
42	CPL-D-F_4-10		150	110	29	20	10.41	30	3.12
43	CPL-E-F_4-1	Sharbati/ARC10063	141	118	28.6	18	9.89	21	2.08
44	CPL-E-F_4-2		120	131	26	14	10.37	18	1.87
45	CPL-E-F_4-3		120	117	27	13	12.41	17	2.11
46	CPL-E-F_4-4		141	104	27	9	10.68	14	1.5
47	CPL-E-F_4-5		147	119	29	20	11.31	18	2.04
48	CPL-E-F_4-6		120	110	30	13	10.95	18	1.97
	Mean		139	116.7	27	15	11	24	2.48
	SEM (±)		4.0	0.97	2.3	3.9	1.2	3.6	0.08
	Range		110–150	84–174	22.3–36	41881	7.16–14.18	1.82–6.76	1.82–6.77
	Heritability (H^2)		0.84	0.88	0.78	0.63	0.59	0.69	0.8

[a]GPC at 12–14% moisture.

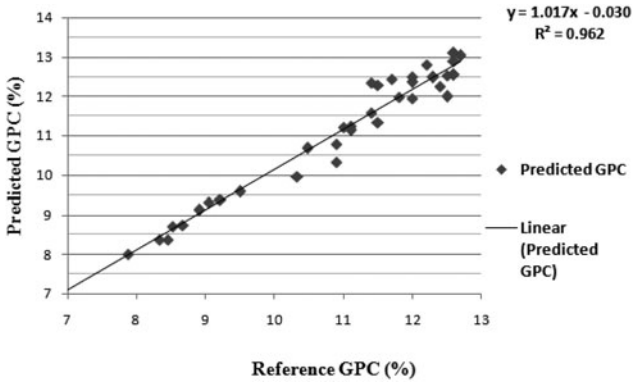

Fig. 1. Predicted against reference grain protein content (GPC).

Rad, USA). Peptide pattern in the gel was visualized through staining with Coomassie brilliant blue and photograph was taken in a gel documentation system.

Detection of alkali spreading value

Gelatinization temperature is estimated by the extent of alkali spreading. The degree of spreading is measured using a seven-point scale ranging from 1 (least spread) to 7 (highest spread): The average score denoted the alkali spreading value of a line. After soaking six milled rice grains of each genotype in 10 ml 1.7% KOH for 23 h at 30 °C, the degree of spreading and clearing values of individual grains were scored. Grain appearance and disintegration were rated visually after incubation. A rating of 5.5–7.0 was classified as low GT (55–69.5 °C); 3.5–5.4 as intermediate (70–74 °C); 2.6–3.4 as intermediate-high and 1.0–2.5 as high (74.5–80 °C) (Juliano 1985).

Statistical analysis

Analysis of variance (ANOVA) and correlation analysis for grain yield and other traits were carried out using standard procedures. The selection differential (S), which is the mean after selection minus the mean before selection, and the response to selection (R), which is the mean of the offspring distribution minus the mean of the distribution before selection, were estimated for the breeding population derived through modified bulk-pedigree selection. In predicting the response to selection for GPC, protein yield and grain yield in all crosses, R was taken as H^2S where H^2 stands for the broad-sense heritability of these traits, defined as the ratio of genotypic and phenotypic variance on an early-mean basis (Falconer & Mackay 1996). The significance of the mean differences of values between F_4 and F_8 generations was estimated using the t-test (Satterthwaite method). Principal component analysis (PCA) was performed to uncover the pattern of the data matrix for determining selection criteria and identifying elite genotypes. Eigen vectors and principal components were calculated and a 2-D biplot was generated between PC1 and PC2 by SAS enterprise guide 4.3 (www.support.sas.com/publishing).

A genotype plus genotype × environment (GGE) biplot analysis was done further to explain the source of variation for genotype (G) and genotype × environment (GE) (Yan *et al.* 2007). A GGE biplot symmetric view was used in the current study to explain the 'which-won-where' patterns for genotypes and environments.

Results

Analysis of F_4 bulk population

Significant variation was observed in yield, the traits contributing to yield, GPC and protein yield from each plant (Table 1) in the F_4 population. The population mean of these derived lines for GPC (10.73%) was significantly ($P < 0.05$) greater than the mean of the low-protein high-yielding parents (8.51%). The mean protein yield of each plant (2.5 g) was also significantly greater in the breeding population as a whole than that of the high-yielding parental lines alone (1.97 g). A moderate variability among genotypes as seen from the genotypic (15.2%) and phenotypic (15.75%) variances (data not presented), as well as a wide range of values (7.31–12.5%) coupled with moderate heritability (59%) and genetic advance (30.27), offer good scope for further improvement of this trait.

Association analysis revealed that protein yield ($r = 0.83$), number of panicles ($r = 0.50$) and panicle length ($r = 0.28$) were correlated significantly ($P < 0.05$) and positively with grain yield, whereas GPC was correlated significantly ($P < 0.05$) and negatively with grain yield ($r = -0.21$), but positively with panicle length ($r = 0.17$) (Table 2).

In PCA of the F_4 population, the first two principal components explained 74.11% of the total variability among the breeding lines. The positive effect of protein yield (0.61), plant yield (0.57) and panicle number (0.44) contributed significantly

Table 2. Correlation matrix for grain yield and yield attributing traits in F_4 population

Parameters	Maturity duration	Plant height (cm)	Panicle length (cm)	Number of panicles/plant	GPC (%)	Protein yield/plant (g)
Maturity duration	1.00					
Plant height (cm)	−0.15	1.00				
Panicle length (cm)	0.06	0.41	1.00			
Number of panicles/plant	0.07	−0.19	0.10	1.00		
GPC (%)	0.14	−0.19	0.17	0.02	1.00	
Protein yield/plant (g)	0.24	−0.09	0.37	0.49	0.36	1.00
Plant yield (g)	0.16*	0.01	0.28*	0.50**	−0.21*	0.83**

GPC, grain protein content.
*Significant at $P < 0.05$ level, **Significant at $P < 0.01$ level.

https://doi.org/10.1017/S0021859618000230 Published online by Cambridge University Press

($P < 0.01$) to the first principal component (PC1), and the negative effect of plant yield (−0.34) and panicle number (−0.23) and the positive effect of GPC (0.83) contributed to PC2. The Eigen vector size also revealed that along with grain yield and GPC, the number of panicles contributed significantly ($P < 0.01$) to the positional distribution of genotypes in the biplot based on PCA (Fig. 2). Therefore, in subsequent pedigree selection cycles, the number of panicles/plant was used as the selection index for yield in the current breeding programme for improving GPC. The bulk breeding lines (Table 1) belonged to the quarter in biplot (Fig. 2) with only the positive effect of PC1, governed mainly by protein yield, and the positive effect of PC2, governed mainly by the GPC, will be preferred from an economic point of view. Protein yield in many F_4 lines was significantly ($P < 0.01$) higher than that in the high-yielding cultivars Swarna, Naveen and IR 64 (Table 1). Lines such as CPL-D-F_4-9 (12.5%, 3.88 g), CPL-D-F_4-4 (12.1%, 3.37 g) and many others recorded significantly ($P < 0.01$) higher GPC and protein yield than Swarna (8.29%, 2.32 g) (Table 1). Cross combinations such as IR64/ARC10063 (CPL-B) and Naveen/ARC10063 (CPL-D) were identified as having a high potential because they proved more capable of delivering a larger number of transgressive segregants with high GPC than the other cross combinations.

Apparent amylose and grain protein content of F_8 grains using near-infrared spectroscopy

The GPC of 1710 breeding lines and six parents in the wet season of 2013 (Environment-1) was estimated using NIRS. Average GPC and protein yield were 11.21% and 2.76 g, respectively (Table 3). The following lines recorded high GPC and protein yield/plant: CPL-D-F_8-9 (13.01%, 5.01 g), CPL-D-F_8-19 (12.8%, 5.2 g) and CPL-A-F_8-67 (12.72%, 4.32 g). Apparent amylose content of the F_8 population varied from 19.32 to 26.7% with a mean value of 22.65%. Most of the F_8 lines contained intermediate levels (20–25%) of amylose, which was desirable for acceptable cooking quality.

Improvement of grain protein content and protein yield in F_8

As in the F_4 generation, the mean value for GPC and protein yield of lines derived from IR64/ARC10063 (CPL-B-F_8) and Naveen/ARC10063 (CPL-D-F_8) were higher than those derived from other crosses and also greater than the mean values of these two crosses in the preceding generation. As expected from the results of the F_4 generation, these two crosses provided a large number of transgressive lines for GPC in the F_8 generation. Grain protein content significantly ($P < 0.05$) improved in the F_8 generation over the F_4 generation, as shown by the *t*-test (Satterthwaite method) (Fig. 3(*a*)). Although grain yield was not significantly greater, protein yield in the F_8 generation was significantly higher ($P < 0.05$) than the mean of the whole F_4 population (Fig. 3(*b*)). However, this pattern was not uniform when individual crosses were considered: GPC was only significantly ($P < 0.05$) higher in CPL-D and CPL-A. Protein yield only increased significantly ($P < 0.05$) in CPL-E, as it recorded greater improvement in grain yield (Fig. 4). The extent of improvement in GPC in the F_8 generation over the parental population was significant ($P < 0.05$) in all the crosses individually and for the breeding population as a whole. On the other hand, the extent of improvement over the F_4 generation was significant ($P < 0.05$) only in CPL-A. Similarly, for protein yield, significant ($P < 0.05$) improvement in the F_4 bulk population over the parental population was evident in all the crosses and in the breeding population as a whole. With the exceptions of CPL-C and CPL-D, the pattern was repeated in the pedigree selection lines in F_8. Apart from CPL-C and

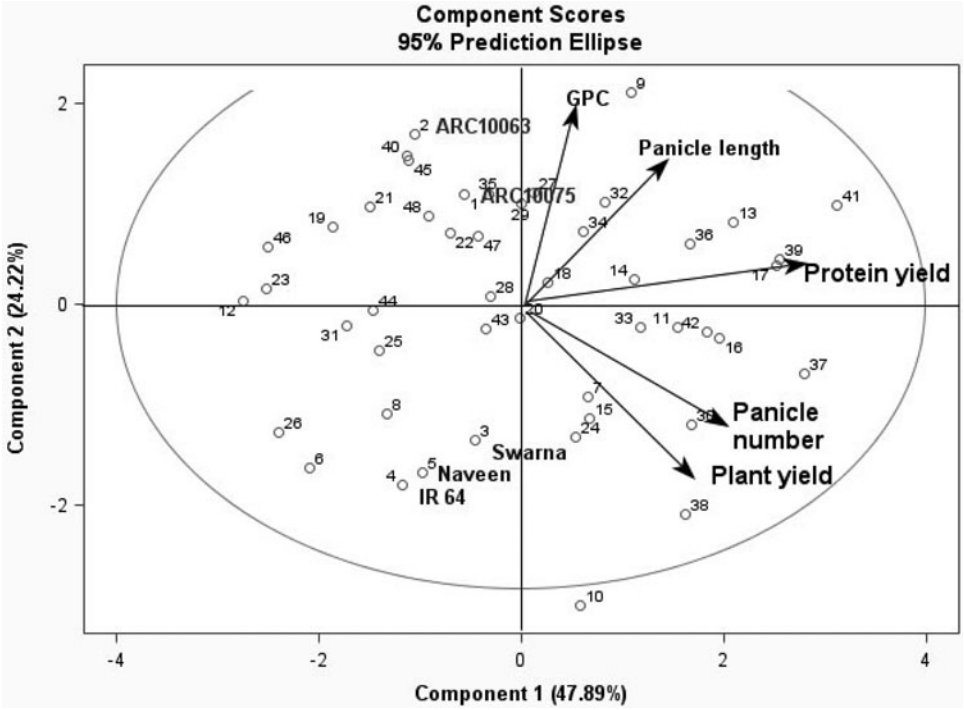

Fig. 2. 2-D plot presentation of traits and genotypes derived through principal component analysis (PCA) of yield, yield contributing traits, grain protein content (GPC) and protein yield of parents and bulk population (F_4) (Listed in Table 1 as Genotype number).

https://doi.org/10.1017/S0021859618000230 Published online by Cambridge University Press

Table 3. Mean grain protein content (GPC), amylose content, plant yield and protein yield of parents and breeding population of rice in *wet season* 2013

Variety/ population	Parent/cross combination	Number of lines	Mean amylose content (%)	Mean GPC (%)	Mean plant yield (g)	Mean protein yield/ plant (g)
Naveen	Parent	1	21.51	9.12	24.50	2.23
Swarna	Parent	1	22.79	8.56	27.60	2.45
IR 64	Parent	1	19.32	7.77	23.50	1.83
ARC10075	Parent	1	21.63	11.40	19.80	2.26
ARC10063	Parent	1	22.24	12.30	18.50	2.28
Sharbati	Parent	1	21.80	8.30	18.90	1.57
CPL-A-F_8	IR64/ARC10075	720	23.16	11.35	24.57	2.80
CPL-B-F_8	IR64/ARC10063	240	21.29	11.22	22.71	2.55
CPL-C-F_8	Swarna/ARC10063	380	23.87	10.65	24.23	2.58
CPL-D-F_8	Naveen/ARC10063	230	21.26	11.71	27.60	3.25
CPL-E-F_8	Sharbati/ARC10063	140	21.87	11.02	23.31	2.60
F_8 Population		1710	22.65	11.13	24.46	2.73
SD			1.64	1.23	6.45	0.81

CPL-B, plant yield was also significantly ($P < 0.05$) greater in the F_8 population than in the parental population, whereas this improvement over the F_4 bulk population was significant ($P < 0.05$) only in CPL-E over the parental population (Table 4).

Selection differential and predicated versus observed response to selection

For GPC, although the selection differential (S) was significant ($P < 0.05$) for all the crosses and for the population as a whole, the observed response to selection was significant ($P < 0.05$) only in CPL-A and in the population as a whole. For grain yield, S was significant ($P < 0.05$) for CPL-E and CPL-D but the response to selection (R) was significant ($P < 0.05$) only in CPL-E. On the other hand, S for protein yield was significant ($P < 0.05$) for all the crosses and for the population as a whole and, as expected, R was significant ($P < 0.05$) for CPL-E and for the population as a whole (Table 5). The predicted response to selection for GPC, grain yield and protein yield individually for each cross and for the population as a whole was derived from the selection differential and the heritability (0.59, 0.69 and 0.80%, respectively) of these traits. The predicted and observed responses to selection for both GPC and protein yield were almost equal in the population as a whole (Fig. 5) but, except for a few cases, were significantly different ($P < 0.05$) for GPC and protein yield (Table 5).

Genotype plus genotype × environment biplot analysis of multi-environmental data

In the wet season of 2013 (Environment-1), 1710 F_8 lines were evaluated and plant yield and GPC from five randomly selected plants from each line were measured. F_9 lines derived from the corresponding F_8 lines were further evaluated in two environments, the wet season of 2014 (Environment-2) and dry season of 2014 (Environment-3) for yield and GPC. Significant differences ($P < 0.05$) for genotype, environments and genotype × environment interaction (GEI) were found. The GGE-biplot (Fig. 6) displays the 'which-won-where' pattern of rice genotypes with high mean GPC (Table 6) in the multi-environment trial (MET). In this biplot, a polygon was formed by connecting the vertex genotypes such as 891 (CPL-D-F_8-891), 347

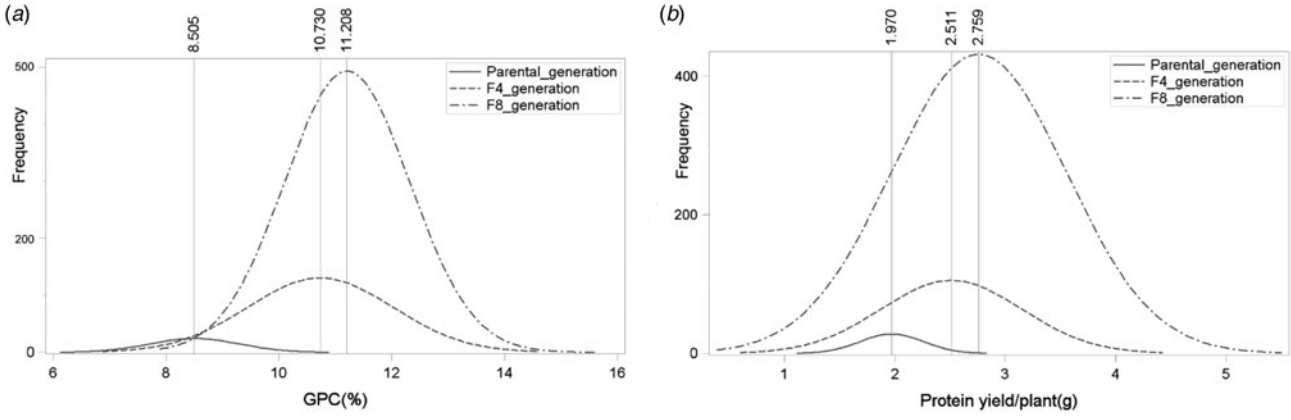

Fig. 3. Distribution of parents, F_4 and F_8 generation and their means for grain protein content (GPC) (*a*) and protein yield (*b*) for the whole breeding population.

https://doi.org/10.1017/S0021859618000230 Published online by Cambridge University Press

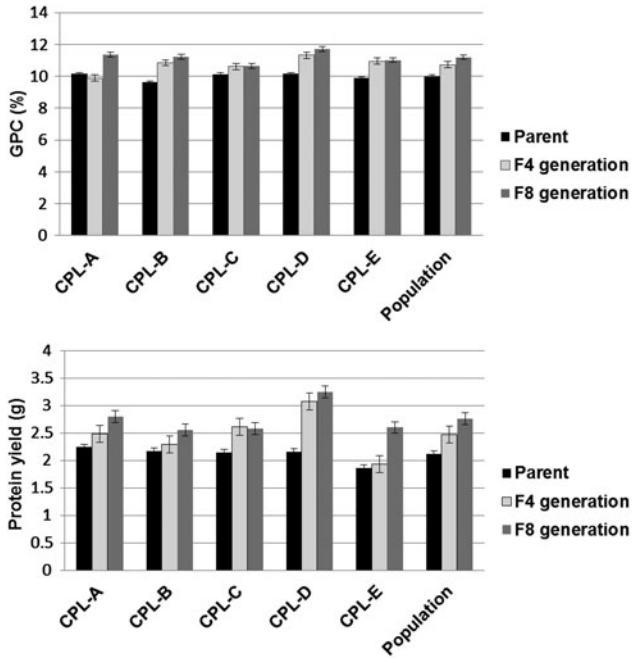

Fig. 4. Mean grain protein content (GPC) (*a*) and protein yield (*b*) of parents, F_4 and F_8 population of individual crosses and breeding population as a whole.

(CPL-B-F_8-347), 137 (CPL-E-F_8-137) IR 64 and 733 (CPL-C-F_8-733) with straight lines and the rest of the high mean genotypes placed within the polygon. These vertex genotypes were the best or worse genotypes in any of the environments. Test environments fell into two sections. The first section contained E1 and E3 and the second section contained E2. High protein genotypes were mostly placed inside these two sections. Another two vertex genotypes, IR 64 and 733 (CPL-C-F_8-733), were lower with respect to protein content.

Similarly, another polygon view (Fig. 7) displays the which-won-where' pattern of rice genotypes with high protein yield in the MET. The biplot was divided into seven sections. The vertex genotype 1091 (CPL-A-F_8-1091) was the best performer and situated in the section where two of the test environments (E1 and E3) fell. Another test environment fell within a section where the vertex genotype 1049 (CPL-A-F_8-1049) had high protein yield. On the other hand, the two lowest protein yield cultivars, IR 64 and ARC 10075, were placed in the vertex of the polygon. The stability of high GPC and PROY genotypes were evaluated. An average environment expressed by the average PC1 and PC2 scores of all environments jointly explained 85.7% variation. This is represented by a small circle in the GGE biplot (Fig. 8). An average environment axis (AEA) line with the arrow pointing to the greater genotype main effect passes through this average environment (small circle) and serves as the abscissa of the AEC. The AEC ordinate is indicated by double arrows, and either direction away from the biplot origin indicates greater GEI effect and reduced stability. A longer projection to the AEC ordinate, regardless of the direction, represents a greater tendency of the GEI of genotype, which means it is more variable and less stable across environments or vice versa. High mean genotypes for GPC and PROY were found to be more or less stable across environments. Among the selected high GPC and PROY genotypes, 972 (CPL-C-F_8-972) and 966 (CPL-C-F_8-966) for GPC and 884 (CPL-D-F_8-884) and 824 (CPL-D-F_8-824) for PROY were found to be highly stable.

Identification of high yielding lines with high grain protein content and good cooking quality

A few high mean lines for GPC and PROY were identified stable through GGE biplot analysis of multi-environment data. Some had desirable agronomic traits and good cooking quality with desired amylose and alkali spreading value (Table 7). Some of the lines that showed high protein yield (g/m^2), high GPC (%) and good cooking quality (with desirable amylose content and alkali spreading value) include: CPL-D-F_8-824 (74.75, 12.55, 23.19, 4), CPL-D-F_8-905 (65.41, 11.88, 21.93, 4), CPL-A-F_8-1049 (49.22, 13.07, 21.76, 5), CPL-C-F_8-972 (46.31, 12.30, 23.83, 3).

Fractionation of grain protein and sodium dodecyl sulphate polyacrylamide gel electrophoresis of some selected lines with high grain protein content and protein yield

Selected high protein lines (Table 7) along with high protein donors in the current breeding programme, ARC10063 and high yielding parent, Swarna were taken and 10% milled rice grains (Fig. 9) were subjected to fractionation of soluble proteins. Significant ($P < 0.05$) variation was observed among all genotypes for the four fractions: albumins, globulins, prolamins and

Table 4. Improvement of grain protein content (GPC), plant yield and protein yield in F_8 population over F_4 bulk population and parental mean

		Percentage of improvement					
		GPC (%)		Plant yield (g)		Protein yield (g)	
F_8 lines	Cross	Over F_4 bulk population mean	Over parental mean	Over F_4 bulk population mean	Over parental mean	Over F_4 bulk population mean	Over Parental mean
CPL-A-F_8	IR64/ARC10075	14.88	19.47	−3.00	10.03	12.90	34.62
CPL-B-F_8	IR64/ARC10063	3.41	16.51	8.14	5.63	11.35	26.24
CPL-C-F_8	Swarna/ARC10063	0.38	5.86	−2.42	5.21	−1.15	14.16
CPL-D-F_8	Naveen/ARC10063	3.45	12.49	0.73	26.90	5.86	45.74
CPL-E-F_8	Sarbati/ARC10063	0.73	14.32	31.92	23.66	34.72	43.65
F_8 Population		3.92	15.46	4.95	10.53	10.26	28.17

https://doi.org/10.1017/S0021859618000230 Published online by Cambridge University Press

Table 5. Mean, range, selection differential (S), and predicted and observed response to selection (R) of F_8 breeding population derived from bulk-pedigree method of selection

Cross	Designation	No. lines	Grain protein content (%)					Plant yield (g)					Protein yield of each plant (g)				
			Mean	Range	S	Observed R	Predicted R	Mean	Range	S	Observed R	Predicted R	Mean	Range	S	Observed R	Predicted R
IR64/ ARC10075	CPL-A-F_8	720	11.35	8.78–15.11	1.74**	1.47**	0.36	24.57	12.5–40.6	1.57	−0.76	0.75	2.80	1.23–4.72	0.63*	0.32	0.27
IR64/ ARC10063	CPL-B-F_8	240	11.22	8.72–13.3	1.1**	0.37	0.23	22.71	9.2–39	0.71	1.71	0.34	2.55	1.11–4.16	0.4*	0.26	0.17
Swarna/ ARC10063	CPL-C-F_8	380	10.65	9.25–12.92	0.49*	0.04	0.10	24.23	17.5–35.6	1.23	−0.60	0.58	2.58	1.72–3.82	0.34*	−0.03	0.14
Naveen/ ARC10063	CPL-D-F_8	230	11.71	10.82–13.01	1.55**	0.39	0.32	27.6	18.6–40.6	5.6*	0.20	2.66	3.25	2.06–5.2	1.09**	0.18	0.46
Sarbati/ ARC10063	CPL-E-F_8	140	11.02	9.6–13.25	1.13**	0.08	0.23	23.31	13.4–35.7	4.31*	5.64*	2.05	2.60	1.33–4.18	0.74**	0.67**	0.31
Total	F_8 population	1710	11.21	8.72–15.11	1.95**	0.48*	0.40	24.54	9.2–40.6	1.54	1.02	0.73	2.76	1.11–5.2	0.67**	0.25*	0.28

*Significant at $P < 0.05$ level, **Significant at $P < 0.01$ level.

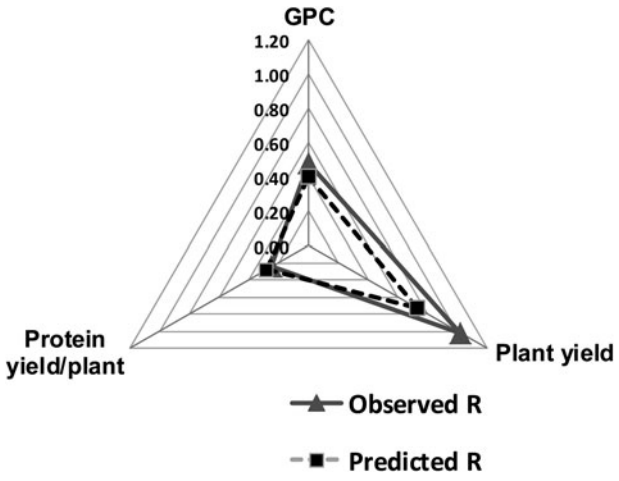

Fig. 5. Predicted and observed response to selection (R) of grain protein content (GPC), plant yield and protein yield.

glutelins. The mean contents of albumin, globulin, glutelin and prolamin fractions of these genotypes in polished rice were 0.19, 0.30, 5.35 and 0.1%, respectively (Table 8). Contents of albumin, globulin, glutelin and prolamin among the genotypes ranged from 0.15 to 0.26%, 0.18 to 0.42%, 2.72 to 7.05% and 0.08 to 0.11%, respectively. All high protein lines had significantly ($P < 0.001$) higher glutelin content than Swarna. The ratio of prolamin to glutelin fractions ranged from 0.01 to 0.03. All high protein lines had either lower or similar values in prolamin/glutelin ratio than Swarna.

The partially purified glutelins and prolamins in all these genotypes were separated on preparative SDS-PAGE gels. The basic sub-unit, α-glutelin was found in an average 29 kD region, while the average molecular weight of β-glutelin unit was 21 kD in these genotypes. Another prolamin band was observed in these genotypes at 13-14 kD region. The differences were observed only in the intensity of banding pattern (Fig. 10). Higher expression in both α- and β-glutelin was observed in all high protein lines except CPL-A-F_8-1045 than the high yielding cultivar Swarna. This banding pattern mostly correlated with the glutelin content observed through fractionation of soluble protein.

Discussion

Assessing the scope for improvement of grain protein content in rice

Rice, being the main source of calories as well as of proteins for billions of people, is ideal for enhancing GPC through hybridization and selection with the aim of providing greater protein in this staple food. Any success on this front will certainly have a positive impact on the health of the poor and the malnourished in developing countries (Mahmoud *et al.* 2008). Early researchers pointed out that high doses of N improve GPC but also increase the degree of translucency significantly (Blumenthal *et al.* 2008). Therefore, lines are required that give high yields and high GPC under standard N management. It has been observed that N translocation from the vegetative tissues plays an important role in the nutrition of rice grains. In addition, it has also been reported that N translocation was positively associated with total accumulation of N in plants at anthesis and added to

https://doi.org/10.1017/S0021859618000230 Published online by Cambridge University Press

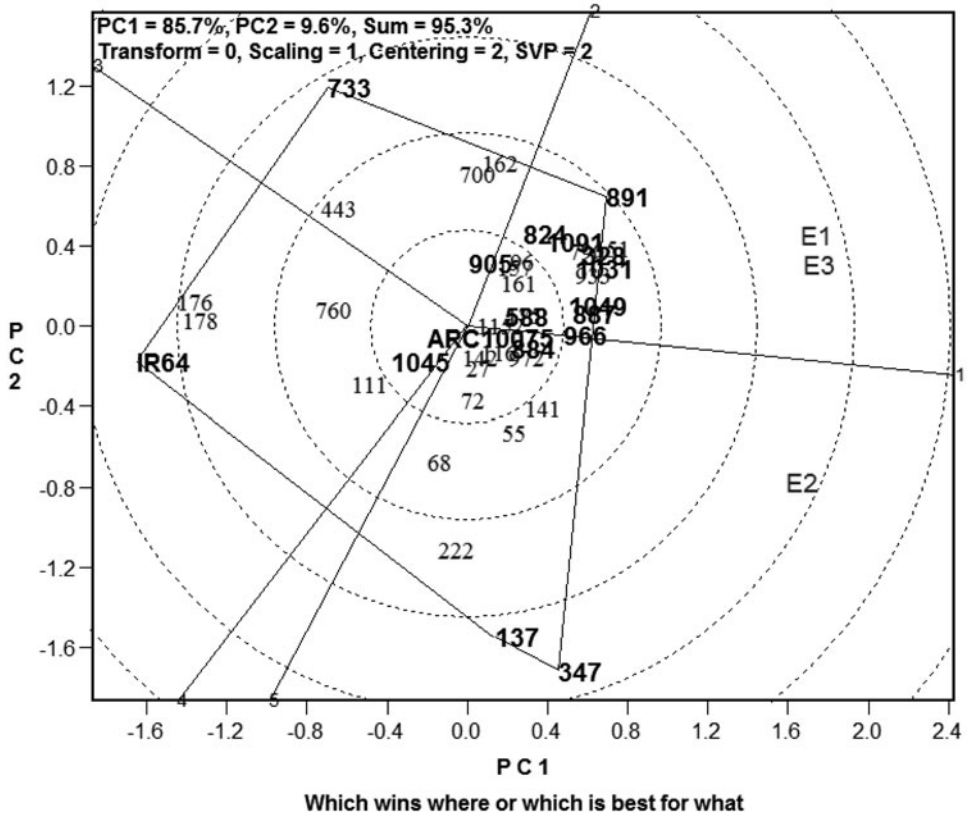

Fig. 6. Polygon in multi-environment trials view displaying 'which won where' pattern of genotype plus genotype × environment (GGE) biplot of selected high grain protein content (GPC) genotypes (listed in Table 6).

grain N at maturity by 30–60%, depending on the genotype (Ntanos & Koutroubas 2002). Using genetic resources with higher potentiality for N accumulation, GPC has been improved in many cereals (Vasal 2002). For improvement of GPC, two landraces with high GPC (>11%) were utilized by Chattopadhyay *et al.* (2011). In spite of linkage with undesirable traits, the transfer of desired alleles with positive impacts on grain quality from agronomically poor landraces has been proved feasible previously (Aluko *et al.* 2004; Li *et al.* 2004) and also in the present breeding programme, as 17 F_4 breeding lines were found with trangressive segregation for GPC (>12%). Many other lines also contained significantly higher GPC (>10%) than the parental population. All these lines were upgraded to pre-breeding lines for high GPC and to advanced elite high-yielding lines with high GPC. However, the economic viability of the breeding programme lies not only in the greater protein content of 100 g grain volume or of a single grain but also in total protein yield per unit area. Accordingly, the selection and fixation of breeding lines in subsequent generations focused on both high GPC and protein yield.

Bulk-pedigree breeding coupled with high throughput selection for improvement of grain protein content

As GPC is a polygenic trait (Mahmoud *et al.* 2008), it generally shows low heritability. The bulk-pedigree method is generally used for improving traits with low heritability. This is a modification of the bulk method, in which individual plants of the F_2 generation are harvested in bulk up to the F_4 generation, from which single-plant selections are made and subsequent generations handled the same way as in the pedigree method: Wynne & Gregory (1981), who pioneered this modification, believed that selection from F_5 onwards through the pedigree method would help in fixing the desirable genes. Although there is no direct effect of GPC on seed yield, the slightly negative association between the two traits indicated the possible risk of reduction of grain yield when selection is restricted solely to higher GPC. Therefore, the pedigree selection cycle of the present programme opted for simultaneous selection for higher GPC and a greater number of panicles per plant, which was the most important component of grain yield in the association analysis and further supported by PCA. The negative association was reduced to a significant level after three pedigree selection cycles in the present breeding programme and this facilitated the simultaneous improvement of these traits, as also reported by the International Rice Research Institute (IRRI 1976).

Breeding for grain quality requires a method that can determine the nutritional composition of a given substance quickly and accurately and yet allows large-scale screening. Near-infrared spectroscopy has been used extensively for rapid and non-destructive estimation of grain or kernel quality, especially of cereals, pulses and oilseeds. The method requires minimum sample preparation and is therefore cost-effective and saves both time and labour. A calibration model was developed for GPC and validated. A high correlation between predicted and reference values indicated that this calibration model could quantify the prediction of GPC of brown rice. In the present breeding programme, GPC and amylose content of a large number of breeding lines were predicted by NIRS.

https://doi.org/10.1017/S0021859618000230 Published online by Cambridge University Press

https://doi.org/10.1017/S0021859618000230 Published online by Cambridge University Press

Table 6. High mean F_8 lines for grain protein content (GPC), plant yield (PY) and protein yield (PROY) based on three environments (E1- *wet season* 2013, E2- *wet season* 2014, E3- *dry season* 2014)

		GPC				PY				PROY			
Genotype number	Designation	E1	E2	E3	Mean	E1	E2	E3	Mean	E1	E2	E3	Mean
222	CPL-B-F_8-222	10.68	12.52	11.40	11.53	30.50	27.33	26.30	28.04	3.26	3.42	3.00	3.23
27	CPL-D-F_8-27	11.22	11.75	12.64	11.87	38.90	30.20	28.30	32.47	4.36	3.55	3.58	3.83
328	CPL-C-F_8-328	13.29	12.53	13.85	13.22	25.60	26.43	22.54	24.86	3.40	3.31	3.12	3.28
824	CPL-D-F_8-824	12.26	11.65	13.74	12.55	33.30	36.50	30.43	33.41	4.08	4.25	4.18	4.17
891	CPL-D-F_8-891	13.39	12.42	14.83	13.55	26.20	25.32	23.81	25.11	3.51	3.14	3.53	3.39
905	CPL-D-F_8-905	11.79	11.22	12.64	11.88	40.00	30.50	34.85	35.12	4.72	3.42	4.41	4.18
1045	CPL-A-F_8-1045	10.63	10.88	11.37	10.96	32.60	31.60	36.80	33.67	3.47	3.44	4.18	3.70
443	CPL-C-F_8-443	10.70	9.23	10.20	10.04	39.10	36.61	40.52	38.74	4.18	3.38	4.13	3.90
55	CPL-A-F_8-55	11.83	12.61	12.40	12.28	24.80	23.59	27.40	25.26	2.93	2.97	3.40	3.10
884	CPL-D-F_8-884	12.72	12.32	11.98	12.34	34.00	31.40	33.20	32.87	4.32	3.87	3.98	4.06
68	CPL-A-F_8-68	10.59	11.86	11.80	11.42	30.50	32.00	30.60	31.03	3.23	3.80	3.61	3.55
700	CPL-C-F_8-700	11.93	10.59	12.76	11.76	36.40	38.40	40.30	38.37	4.34	4.07	5.14	4.52
887	CPL-D-F_8-887	13.29	12.80	13.18	13.09	27.20	30.40	29.60	29.07	3.61	3.89	3.90	3.80
72	CPL-A-F_8-72	11.77	12.00	11.50	11.76	30.40	37.20	33.60	33.73	3.58	4.46	3.86	3.97
733	CPL-C-F_8-733	10.96	8.57	10.83	10.12	38.00	42.50	40.50	40.33	4.16	3.64	4.39	4.06
760	CPL-D-F_8-760	10.49	9.78	9.65	9.97	39.00	40.50	31.70	37.07	4.09	3.96	3.06	3.70
790	CPL-D-F_8-790	12.67	12.28	14.40	13.12	29.20	30.80	27.60	29.20	3.70	3.78	3.97	3.82
933	CPL-C-F_8-933	13.30	12.60	13.50	13.13	22.70	26.40	25.00	24.70	3.02	3.33	3.38	3.24
96	CPL-A-F_8-96	12.73	11.78	12.50	12.34	29.50	27.90	30.20	29.20	3.76	3.29	3.78	3.61
111	CPL-B-F_8-111	10.82	10.65	9.74	10.40	30.10	33.30	42.90	35.43	3.26	3.55	4.18	3.66
116	CPL-B-F_8-116	11.49	11.85	12.83	12.06	24.30	26.40	25.90	25.53	2.79	3.13	3.32	3.08
972	CPL-C-F_8-972	12.69	12.34	11.86	12.30	30.10	33.90	37.40	33.80	3.82	4.18	4.44	4.15
137	CPL-E-F_8-137	11.21	13.70	11.73	12.21	31.50	30.70	28.60	30.27	3.53	4.21	3.35	3.70
966	CPL-C-F_8-966	12.69	12.75	13.64	13.03	28.00	29.80	25.80	27.87	3.55	3.80	3.52	3.62
588	CPL-C-F_8-588	11.52	11.85	13.74	12.37	37.20	40.00	40.40	39.20	4.29	4.74	5.55	4.86
141	CPL-E-F_8-141	11.64	12.65	13.53	12.61	25.30	28.90	31.60	28.60	2.94	3.66	4.28	3.63
142	CPL-E-F_8-142	11.42	11.67	12.32	11.80	37.20	41.80	36.70	38.57	4.25	4.88	4.52	4.55
1049	CPL-A-F_8-1049	13.01	12.67	13.53	13.07	38.50	40.00	34.00	37.50	5.01	5.07	4.60	4.89
347	CPL-B-F_8-347	11.60	14.60	12.81	13.00	27.00	29.80	24.60	27.13	3.13	4.35	3.15	3.54
151	CPL-E-F_8-151	12.95	12.56	14.75	13.42	26.70	20.40	22.60	23.23	3.46	2.56	3.33	3.12

(*Continued*)

Table 6. (*Continued.*)

Genotype number	Designation	GPC				PY				PROY			
		E1	E2	E3	Mean	E1	E2	E3	Mean	E1	E2	E3	Mean
1152	CPL-A-F_8-1152	11.56	11.65	12.79	12.00	33.30	30.20	27.50	30.33	3.85	3.52	3.52	3.63
1091	CPL-A-F_8-1091	12.80	12.00	13.59	12.80	40.60	33.20	35.40	36.40	5.20	3.98	4.81	4.66
157	CPL-E-F_8-157	11.52	11.48	13.80	12.27	34.80	30.50	27.50	30.93	4.01	3.50	3.80	3.77
161	CPL-E-F_8-161	11.72	11.64	13.53	12.30	35.70	30.59	32.63	32.97	4.18	3.56	4.41	4.05
162	CPL-E-F_8-162	12.67	10.86	12.40	11.98	31.00	34.70	30.29	32.00	3.93	3.77	3.76	3.82
1031	CPL-A-F_8-1031	13.25	12.53	13.62	13.13	28.50	26.57	29.32	28.13	3.78	3.33	3.99	3.70
174	ARC10075	11.40	11.20	11.45	11.35	20.00	22.00	19.60	20.53	2.28	2.46	2.24	2.33
175	ARC10063	12.30	12.00	12.63	12.31	18.00	21.50	20.00	19.83	2.21	2.58	2.53	2.44
176	Swarna	8.56	8.03	8.40	8.33	28.00	32.50	30.50	30.33	2.40	2.61	2.56	2.52
177	IR 64	7.77	7.87	8.00	7.88	26.00	25.60	26.80	26.13	2.02	2.01	2.14	2.06
178	Naveen	8.58	8.20	8.40	8.39	26.00	27.90	28.60	27.50	2.23	2.29	2.40	2.31

Assessment of breeding population based on response to selection for both grain protein content and grain yield

The gain in GPC and protein yield was noted in the F_8 generation as compared with that in the F_4 bulk population. Significant yield advantage was observed in CPL-E-F_8 over the F_4 bulk. Ali (2011) found that pedigree selection for two cycles showed a significant increase in yield over the bulk selection. However, in the current work, it was noted that mainly GPC and grain yield increased significantly through bulking of F_2 plants as shown in the F_4 bulk lines in most of the cross combinations and in the breeding population as a whole. This improvement was then stabilized through pedigree selection cycles. Hence, the modified bulk-pedigree method of selection proved effective in improving both yield and grain quality simultaneously.

Grain protein content and protein yield of individual plants in the breeding population as a whole also showed a substantial response to bulk-pedigree selection as shown by the significant *S* and *R* values. In addition, the population of individual crosses also showed a significant response for these traits. However, the observed and expected *R* values diverged greatly in individual crosses. This discrepancy arose probably because the estimates of the extent of heritability calculated by the components of variance method based on data from one generation and year were applied to the next generation. Under such circumstances, a large discrepancy between the actual gains obtained in the F_8 and those predicted in the F_4 generation is only to be expected by the genotype × year interaction (Frey & Horner 1955). However, the current results are in agreement with those obtained by Gaido *et al.* (2000), who used recurrent selection for improving grain yield. The actual response to selection for grain yield and protein yield in the cross CPL-E and that for GPC in the cross CPL-A was greater than the predicted response, indicating that dominant-gene effects are involved in the inheritance of these traits. Whereas Lal & Singh (2012) found that gene action for grain yield in rice was mainly non-additive, Singh & Singh (1982) found that gene action for protein content in rice consisted of both additive and non-additive components of variance.

The current study found that when the whole F_8 population was considered for analysis, the observed and expected *R* for GPC and protein yield were not significantly different. This finding serves to validate the method adopted in the breeding programme and proved once again that the bulk-pedigree method is effective on traits significantly influenced by GEI. As suggested by Frey & Horner (1955), a close agreement between the expected and actual gains indicates that the gene action involved in the selection is largely additive. In the present experiment, actual grain yield was greater than the predicted yield, a result that confirms the superiority of the dominant-gene effects (El-Ameen *et al.* 2013).

Evaluation of cooking quality of breeding lines with elevated grain protein content

Amylose content greatly influences the cooking characteristics of rice and other features related to grain quality. Amylose is also responsible for the way rice hardens on cooling (Allahgholipour *et al.* 2006). Protein is the second most important component in the endosperm, which affects both milling and cooking qualities of rice grain. Grain proteins inhibit the gelatinization of starch and disruption of protein structure during cooking makes the cooked grains stickier (Cameron & Wang 2005;

https://doi.org/10.1017/S0021859618000230 Published online by Cambridge University Press

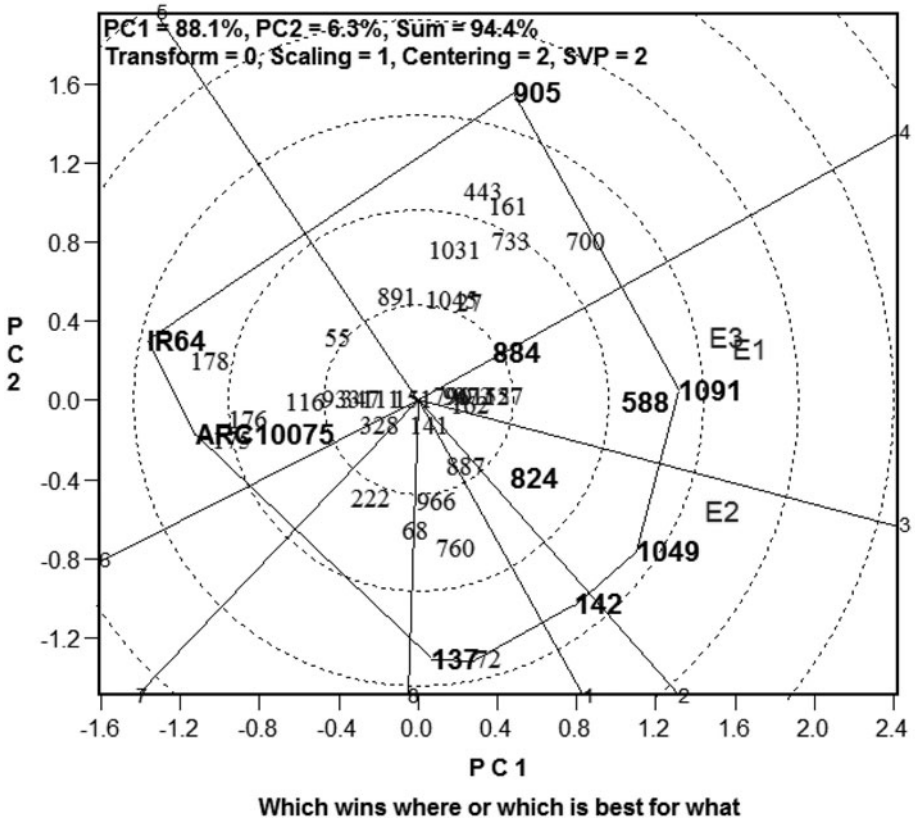

Fig. 7. Polygon view in multi-environment trials displaying 'which won where' pattern of genotype plus genotype-by-environment (GGE) biplot of selected high protein yield (PROY) genotypes (listed in Table 6).

Derycke *et al.* 2005). A negative correlation of grain protein and amylose content has been observed in the current study. This is in keeping with an early observation (Yang *et al.* 2004). The PCA also revealed that the Eigen vector for grain yield and amylose content occupied the same quarter, whereas GPC belonged to a different quarter in the biplot. In general, the anti-waxy gene increased protein content and lowered the amylose content of rice grains (Li *et al.* 2009). Therefore, selecting for

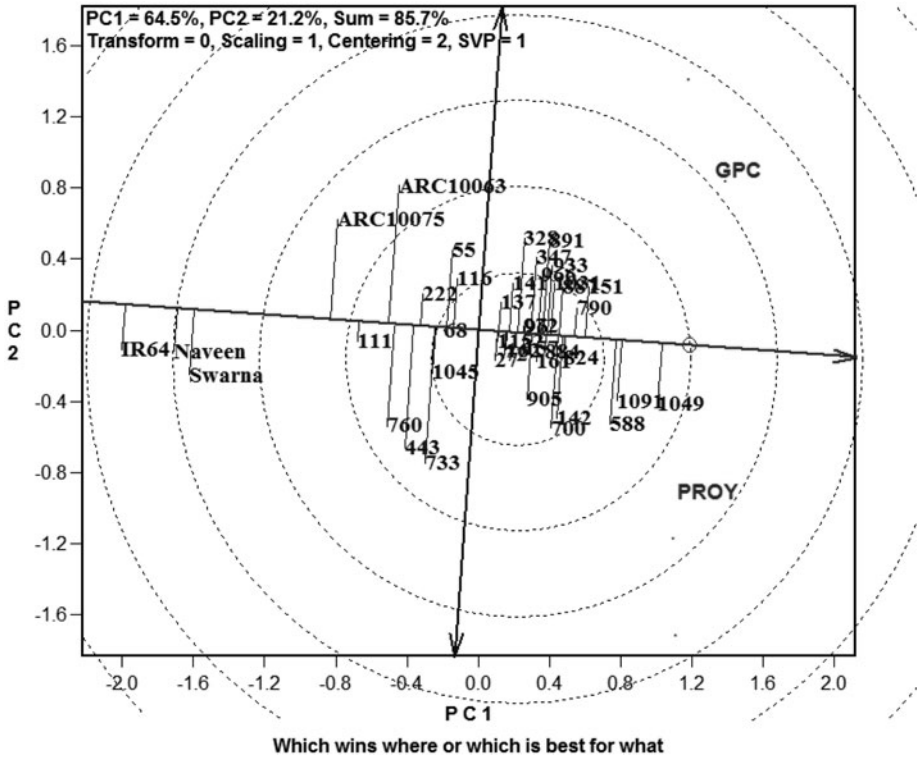

Fig. 8. Average environment coordination view of the GGE-biplot for the performance and stability of the high mean genotypes for grain protein content (GPC) and protein yield (PROY) (Listed in Table 6).

https://doi.org/10.1017/S0021859618000230 Published online by Cambridge University Press

Table 7. Grain protein content (GPC), yield and yield attributing traits and protein yield of selected high protein rice genotypes over the environments

Genotype	GPC	Plant height (cm)	Number of panicles/plant	Panicle length (cm)	Grain yield (g/m^2)	Protein yield (g/m^2)	Amylose content (%)	Alkali spreading value (ASV)
CPL-D-F_8-824	12.55	107.50	8.50	25.53	595.6	74.75	23.19	4
CPL-A-F_8-1045	10.96	114.70	8.23	25.72	540.5	59.24	22.23	4
CPL-D-F_8-891	13.55	133.80	6.60	28.15	402.7	54.57	24.06	4
CPL-D-F_8-905	11.88	125.70	8.73	26.55	550.6	65.41	21.93	4
CPL-D-F_8-884	12.34	130.60	7.11	26.46	366.8	45.26	22.83	5
CPL-D-F_8-887	13.09	135.30	9.42	28.23	359.4	47.05	22.65	5
CPL-C-F_8-972	12.30	98.70	12.90	27.13	376.5	46.31	23.83	3
CPL-C-F_8-966	13.03	115.70	7.45	24.67	343.4	44.75	24.17	3
CPL-C-F_8-588	12.37	90.40	6.82	22.91	364.8	45.13	21.94	5
CPL-A-F_8-1049	13.07	122.60	6.67	25.58	376.6	49.22	21.76	5
CPL-A-F_8-1091	12.8	125.60	6.93	24.16	362.6	46.41	21.08	5
CPL-A-F_8-1031	13.13	113.20	7.33	28.74	380.4	49.95	20.83	5
Swarna	8.33	101.60	10.22	24.72	470.3	39.18	23.8	4
ARC10063	12.31	160.60	6.75	27.62	298.6	36.76	21.5	5
General Mean	12.27	119.71	8.11	26.16	413.49	50.28	22.56	4.36
CD (5%)	0.34	4.8	0.6	0.6	53.4	4.2	0.9	0.4

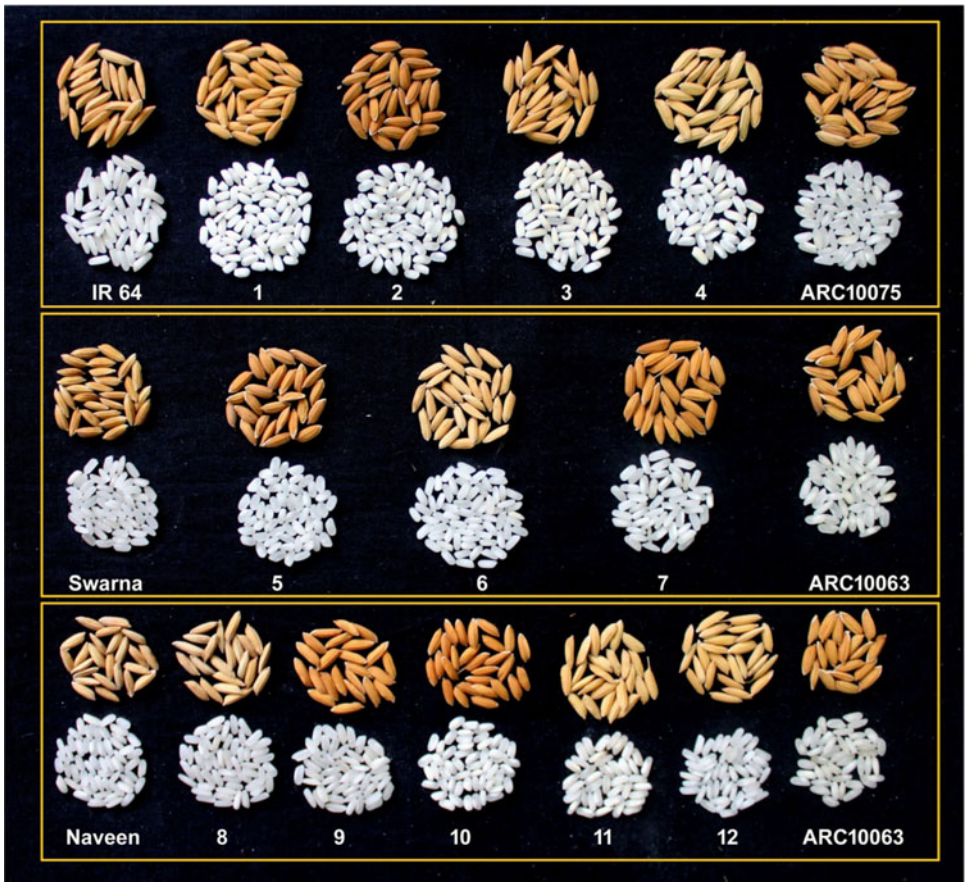

Fig. 9. Paddy and polished rice of parents and their high protein offsprings, viz. 1: CPL-A-F_8-1045, 2: CPL-A-F_8-1049, 3: CPL-A-F_8-1091, 4: CPL-A-F_8-1031, 5: CPL-C-F_8-972, 6: CPL-C-F_8-966, 7: CPL-C-F_8-588, 8: CPL-D-F_8-824, 9: CPL-D-F_8-891, 10: CPL-D-F_8-905, 11: CPL-D-F_8-884, 12: CPL-D-F_8-887.

https://doi.org/10.1017/S0021859618000230 Published online by Cambridge University Press

Table 8. Fractionation of soluble protein in selected high protein rice lines and high and low protein checks

Genotype	Albumin %	Globulin %	Glutelin %	Prolamin %	Prolamin/Glutelin ratio
CPL-D-F_8-824	0.22	0.18l	4.16	0.09	0.02
CPL-A-F_8-1045	0.22	0.29	4.41	0.11	0.02
CPL-D-F_8-891	0.26	0.26	3.93	0.10	0.03
CPL-D-F_8-905	0.18	0.42	5.34	0.10	0.02
CPL-D-F_8-884	0.17	0.25	7.05	0.11	0.02
CPL-D-F_8-887	0.21	0.35	6.37	0.08	0.01
CPL-C-F_8-972	0.18	0.34	6.24	0.11	0.02
CPL-C-F_8-966	0.15	0.30	6.32	0.10	0.02
CPL-C-F_8-588	0.19	0.41	5.88	0.09	0.02
CPL-A-F_8-1049	0.18	0.22	5.55	0.11	0.02
CPL-A-F_8-1091	0.18	0.32	5.34	0.10	0.02
CPL-A-F_8-1031	0.15	0.35	6.58	0.10	0.02
Swarna	0.19	0.25	2.72	0.08	0.03
ARC10063	0.20	0.32	4.94	0.10	0.02
General Mean	0.19	0.3	5.35	0.1	0.02
P value	<0.001	<0.001	<0.001	<0.001	–
CV (%)	1.18	0.58	0.17	1.89	–
SE(d)	0.002	0.001	0.007	0.002	–
Tukey HSD at 5%	0.0068	0.0053	0.0267	0.0057	–

high GPC may produce more genotypes with unacceptably low levels of amylose in the rice grain. Therefore, the grain amylose content of all the breeding lines was also estimated. The alkali test has been employed extensively in rice quality studies as an indirect estimation of the gelatinization temperature. Therefore, along with amylose content, ASV also determines the cooking quality. At the end of the breeding cycle, elite lines with high GPC but also with desirable levels of ASV and amylose content were detected. The present breeding programme has significantly improved GPC in a high-yielding background with desirable cooking quality.

Genotype × environment interaction and identification of stable high protein elite lines

Grain protein content is highly influenced by the environment. A GGE biplot analysis is an effective way to identify stable high-protein genotypes partitioning genotype and GEI effects in many cereals (Khazratkulova *et al.* 2015). The current study presents the polygon view of the biplot where selected high mean genotypes for GPC and PROY are placed. The vertex genotypes are the best or the poorest genotypes in some or all of the environments because they are farthest from the origin of the biplot (Yan & Kang 2003). The stability of high GPC and PROY genotypes were evaluated by an average environment coordination (AEC) method (Yan 2002). Some of the high mean genotypes for GPC and PROY were found to be adaptable for different environments and the cooking quality of those lines were tested. Among the tested lines, a few were detected with desirable intermediate amylose content (20–25%) and ASV (3–5). They could be used either as improved introgression lines in breeding for grain quality or directly as biofortified rice varieties. Two genotypes (CPL-D-F_8-824, CPL-C-F_8-972) derived from the present breeding programme were nominated in the national biofortification trial for multi-locational testing.

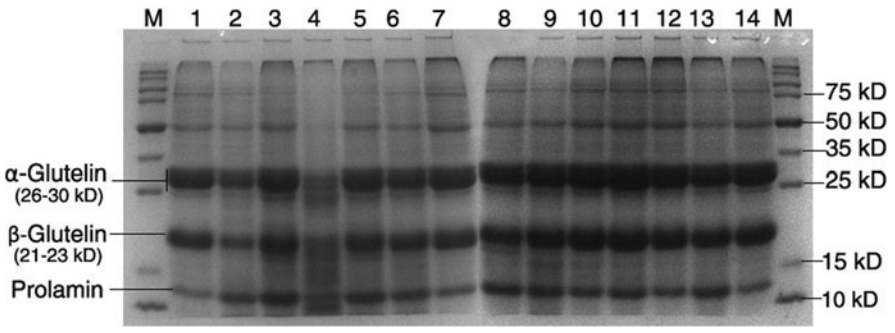

Fig. 10. The sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) profiles of partially purified glutelin fraction showing α (~29 kD) and β-glutelin (~21 kD) sub-unit and a prolamin band (~13-14 kD) in rice genotypes, viz., lane1: ARC1063, lane 2: Swarna, lane3: CPL-D-F_8-824, lane 4: CPL-A-F_8-1045, lane 5: CPL-D-F_8-891, lane 6: CPL-D-F_8-905, lane 7: CPL-D-F_8-884, lane 8: CPL-D-F_8-887, lane 9: CPL-C-F_8-972, lane 10: CPL-C-F_8-966, lane 11: CPL-C-F_8-588, lane 12: CPL-A-F_8-1049, lane 13: CPL-A-F_8-1091, lane14: CPL-A-F_8-1031 and lane M: ladder of 10-225 kD molecular weight. Glutelin samples were loaded at 15 µg on each lane and proteins were detected with standard Coomassie brilliant blue stain after electrophoresis.

https://doi.org/10.1017/S0021859618000230 Published online by Cambridge University Press

Protein fractionation and assessment of nutritional quality of protein in elite lines

It has been reported earlier that improvement of GPC reduced the protein quality and resulted in a hardening of the cooked rice (Derycke *et al.* 2005). Therefore, assessment of the quality of protein was required in high-protein elite lines. Hence, grain proteins of selected stable lines were fractionated into albumins, globulins, prolamins and glutelins. The results were similar to those of Kim *et al.* (2013) in waxy brown rice, which contained 1.86, 0.50, 7.31 and 0.05% albumin, globulin, glutelin and prolamin fractions, respectively. Values in the current study were relatively lower, perhaps because the extraction was done using milled rice or because of genotypic differences. Variations in fractionation procedure might also have contributed to the difference in values.

In rice endosperm, two types of PB are found where storage protein accumulates. Regular, spherical PB-I contains mostly prolamins while irregularly shaped PB-II contains mostly glutelins and globulins (Ogawa *et al.* 1987). It is known that PB-II is more digestible than PB-I. Therefore, glutelin is nutritionally more important than prolamins for human beings in respect of digestibility. Hence, the protein quality of rice could be enhanced further by increasing the glutelin and globulin contents (Kumamaru *et al.* 1988). Higher expression of the major storage protein, the glutelin fraction, ultimately indicates the higher level of PB-II in rice endosperm.

The SDS-PAGE profile of genotypes in the current study was found to be similar to the banding pattern obtained by Wen & Luthe (1985), where an α-glutelin polypeptide group with three bands had an average molecular weight of 29.6 kD and the smaller group, β-glutelin with two bands, had an average molecular weight of 21.1 kD. But other researchers (Mahmoud *et al.* 2008; Jiang *et al.* 2014; Pal *et al.* 2016) obtained slightly higher molecular weights forα-glutelin. This might be due to differences in the gel systems. The higher intensity of both groups of bands in high-protein lines compared with Swarna indicates that the enhanced GPC is due primarily to higher accumulation of the glutelin fraction as a storage protein. As glutelin is rich in all essential amino acids, the nutritional quality of these lines was also supposedly enhanced over the high-yielding parent, Swarna.

It has been reported earlier that prolamins generally consist of three polypeptide sub-units with the apparent molecular weight of 10, 13 and 16 kD. The band at approximately 13–14 kD region is predominant (Ogawa *et al.* 1987; Mahmoud *et al.* 2008). This prolamin band at a similar position (13–14 kD) appeared in the current SDS-PAGE profile of partially purified glutelin fraction of all the genotypes. Krishnan & Okita (1986) also reported that this prolamin band appeared as the major contaminant of the glutelin fraction. Mahmoud *et al.* (2008) reported that in a hybrid between IR 64 and *Oryza nivara* the expression of this prolamin band was much higher than that in the two parents. But the nutritional value of prolamins is inferior to glutelins for its low digestibility and negative influence on cooking quality, which increases the hardness of cooked rice (Ogawa *et al.* 1987; Furukawa *et al.* 2003). The breeding lines used in the current work showed no significant changes in the intensity of this prolamin band. Moreover, the similar or lower prolamin/glutelin ratio as observed in high-protein lines, compared with Swana, indicates that the protein quality of the high-protein rice lines remained unchanged if not improved by this breeding programme initiated for quantitative improvement in protein content.

Conclusion

Effective selection through calibrated NIRS and modification of bulk-pedigree rice breeding in grain protein improvement in both quantitative and qualitative basis were established through the current investigation, carried out for the past 6 years. Some of the biofortified lines (as pre-breeding lines) with high GPC and good cooking quality can be used for further improvement of high yielding cultivars. After multi-location testing, the stable high-yielding lines with high GPC can also be included for cultivation in the 'nutri-farms' developed in India to combat the problem of malnutrition in millions of underprivileged school-going children.

Acknowledgements. The authors acknowledge the research guidance received from Dr T. Mohapatra, Honourable Director General, Indian Council of Agricultural Research (ICAR), New Delhi.

Financial support. Financial assistance received from the Director, ICAR-NRRI and 'CRP on Biofortification of crops' project, ICAR, India.

Conflicts of interest. None.

References

Ali MA (2011) Response to pedigree selection for earliness and grain yield in spring wheat under heat stress. *Asian Journal of Crop Science* **3**, 118–129.

Allahgholipour M, *et al.* (2006) Relationship between rice grain amylose and pasting properties for breeding better quality rice varieties. *Plant Breeding* **125**, 357–362.

Aluko G, *et al.* (2004) QTL mapping of grain quality traits from the interspecific cross *Oryza sativa* × *O. glaberrima*. *Theoretical and Applied Genetics* **109**, 630–639.

Bagchi TB, Sharma SG and Chattopadhyay K (2015) Development of NIRS models to predict protein and amylose content of brown rice and proximate compositions of rice bran. *Food Chemistry* **191**, 21–27.

Blumenthal JM, *et al.* (2008) Importance and effect of nitrogen on crop quality and health. In Hatfield JL and Follett RF (eds). *Nitrogen in the Environment: Sources, Problems and Management*. pp. 51–70. Amsterdam, the Netherlands: Academic Press/Elsevier.

Cameron DK and Wang Y (2005) A better understanding of factors that affect the hardness and stickiness of long grain rice. *Cereal Chemistry* **82**, 113–119.

Chattopadhyay K, Das A and Das SP (2011) Grain protein content and genetic diversity of rice in north eastern India. *Oryza* **48**, 73–75.

Derycke V, *et al.* (2005) Impact of proteins on pasting and cooking properties of nonparboiled and parboiled rice. *Cereal Chemistry* **82**, 468–474.

El-Ameen T, Hossain A and da Silva JAT (2013) Genetic analysis and selection for bread wheat (*Triticum aestivum* L.) yield and agronomic traits under drought conditions. *International Journal of Plant Breeding* **7**, 61–68.

Falconer DS and Mackay TFC (1996) *Introduction to Quantitative Genetics*, 4th edn., Harlow, UK: Longman.

Fitzgerald MA, McCouch SR and Hall RD (2009) Not just a grain of rice: the quest for quality. *Trends in Plant Science* **14**, 133–139.

Frey KJ and Horner T (1955) Comparison of actual and predicted gains in barley selection experiments. *Agronomy Journal* **47**, 186–188.

Furukawa S, *et al.* (2003) Distribution of storage proteins in low-glutelin rice seed determined using a fluorescent antibody. *Journal of Bioscience and Bioengineering* **96**, 467–473.

Gaido AZ, Maich RH and Dubois MF (2000) Baking quality in lines of bread wheat with growing potential grain yield. *Phyton, International Journal of Experimental Botany* **69**, 85–90.

International Rice Research Institute (1976) *Annual Report for 1975*. Los Baños, Philippines: IRRI.

Jiang C, *et al.* (2014) Proteomic analysis of seed storage proteins in wild rice species of the *Oryza* genus. *Proteome Science* **12**, article no. 51. doi: 10.1186/s12953-014-0051-4.

https://doi.org/10.1017/S0021859618000230 Published online by Cambridge University Press

Ju Z, Hettiarachchy N and Rath N (2001) Extraction, denaturation and hydrophobic properties of rice flour proteins. *Journal of Food Science* **66**, 229–232.

Juliano BO (1980) Rice: recent progress in chemistry and nutrition. In Inglett GE and Munck L (eds). *Cereals for Food and Beverages, Recent Progress in Cereal Chemistry*, pp. 409–428. New York, USA: Academic Press.

Juliano B. O. (1985) Criteria and tests for rice grain qualities. In Juliano BO (ed). *Rice Chemistry and Technology*, 2nd edn., pp. 43–524. St. Paul, MN, USA: American Association of Cereal Chemists, Inc.

Khazratkulova S, *et al.* (2015) Genotype × environment interaction and stability of grain yield and selected quality traits in winter wheat in Central Asia. *Turkish Journal of Agriculture and Forestry* **39**, 920–929.

Khush GS and Juliano BO (1984) Status of rice varietal improvement for protein content at IRRI. In *Nuclear Techniques for Cereal Grain Protein Improvement: Proceedings of Research Coordination Meeting*, pp. 199–202. Vienna, Austria: IAEA.

Kim JW, *et al.* (2013) Protein content and composition of waxy rice grains. *Pakistan Journal of Botany* **45**, 151–156.

Krishnan HB and Okita TW (1986) Structural relationship among the rice glutelin polypeptides. *Plant Physiology* **81**, 748–753.

Kumamaru T, *et al.* (1988) Mutants for rice storage proteins. 1. Screening of mutants for rice storage proteins of protein bodies in the starchy endosperm. *Theoretical and Applied Genetics* **76**, 11–16.

Lal M and Singh DP (2012) Genotype × environment interaction in rice (Oryza sativa L.). *Annals of Biology* **28**, 53–55.

Li JM, *et al.* (2004) QTL detection for rice grain quality traits using an interspecific backcross population derived from cultivated Asian (*O. sativa* L.) and African (*O. glaberrima* S.) rice. *Genome* **47**, 697–704.

Li J, *et al.* (2009) Breeding elite *japonica*-type soft rice with high protein content through the introduction of the anti-*Waxy* gene. *African Journal of Biotechnology* **8**, 161–166.

Lowry OH, *et al.* (1951) Protein measurement with the Folin Phenol reagent. *Journal of Biological Chemistry* **193**, 265–275.

Mahmoud AA, Sukumar S and Krishnan HB (2008) Interspecific rice hybrid of *Oryza sativa* × *Oryza nivara* reveals a significant increase in seed protein content. *Journal of Agricultural and Food Chemistry* **56**, 476–482.

Ntanos DA and Koutroubas SD (2002) Dry matter and N accumulation and translocation for Indica and Japonica rice under Mediterranean conditions. *Field Crops Research* **74**, 93–101.

Ogawa M, *et al.* (1987) Purification of protein body-I of rice seed and its polypeptide composition. *Plant and Cell Physiology* **28**, 1517–1527.

Pal P, *et al.* (2016) Effect of nonthermal plasma on physico-chemical, amino acid composition, pasting and protein characteristics of short and long grain rice flour. *Food Research International* **81**, 50–57.

Sambrook J and Russell D (2001) *Molecular Cloning: A Laboratory Manual*. New York, USA: CSHL Press.

Shao Y, *et al.* (2011) Infrared spectroscopy and chemometrics for the starch and protein prediction in irradiated rice. *Food Chemistry* **126**, 1856–1861.

Singh NB and Singh HG (1982) Gene action for quality components in rice. *Indian Journal of Agricultural Science* **52**, 485–488.

Sugimoto T, Tanaka K and Kasai Z (1986) Improved extraction of rice prolamin. *Agricultural and Biological Chemistry* **50**, 2409–2411.

Ufaz S and Galili G (2008) Improving the content of essential amino acids in crop plants: goals and opportunities. *Plant Physiology* **147**, 954–961.

Vasal SK (2002) The role of high lysine cereals in animal and human nutrition in Asia. In *Protein Sources for the Animal Feed Industry*, pp. 167–184. Rome, Italy: FAO.

Wen TN and Luthe DS (1985) Biochemical characterization of rice glutelin. *Plant Physiology* **78**, 172–177.

Wynne JC and Gregory WC (1981) Peanut breeding. *Advances in Agronomy* **34**, 39–71.

Yan W (2002) Singular-value partitioning in biplot analysis of multienvironment trial data. *Agronomy Journal* **94**, 990–996.

Yan W and Kang MS (2003) *GGE Biplot Analysis: A Graphical Tool for Breeders, Geneticists, and Agronomists*. Boca Raton, FL, USA: CRC Press.

Yan W, *et al.* (2007) GGE biplot vs. AMMI analysis of genotype-by-environment data. *Crop Science* **47**, 641–653.

Yang LJ, Xu L and Li JY (2004) Analysis of correlation between protein content, amylose content in the unpolished rice and 1000-grain weight in six different cultivars' rice. *Journal of Shanghai Normal University* (*Natural Sciences*) **10**(suppl.), 55–58.

Yang Y, *et al.* (2015) Identification of quantitative trait loci responsible for rice grain protein content using chromosome segment substitution lines and fine mapping of qPC-1 in rice (Oryza sativa L.). *Molecular Breeding* **35**, article no. 130, http//dx.doi:10.1007/s11032-015-0328-z.

Yoshida S, *et al.* (1976) *Laboratory Manual for Physiological Studies of Rice*, 3rd edn. Los Banos, the Philippines: IRRI.

Yoshida S, *et al.* (2002) QTL analysis for plant and grain characters of sake-brewing rice using a doubled haploid population. *Breeding Science* **52**, 309–317.

Zheng LN, *et al.* (2011) Dynamic QTL analysis of rice protein content and protein index using recombinant inbred lines. *Journal of Plant Biology* **54**, 321–328.

https://doi.org/10.1017/S0021859618000230 Published online by Cambridge University Press