Theor Appl Genet (2010) 121:741–750
DOI 10.1007/s00122-010-1345-2

ORIGINAL PAPER

QTL mapping of protein content in rice using single chromosome segment substitution lines

Guoyou Ye · Shanshan Liang · Jianmin Wan

Received: 1 December 2008 / Accepted: 18 April 2010 / Published online: 16 May 2010

Abstract Protein content (PC) is an important component of rice nutritional quality. In order to better understand the genetic basis of this trait and increase related breeding efficiency, 21 single chromosome segment substitution (SCSS) lines grown in four sites over two growing seasons (regarded as eight environments) were used to associate PC with particular chromosome segments. Segments from 15 chromosomes were found to contain quantitative trait loci (QTLs) for PC in at least one environment. These included segments from chromosome 3 and 8, in which QTLs for PC had not previously been identified. The segment of chromosome 8 in CSSL-48 had the largest positive effect across all environments. The interaction between substitution and environment was highly significant. Some substitutions had large effects in one environment, but no effect in another (i.e. CSSL-08 and CSSL-17), while some substitutions significantly increased PC in one environment but decreased it in another (i.e. CSSL-41 and CSSL-43). By biplot and clustering analysis, the eight environments were grouped into two contrasting environment types, that is, Hainan and Jiangsu. The segment of chromosome 8 in CSSL-48 had PC-enhancing QTLs in both of the environment types. The segments in CSSL-34 had QTLs which increase PC in the Jiangsu environment but have no effect in the Hainan environment. For enhancing PC, CSSL-48 could be explored in breeding for wide adaptation across all environments, while CSSL-12, CSSL-14, CSSL-17, CSSL-41 and CSSL-43, and that in CSSL-34 could be explored in breeding for specific adaptation to the Hainan and Jiangsu environments, respectively. Near isogenic lines are under development to validate the QTLs with large effects in a range of genetic backgrounds relevant to Jiangsu and Hainan breeding programs. Secondary mapping populations are also being developed for further localising the responsible QTLs in CSSL-14, CSSL-34 and CSSL-48.

Communicated by M. Cooper.

G. Ye
Crop Research Informatics Laboratory,
International Rice Research Institute,
Los Baños, Laguna, Philippines

S. Liang
Plant Breeding, Genetics, and Biotechnology Division,
International Rice Research Institute (IRRI), Los Baños,
Laguna, Philippines

J. Wan (✉)
Institute of Crop Science and The National Key
Facility for Crop Gene Resources and Genetic Improvement,
Chinese Academy of Agricultural Sciences,
Beijing 100081, China
e-mail: wanjm@caas.net.cn

Introduction

Rice (*Oryza sativa*) is one of the most important crops in the world, and used as staple food by more than half the world's population. Traditionally, plant breeders concentrated on breeding for high yield and pest resistance. As countries reached self-sufficiency in rice production, the demand by consumers for better quality rice has increased. Breeding for improved grain quality is now a major objective of many breeding programs worldwide. Rice grain quality is a complex character consisting of many components such as grain appearance, milling, cooking, palatability and nutritional qualities. Protein content (PC) of rice grain influences the taste of cooked rice. It also affects the physical–chemical properties of cooked rice. It is generally thought that the higher the PC, the more the nutritional value, but in terms of taste reduced PC appears desirable (Wan et al. 2004).

Substantial variation for PC has been detected among *indica* varieties with a range from 4.9 to 19.3% and *japonica* varieties from 5.9 to 16.5% (Lin et al. 2000). Classical genetic analysis suggested that PC was quantitatively inherited (Shenoy et al. 1991; Shi et al. 1999). PC is sensitive to environmental conditions like the level of nitrogen fertilization (Perez et al. 1996), which has contributed to the slow progress of breeding efforts to improve PC.

To design effective and efficient breeding methods for improving PC, several QTL mapping studies have been conducted using different mapping populations. Using a population of recombinant inbred lines derived from a cross between two elite varieties, Tan et al. (2001) identified two PC QTLs on chromosome 6 and 7, which accounted for 13.0 and 4.7% of the phenotypic variation, respectively. The QTL on chromosome 6 was in the region of the waxy (*wx*) gene. Aluko et al. (2004) identified four QTLs on chromosomes 1, 2, 6 and 11 using a double haploid population derived from an interspecific cross between *O. sativa* and *O. glaberrima*. The QTL on chromosome 6 was also in the region of the *wx* gene. Hu et al. (2004) identified five main-effect QTLs on chromosomes 1, 4, 5, 6 and 7, and a pair of interacting QTL on chromosome 10 and 11 using a double haploid population derived from a cross between *indica* and *japonica* varieties. The five main-effect QTLs collectively explained 74% of the phenotypic variation. Both of the parental lines contributed alleles with increased PC. Thus, 8 of the 12 rice chromosomes (1, 2, 4, 5, 6, 7, 10 and 11) have so far been shown to have QTLs for PC. All these studies used phenotyping data from only a single environment, which did not allow the investigation of the possible effects of QTL-by-environment interaction on trait performance.

The objectives of this study are to (1) test the effects of various chromosome segment substitutions on PC in different environments; (2) investigate the interaction between chromosome substitution and testing environments; (3) identify lines that have large and stable effects for further locating QTLs responsible for the observed effects. The biplot is used to investigate the substitution-by-environment interaction pattern including (1) the magnitude of the substitution effect, (2) the similarity among substitutions in effect and response to environments, (3) the average effect of a substitution across environments and its stability, and (4) the similarity among environments in discriminating substitution effects.

Materials and methods

Genetic population

Tsunematsu et al. (1996) developed a set of 65 chromosome segment substitution lines using successive backcrossing and marker-assisted selection. The *japonica* variety Asominori was used as the background parent and the *indica* rice variety IR24 as the donor parent (Kubo et al. 2002). Twenty-one of these lines, which have only a single segment substitution (Table 1), and the recurrent parent Asominori were grown in eight environments (denoted as "E") involving four locations in two different years (Table 2). A randomized complete block design with two replications was used to layout the trial in all environments (years and locations). Each plot consisted of ten ten-plant rows. Trials were managed according to standard local practices. At maturity, each entry was harvested in bulk. After drying, grains were stored at room temperature for 3 months before processing. Protein content was determined by a nitrogen gas analyser. Samples of 1 mg were placed into a quartz combustion tube in an induction furnace at 900°C. A nitrogen conversion of 5.95 was used to calculate the nitrogen content.

Data analysis

For within-environment analysis, the linear model for the PC of genotype i in block k is

$$y_{ik} = \mu + g_i + b_k + \varepsilon_{ik}$$

where y_{ik} is the observed PC of genotype i, in block k, μ is the overall mean, g_i is the fixed effect of genotype i, b_k is the random effect of block j, ε_{ik} is the random error term associated with observation y_{kj}.

Analysis of variance was performed using the GLM procedure of SAS (SAS Institute 2001).

The differences between the SCSS lines and the recurrent parent Asominori were tested for significance using the Dunnett approach (Dunnett 1964).

For across-environment analysis, the linear model for the PC of genotype i in block k in environment j is

$$y_{ijk} = \mu + g_i + e_j + ge_{ij} + (b/e)_{k(j)} + \varepsilon_{ijk}$$

where y_{ijk} is the PC of genotype i, in environment j and block k, μ is the overall mean, g_i is the effect of genotype i, e_j is the effect of the environment j, ge_{ij} is the effect of the interaction of genotype i and environment j, $(b/e)_{k(j)}$ is the effect of block k within the environment j, ε_{ijk} is the random error term associated with observation y_{ijk}.

Prior to the analysis, the homogeneity among error variances estimated from each of the environments was checked with a Bartlett's test (1937) ($P = 0.05$). Variance components were estimated by REML method by regarding all effects as random. To compare the SCSS lines against the recurrent parent across environments, an analysis was conducted using the same model but regarding the effect of genotype as fixed. Both the analyses were

Table 1 Asominori/IR24 single chromosome segment substitution lines

Code	Chromosome	Segment length (cm)	Marker composition
CSSL-03	1	38.55	C970, C955
CSSL-08	1	54.10	C1370, C2340, C86
CSSL-10	2	51.90	R459, G1340, X132
CSSL-12	2	53.35	X132, X67, C747, R3393
CSSL-14	2	11.50	C1470
CSSL-17	3	73.65	C515, R518, C563, R3156, C1677, R19
CSSL-22	4	23.95	R1854, R228
CSSL-24	4	51.25	C891, C335, X331
CSSL-31	5	45.40	R569, R2289
CSSL-32	5	79.95	R3166, R569, R2289, C128
CSSL-33	5	11.5	C1447
CSSL-34	6	42.15	X209, C688, R2171
CSSL-41	7	49.80	C1057, R2829, C39
CSSL-43	7	44.85	C39, R2394
CSSL-46	7	28.25	R1789, X278
CSSL-47	8	13.05	X278
CSSL-48	8	54.45	X41, R727, G1149, X397
CSSL-57	10	76.00	R1629, C1166, R1877, C1361, C809
CSSL-58	10	86.65	C1166, R1877, C1361, C809, C405
CSSL-59	11	4.80	X 52, R2918
CSSL-62	11	83.70	C1350, X257, C1172, C1465, R543

Table 2 Environments (combinations of two growing seasons and four locations) for testing 21 Asominori/IR24 single chromosome segment substitution lines and Asominori

Code	Season	Location
E1	May–Oct 2001	Nanjing Agricultural University, Nanjing, China, N 31.2°, E 118.4 Â°
E2	May–Oct 2001	Jinhu County, Jiangsu, China, N 32.7°, E 119.6°
E3	June–Nov 2001	Donghai County, Jiangsu, China, N 35.1°, E 118.4°,
E4	Dec 2001–May 2002	Lingshui County, Hainan, China, N 18.2°, E 108.9°
E5	May–Oct 2002	Nanjing Agricultural University, Nanjing, China, N 31.2°, E 118.4°
E6	May–Oct 2002	Jinhu County, Jiangsu, China, N 32.7°, E 119.6°
E7	June–Nov 2002	Donghai County, Jiangsu, China, N 35.1°, E 118.4°
E8	Dec 2002–May 2003	Lingshui County, Hainan, China, N 18.2°, E 108.9°

conducted using the MIXED procedure of the SAS software (SAS Institute 2001).

The substitution effect was computed as the difference between an SCSS line and the recurrent parent and used for environment and line grouping. Classification of lines and environments was performed using an agglomerative hierarchical clustering procedure with squared Euclidean distance as the dissimilarity measure (Williams 1976), and Ward's method, which uses incremental sums of squares as the clustering strategy (Ward 1963). The R package pvclust (Shimodaira 2004) was used for assessing the uncertainty in hierarchical cluster analysis. For each cluster in hierarchical clustering, the approximately unbiased *P* value computed by multiscale bootstrap resampling and the bootstrap probability value computed by normal bootstrap resampling were calculated (Shimodaira 2004). The *P* value of a cluster is a value between 0 and 1, which indicates how strong the cluster is supported by data. Ordination was done on the environment centred mean data using a singular value decomposition algorithm with results represented by a biplot (Gabriel 1971; Kempton 1984) using the R (R Development Core Team 2008) and GGEBiplot software (Yan 2001).

Results

Effects of chromosome segment substitutions in eight testing environments

The genotype (line) effect was significant for all the eight trials. All the 21 SCSS lines were not significantly different

from the recurrent parent Asominori at E_1, E_4 and E_6 mainly due to large experimental error (Table 3). At E_4, line 58 was different from the Asominori at 0.1 probability level. The numbers of lines that were significantly different from Asominori in E_2, E_3, E_5, E_7 and E_8, were five, one, four, two and seven, respectively.

The segment of chromosome 1 in CSSL-03 had a significantly negative effect at E_3 and E_7, while the segment of chromosome 1 in CSSL-08 had a significantly positive effect at E_2. The segment of chromosome 2 in CSSL-10 had a significantly positive effect at E_2, but a nearly significantly negative effect at E_3. The segment of chromosome 2 in CSSL-12 had a significant positive effect at E_8. The two segments overlapped with the common marker 13. Therefore, the gene responsible for the observed response was likely to be between marker 13 and 14. The single marker substitution in CSSL-14 increased PC highly significantly at E_8, where it had the second largest difference from Asominori among all the SCCS lines. The segment of chromosome 3 in CSSL-17 had a significantly negative effect at E_7 but a significantly positive effect at E_8. The two segments of chromosome 4 (in CSSL-22 and CSSL-24, respectively) had significantly positive effects. The effect was detected at E_2, E_5 and E_8 for CSSL-22, but only at E_2 for CSSL-24. The segment of chromosome 5 in CSSL-31 had no effect, while that in CSSL-32 significantly reduced PC at E_5. The single marker substitution of chromosome 5 in CSSL-33 significantly reduced PC at E_5. The segment of chromosome 6 in CSSL-34 had no significant effect in any of the environments. The three segments of chromosome 7 (CSSL-41, CSSL-43 and CSSL-46) spanned markers 48–53. Significant and near significant effects were only detected for CSSL-41 and CSSL-43 at E_8 and E_2, indicating that the underlying gene was in between marker 50 and 51. CSSL-47, which has a single marker substitution from chromosome 8 was not significantly different from Asominori in all testing environments. The segment of chromosome 8 in CSSL-48 had a significantly positive effect at E_2, E_7 and E_8, where it also had the largest positive difference among all the SCCS lines. The two segments of chromosome 10 in CSSL-57 and CSSL-58 had no significant effects on PC in any of the environments. The segment of chromosome 11 in CSSL-59 had a significant effect only in E_5, where it reduced PC. The segment of chromosome 11 in CSSL-62 had a significant effect only in E_2, where it increased PC.

Across environments analysis

Environment main effect was the predominant source of variation, followed by line-by-environment interaction. The line-by-environment interaction variance component was 2.13 times the genotype component, indicating a large genotype-by-environment interaction (Table 4). Single degree of freedom *t* test conducted by the 'Estimate' statement of Proc Mixed identified CSSL-48, CSSL-34 and CSSL-14 had significantly lower or higher PC than the recurrent parent, indicating the presence of QTLs with main effects.

Substitution with major effect

The biplot in Fig. 1 is based on the first two principal components from the line-focused singular value partitioning, where SCSS lines were presented as vectors (lines from the biplot origin to the SCSS line coordinates) and environments were presented as points. The first two components account for 83.0% of the total variation. The length of a line vector is a measure of the magnitude of the effect of the substituted segment in the line. A long vector indicates that the substituted segment has a large effect in one or more environments. Thus, segments in CSSL-08, CSSL-12, CSSL-14, CSSL-17, CSSL-22, CSSL-34, CSSL-41, CSSL-43 and CSSL-48 are estimated to have large effects (Fig. 1). The segments in CSSL-46 and CSSL-47, which were not significantly different from Asominori in any environment, have short vector lengths.

Similarity among substitution effects

The distance between two lines in Fig. 1 is an indicator of dissimilarity between them, which is determined by both the length of their vectors and the angle between them. It can be seen that the segment in CSSL-34 is different from most of the segments in the other lines. The segments in CSSL-12, CSSL-14, CSSL-17 and CSSL-43 are similar. Similarly, the segment in CSSL-33 is similar to those in CSSL-59 and CSSL-41, and the segment in CSSL-62 is similar to that in CSSL-47. Since the biplot analysis is conducted using the raw line-by-environment table, the environment main effect, which is not of interest, may have large effect on the classification of the substitution effects. Cluster analysis based on environment standardized data was conducted to classify lines into groups within which all lines have relatively similar ranking of performance. The 21 lines were classified into five groups without losing much information, since 84% of the line sum of squares was captured by the among line group sum of squares. The result of clustering is given in Table 5, which shows a good coincidence with the information in Fig. 1 for most of the lines. However, CSSL-48 is separated out as a single member group, while CSSL-34 is grouped together with CSSL-47 and CSSL-57.

Table 3 Protein content (PC, %) of the two parental lines and the effect of single chromosome segment substitution on rice PC (%) in eight testing environments

Line	Chromosome	Environment															
		E1 (0.28)[a]		E2 (0.08)		E3 (0.15)		E4 (0.09)		E5 (0.05)		E6 (0.35)		E7 (0.08)		E8 (0.14)	
		Effect[b]	*P*	Effect	*P*	Effect	*P*	Effect	*P*	Effect	*P*	Effect	*P*	Effect	*P*	Effect	*P*
Asominori[c]	–	9.22	–	9.35	–	9.05	–	9.21	–	9.02	–	9.92	–	9.13	–	9.14	–
IR 24[c]	–	10.49	–	10.67	–	10.53	–	10.56	–	10.04	–	10.03	–	9.89	–	9.57	–
CSSL-03	1	0.44	0.997	−0.27	0.991	−1.30	0.034	−0.42	0.861	−0.59	0.180	−1.04	0.609	−1.03	0.018	0.54	0.831
CSSL-08	1	0.37	1.000	1.30	0.003	−0.27	1.000	0.11	1.000	−0.30	0.877	−0.28	1.000	0.34	0.926	0.04	1.000
CSSL-10	2	0.47	0.993	0.95	0.039	−1.12	0.094	0.42	0.852	0.07	1.000	−0.59	0.986	−0.49	0.611	0.46	0.933
CSSL-12	2	0.56	0.973	0.88	0.067	−0.87	0.304	0.81	0.163	−0.48	0.396	−0.80	0.888	−0.66	0.247	1.27	0.037
CSSL-14	2	0.46	0.996	0.84	0.086	−0.53	0.863	0.60	0.475	−0.21	0.993	−0.75	0.914	−0.03	1.000	2.31	<0.0001
CSSL-17	3	−0.09	1.000	0.002	1.000	−0.71	0.543	−0.40	0.894	−0.89	0.012	−0.82	0.853	−0.77	0.121	1.72	0.003
CSSL-22	4	0.42	0.998	1.31	0.002	−0.42	0.968	−0.42	0.861	−0.84	0.018	−1.30	0.336	−0.32	0.951	1.38	0.020
CSSL-24	4	0.09	1.000	0.90	0.058	−0.22	1.000	0.06	1.000	−0.25	0.967	−0.62	0.979	−0.52	0.525	0.74	0.467
CSSL-31	5	−0.05	1.000	0.41	0.822	−0.58	0.785	−0.28	0.993	−0.55	0.235	−1.03	0.627	−0.11	1.000	1.05	0.122
CSSL-32	5	0.14	1.000	−0.21	0.999	−0.70	0.552	−0.63	0.405	−0.90	0.011	−0.67	0.958	0.70	0.232	0.81	0.358
CSSL-33	5	−0.37	1.000	−0.33	0.956	−0.80	0.394	0.15	1.000	−0.71	0.060	−0.62	0.979	−0.42	0.772	0.51	0.875
CSSL-34	6	0.86	0.687	0.78	0.133	0.96	0.186	−0.34	0.964	−0.04	1.000	−0.87	0.808	0.48	0.636	0.00	1.000
CSSL-41	7	−0.13	1.000	−0.36	0.920	−0.73	0.508	0.59	0.496	−0.68	0.082	−1.02	0.633	−0.14	1.000	1.70	0.003
CSSL-43	7	−0.37	1.000	0.39	0.862	−0.51	0.883	0.26	0.997	−0.68	0.085	−0.75	0.911	−0.34	0.926	1.49	0.010
CSSL-46	7	−0.38	1.000	0.60	0.396	−0.80	0.394	−0.02	1.000	−0.36	0.745	−0.68	0.953	−0.13	1.000	0.23	1.000
CSSL-47	8	0.34	1.000	0.43	0.779	0.26	1.000	0.21	1.000	−0.24	0.976	−1.02	0.639	0.41	0.795	0.48	0.918
CSSL-48	8	1.040	0.449	1.66	0.000	0.691	0.561	0.71	0.270	0.010	1.000	−0.12	1.000	0.93	0.039	2.67	<0.0001
CSSL-57	10	−0.07	1.000	0.13	1.000	0.072	1.000	0.35	0.960	0.000	1.000	−0.91	0.754	−0.30	0.970	1.05	0.122
CSSL-58	10	−0.40	0.999	−0.09	1.000	−0.135	1.000	−0.88	0.103	−0.09	1.000	−0.27	1.000	−0.20	1.000	0.17	1.000
CSSL-59	11	−0.36	1.000	0.08	1.000	−0.790	0.409	0.46	0.794	−0.78	0.032	−0.84	0.838	−0.65	0.262	0.91	0.244
CSSL-62	11	−0.45	0.997	1.16	0.008	0.090	1.000	−0.15	1.000	0.14	1.000	−0.48	0.998	0.01	1.000	0.71	0.528

[a] Figures in parenthesis is the residual variance

[b] A negative value indicates that it has a lower PC than Asominori

[c] Protein content (%)

Table 4 Variance components estimated from a trial of 21 rice single chromosomal segment substitution lines tested in 8 environments (4 locations over 2 years)

Source of variation	Component		Percentage of total variance
	Estimate	Error	
Genotype (G)	0.058	0.026	9.94
Environment (E)	0.243	0.136	41.99
G-by-E	0.123	0.025	21.17
Block (E)	0.003	0.005	0.47
Residual	0.153	0.017	26.44

Mean effect of a substitution and its stability across environments

Figure 2 displays the mean effect of a substitution and its stability across testing environments. The small circle indicates the average environment defined by the average of PC_1 and PC_2 across environments. The horizontal line, which passes through the circle and the biplot origin, is the average environment axis. The vertical line, which is perpendicular to the average environment axis and passes through the origin of the biplot, separates lines with below average PC from those with above average PC. Projection of a line point to the horizontal average environment axis approximates the mean effect of the substitution in the line (Yan 2001). Thus, the segment in CSSL-48 has the largest mean effect. Other segments with large effect are CSSL-03, CSSL-14, CSSL-33 and CSSL-59. As expected, most of them are also the substitutions with major effects. Substitutions in lines 08, 14, 34 and 48 have above average PC, while those in CSSL-03, CSSL-33, CSSL-58 and CSSL-59 have below average PC. A longer projection to the vertical line, regardless of the direction, represents a greater tendency of line-by-environment interaction of a line, which means the line is more variable and less stable across environments (Yan 2001). Thus, the effects of substitutions in CSSL-48, CSSL-22 and CSSL-62 correspond to the most stable lines that have large positive effects.

Similarity among testing environments and identification of lines for different environment types

Biplot analysis based on environment-centred data is used to visualise the relationships among testing environments (represented by vectors) (Fig. 3). The length of an environment vector indicates the magnitude of the difference among segment substitution (lines) acting in the environment. It can be seen that most substitutions have different effects in E_3 versus E_8. The distance between two environments is an indicator of dissimilarity between them, which is determined by both the length of the vectors and the angle between them. From Fig. 3, E_1 and E_2 are similar; E_4 and E_8 are very similar; and E_3, E_5, E_6 and E_7 are similar. The eight testing environments appear to belong to three environment groups: $E_1 + E_2$, $E_4 + E_8$ and $E_3 + E_5 + E_6 + E_7$. The dendrogram of the hierarchal cluster analysis gives similar grouping (Fig. 4). Cluster stability assessment using bootstrap resampling supports two environment groups (Fig. 4). The combined results from the interpretation of the biplot and the cluster stability analysis indicated two environment groups for further investigation of the genotype-by-environment interaction. E_4 and E_8 are in one group and the remaining environments are in the other group. E_4 and E_8 are the two growing seasons of the testing site in Hainan province (Hainan environment). The other six environments are the two growing seasons of the three testing sites in Jiangsu province (Table 2) (Jiangsu environment).

To ease the identification of lines for the two environment types, the result of a biplot analysis using the two identified environment groups is given in Fig. 5. The segment in CSSL-48 is good for both the environment types. Segment in CSSL-12 and CSSL-14 increases PC for the Hainan but has essentially no effect for the Jiangsu environment. Conversely, the substitution in CSSL-34 has no effect for the Hainan environment but increases PC for the Jiangsu environment. The QTLs contained by the segments in CSSL-12, CSSL-14, CSSL-17, CSSL-41 and CSSL-43 can be explored for the Hainan environment, while those in CSSL-08 and CSSL-34 are of interest for the Jiangsu environment.

Discussion

In this study, twenty-one SCSS lines were used for detecting segments containing QTLs for PC across eight environments. SCSS lines have several distinct advantages over primary mapping populations such as F_2, F_3, recombinant inbred lines, and double haploids in detecting QTLs for complex traits. First, detection power is increased because of reduced effects of interferences from genetic background. An SCSS line carries only a single introduced segment in the near-isogenic background of a recurrent genotype. Interactions between donor alleles are limited to those between QTLs on homozygous substituted tracts, and thus are substantially reduced (Howell et al. 1996; Yano et al. 2000). This may be best illustrated by the fact that in the E_8 trial, which had the lowest experimental error, as many as seven substitutions were found to significantly increase PC, while no previous studies using much larger DH populations identified more than five QTLs for PC. Second, high-resolution mapping of putative QTLs as Mendelian factors and further map-based cloning will be

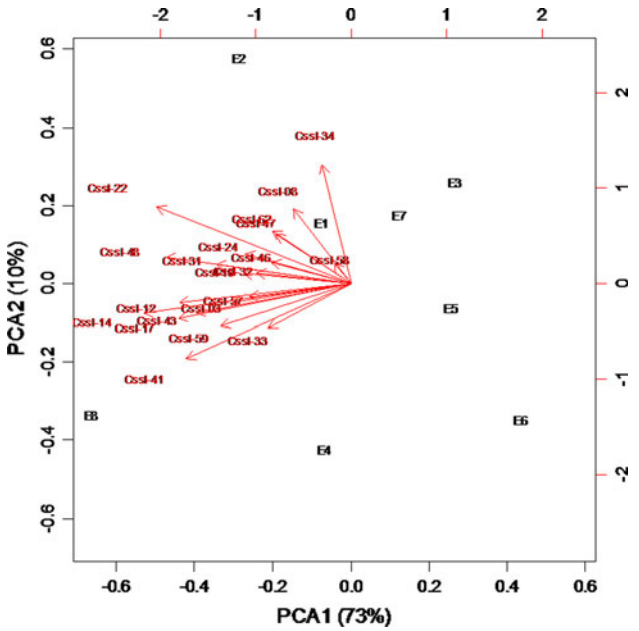

Fig. 1 Biplot based on the first two principal components for comparing SCSS lines and for visualising similarities among them (The substitution lines are represented by *vectors*. The environments are represented by *points*. The length of a vector indicates the predicted effect of the substitution line. The angle between two vectors indicates the similarity between two substitution lines. A substitution line with a short vector is not well presented by the first two principal components and more components are needed)

feasible using a secondary F_2 population derived from a cross between a QTL-containing SCSS line and the recurrent parent (Eshed and Zamir 1995; Frary et al. 2000; Nadeau et al. 2000). In addition, segregating populations can be purposely created to detect the effect of interaction between QTLs (epistasis) (Lin et al. 2000). SCSS lines are also useful for breeding purposes, as they contain only favourable donor alleles and a low percentage of donor genome and these can thus be easily and rapidly isolated and transferred into elite varieties.

Fifteen of the chromosome segments in the 21 SCSS lines were shown to have significant or nearly significant effects on PC in at least one of the eight testing environments. These 15 segments were from 10 of the 12 rice chromosomes. As summarized in the introduction, 8 of the 12 rice chromosomes (1, 2, 4, 5, 6, 7, 10 and 11) have been previously shown to contain QTLs for PC. This study added two more chromosomes (3 and 8) to the list. The segment of chromosome 3 in CSSL-17 had no effect at E_1 and E_2 and increased PC by 1.72% at E_8, but reduced it by different extents at the other five environments. The segment of chromosome 8 in CSSL-48 had no effect at E_5 and slightly reduced PC at E_6, but increased it at other environments. Indeed, the segment in CSSL-48 had the largest positive effect in most of the environments. It is possible that chromosomes 9 and 12 may also have QTLs for PC, since no SCSS lines contain substituted segments from these two chromosomes in this study.

The numbers of lines that were significantly different from Asominori were different for different environments. In E_1, E_4 and E_6, no significant difference was found for any of the lines, while five, one, four, two and seven lines were significantly different from Asominori in E_2, E_3, E_5, E_7 and E_8, respectively. The effect of substitution is environment-specific. The difference between an SCSS line and Asominori was essentially zero in one environment, but large in another environment. For instance, the effect of the substitution in CSSL-62 was 0.005% at E_8 but 1.160% at E_2. Similarly, the effect of substitution in CSSL-08 was 0.036% at E_7 but 1.302% at E_2.

Many substitutions were found to reduce PC, suggesting that Asominori also contains PC-enhancing QTLs. A set of reciprocal substitution lines (if available) will not only help the identification of PC-enhancing QTLs contained in Asominori, but also make it possible to detect the effect of QTL-by-background interaction.

As QTL mapping using other types of biparental populations, the desirable QTLs are only desirable for the genetic background where it resides. Validation of the QTLs in a range of genetic backgrounds relevant to a target breeding program is required. CSSL-48, CSSL-14 and CSSL-08 are being used as donor parents to transfer the QTLs with large effects into 5–10 elite cultivars/breeding lines from the Hainan and Jiangsu breeding programs. It is expected that near isogenic lines with a wide range of genetic backgrounds will be developed quickly by marker assisted backcrossing since the segments containing QTLs are short. Once the effects of the QTLs are validated, the near isogenic lines can also be used as parental lines for

Table 5 SCSS line grouping resulting from hierarchical cluster analysis using Ward method

Group	Lines
I	CSSL-48
II	CSSL-10, CSSL-12, CSSL-14
III	CSSL-34, CSSL-47, CSSL-57
IV	CSSL-08, CSSL-24, CSSL-46, CSSL-58, CSSL-62
V	CSSL-03, CSSL-17, CSSL-22, CSSL-31, CSSL-32, CSSL-33, CSSL-41, CSSL-43, CSSL-59

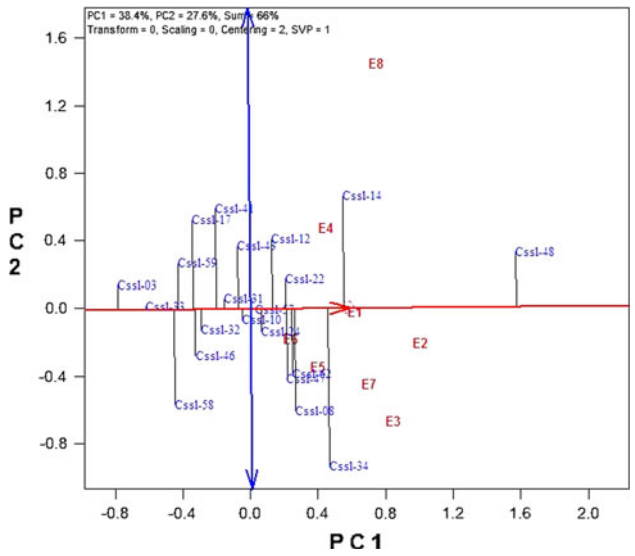

Fig. 2 Mean effect and stability of the SCSS lines across testing environments [The genotypes are ranked along the average-environment axis (*the longest horizontal line*), with the *arrow* pointing in the direction of greater values. The *longest vertical line* separates substitution lines with below-average means from those with above-average means. A substitution line with a longer projection to the *vertical line* is more variable and less stable across environments]

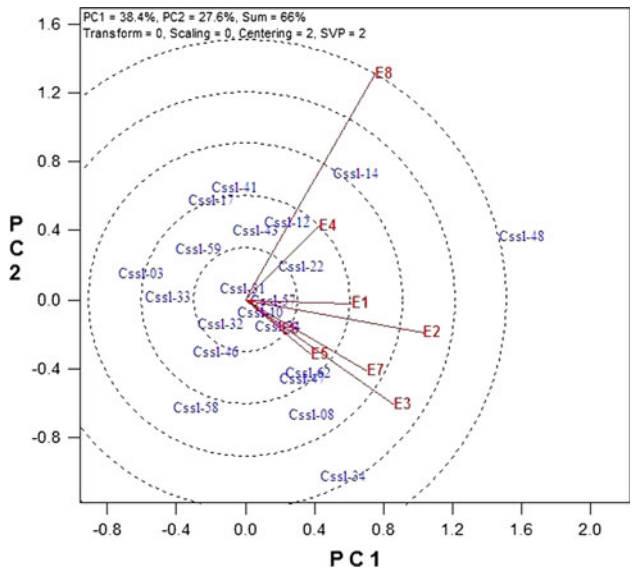

Fig. 3 Biplot based on the first two principal components for showing the similarities among the testing environments (Testing environments are represented by *vectors*. The substitution lines are represented by *points*. Environments with longer vectors are more discriminative of the substitution lines. The angle between two vectors indicates the similarity between the two testing environments)

pyramiding these QTLs, as discussed by Ye and Smith (2010).

The two parental cultivars used in the development of these SCSS lines are elite cultivars of the two rice

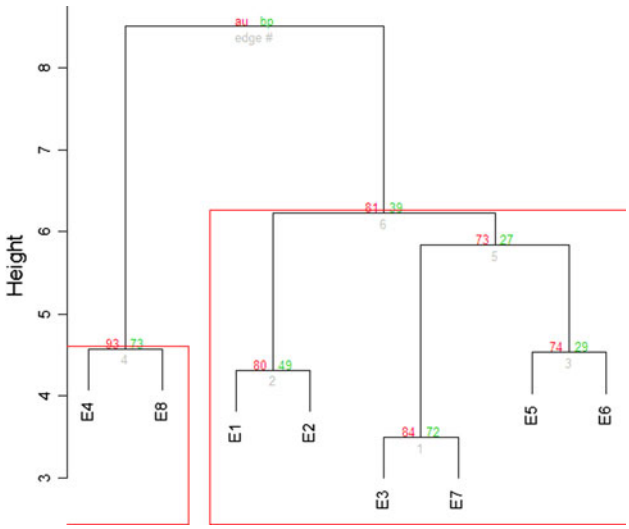

Fig. 4 Cluster dendrogram for environments with *P* value (AU is approximately unbiased *P* value computed by multiscale bootstrap resampling and BP is the Bootstrap Probability value computed by normal bootstrap resampling)

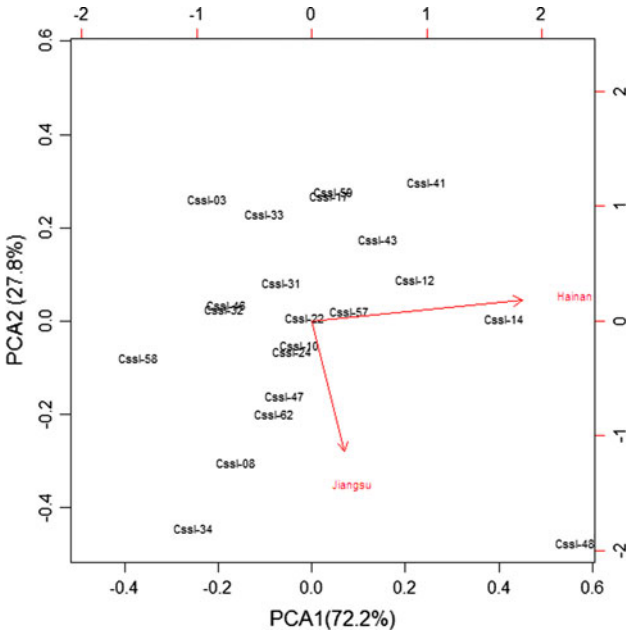

Fig. 5 Biplot based on mean substitution effects on protein contents in two contrasting environment groups

subspecies. Field observations for other key agronomic traits suggested that it possible to develop cultivars with acceptable agronomic performance and, further, enhanced PC directly using the SCSS lines containing desirable QTLs. We are developing populations by crossing superior lines (CSSL-48, CSSL-14 and CSSL-08) to pyramid the positive alleles. These populations will also be used to investigate the possible epistatic interactions between QTLs.

The segment of chromosome 8 contained in CSSL-48 had the largest effect in most of testing environments, suggesting the presence of QTL with a large effect. Moreover, QTL for PC has never been identified on chromosome 8. Therefore, we are developing a secondary population between CSSL-48 and the recurrent parent to fine map this region of chromosome 8. Similarly, secondary populations will also be developed to fine map QTLs contained in CSSL-08 and CSSL-14.

The eight testing environments were grouped into two independent types of environments, that is, Jiangsu and Hainan. This is consistent with the different climate types: Jiangsu has a temperate climate type, while Hainan has a tropical climate type. The two environment types were independent of each other. This implies that (1) it is possible to select substitutions that are good for both environment types; and (2) substitutions with a desired effect in one environment type may not necessarily have the desired effect in the other. The substitution in CSSL-48 increased PC at both of the environment types. Substitution in CSSL-14 increased PC at the Hainan environment, but had minimal effect at the Jiangsu environment. Conversely, substitution in CSSL-08 had no effect at the Hainan environment but had a large positive effect at the Jiangsu environment.

Breeding for specific adaptation where the intended commercial target area is sufficiently large to make it economically worthwhile should be effective and efficient, because many QTLs only have a desirable effect at one of the environments. Both Jiangsu and Hainan are important rice production regions in China. The growing areas in each of the regions are large enough to support separate breeding programs. In fact, for both Hainan and Jiangsu there are several rice breeding programs. Although it may be argued that this is not an efficient way to use the resources for breeding, it does make it possible to capture genetic gain from better exploring specific adaptation. The segments in CSSL-08, CSSL-34, CSSL-48 and CSSL-62 could be explored to enhance PC for the Jiangsu environment, while those in CSSL-14, CSSL-17, CSSL-41, CSSL-43 and CSSL-48 for the Hainan environment. Since only one testing site and two growing seasons were sampled from Hainan, it may not represent the target population of environments of breeding programs in Hainan. The lines showed significantly enhanced PC should be further tested in more sites and years in Hainan.

References

Aluko G, Martinez C, Tohme J, Castano C, Bergman C, Oard JH (2004) QTL mapping of grain quality traits from the interspecific cross *Oryza sativa* × *O. glaberrima*. Theor Appl Genet 109:630–639

Bartlett MS (1937) Properties of sufficiency and statistical tests. Proc R Stat Soc Ser A 160:268–282

Dunnett CW (1964) New tables for multiple comparisons with a control. Biometrics 20:482–491

Eshed Y, Zamir D (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics 141:1147–1162

Frary AN, Nesbitt TC, Frary AM, Grandillo S, Knaap EVD, Cong B, Liu JP, Meller J, Elber R, Alpert KB, Tanksley SD (2000) *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. Science 289:85–88

Gabriel KR (1971) The bi-plot-graphical display of matrices with application to principal component analysis. Biometrika 58:453–467

Howell PM, Marshall DF, Lydiate DJ (1996) Towards developing intervarietal substitution lines in *Brassica napus* using marker-assisted selection. Genome 39:348–358

Hu ZL, Li P, Zhou MQ, Zhang ZH, Wang LX, Zhu LH, Zhu YG (2004) Mapping of quantitative trait loci (QTLs) for rice protein and fat content using doubled haploid lines. Euphytica 135:47–54

Kempton RA (1984) The use of bi-plots in interpreting variety by environment interactions. J Agric Sci 103:123–135

Kubo T, Aida Y, Nakamura K, Tsunematsu H, Doi K, Yoshimura A (2002) Reciprocal chromosome segment substitution series derived from *Japonica* and *Indica* cross of rice (*Oryza sativa* L.). Breed Sci 52:319–325

Lin HX, Yamamoto T, Sasaki T, Yano M (2000) Characterization and detection of epistatic interactions of 3 QTLs, *Hd-1*, *Hd-2* and *Hd-3*, controlling heading date of rice using nearly isogenic lines. Theor Appl Genet 101:1021–1028

Nadeau JH, Singer JB, Matin A, Lander ES (2000) Analysing complex genetic traits with chromosome substitution strains. Nat Genet 24:221–225

Perez CM, Juliano BO, Alcantara JM, Cassman KG (1996) Effects of late nitrogen fertilization application on head rice yield, protein content and grain quality of rice. Cereal Chem 73:556–560

Shenoy VV, Seshu DV, Sachan JKS (1991) Inheritance of protein per grain in rice. Indian J Genet 52:214–220

Shi CH, Zhu J, Yang XE, Yu YG, Wu J (1999) Genetic analysis for protein content in *indica* rice. Euphytica 107:135–140

Shimodaira H (2004) Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. Ann Stat 32:2616–2641

Tan YF, Sun M, Xing YZ, Hua JP, Sun XL, Zhang QF, Corke H (2001) Mapping quantitative trail loci for milling quality, protein content and color characteristics of rice using a recombinant inbred line population derived from an elite rice hybrid. Theor Appl Genet 103:1037–1045

R Development Core Team (2008) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org

Tsunematsu H, Yoshimura A, Harushima Y, Nagamura Y, Kurata N, Yano M, Sasaki T, Iwata N (1996) RFLP framework map using recombinant inbred lines in rice. Breed Sci 46:279–284

Wan XY, Wan JM, Su CC, Wang CM, Shen WB, Li JM, Wang HL, Jiang L, Liu SJ, Chen LM, Yasui H, Yoshimura A (2004) QTL detection for eating quality of cooked rice in a population of chromosome segment substitution lines. Theor Appl Genet 110:71–79

Ward JH (1963) Hierarchical grouping to optimize an objective function. J Am Stat Assoc 58:236–244

Williams WT (1976) Pattern analysis in agricultural science. Elsevier, Amsterdam

Yan W (2001) GGEBiplot—a Window application for graphical analysis of multi-environment trail data and other types of two-way data. Agron J 93:1111–1118

Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y, Sasaki T (2000) *Hd-1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS*. Plant cell 12:2473–2483

Ye G, Smith KF (2010) Marker-assisted gene pyramiding for cultivar development. Plant Breed Rev 33:219–256