BMC Genomic Data

RESEARCH

Open Access

Correlation of rice yield based on RILs population QTL analysis

Junrong Liu[1†], Xinyi Lou[1†], Lin Zhang[1], Tiangang Hou[2], Xin Xin[1], Yan Wang[1], Shu Wang[1], Yuancai Huang[1], Chanchan Zhou[1], Baoyan Jia[1*] and Yue Feng[3*]

Abstract

Rice production has been a primary concern in crop quality breeding. In this study, India japonica variety M494 and indica variety Z9B were used as parents. Hybridization and selfing were conducted to obtain recombinant inbred lines (RILs) as the experimental material. The F_3 and F_7 populations were analyzed to determine six yield-related traits, including panicle length, effective panicle number, number of grains per panicle, seed setting rate, yield per plant, and grain density. QTL mapping of rice yield-related traits and tillering angle was performed using the SSR molecular marker linkage map, resulting in the identification of 19 QTLs controlling panicle length, grain number per panicle, effective panicle number, seed setting rate, grain density.

Additionally, multiple regression analysis and path analysis were employed to investigate the relationship between different agronomic traits and rice yield in the F_7 population. An optimal regression equation, $Y_{YPP} = -24.515 + 0.694X_{PL} + 1.273X_{PN} + 0.007X_{PPG} + 18.981X_{SSR}$ was derived, and it was concluded that SSR was the trait with the greatest impact on YPP, followed by PL.

Keywords Rice, Yield related traits, QTL, Multiple regression analysis, Path analysis

Introduction

The effective panicle number of rice and grain number per panicle are both major factors affecting rice yield [9]. Generally, the effective panicle number of rice shows a proportional relationship with production, while grain number per panicle, panicle length, and grain density are closely related.

Several studies have extensively reported on important agronomic traits and QTLs in rice and other crops. The application of molecular markers for QTL analysis in tomatoes followed later [15]. Effective panicle number, which is influenced by the number of tillers and individual spikelets, has a direct impact on overall production. Therefore, studying the QTLs controlling effective panicle number is of great importance. For instance, through the use of the NIL-F_2 (F_3) population, the QTL-*qPN1* controlling effective panicle number was successfully localized to a specific 34.4 kb region on chromosome 1

†Junrong Liu and Xinyi Lou contributed equally to this work.

*Correspondence:
Baoyan Jia
nxyjby2004@syau.edu.cn
Yue Feng
fy_555500@163.com
[1]Agricultural College, Shenyang Agricultural University, Shenyang, Liaoning 110866, China
[2]College of Engineering, Shenyang Agricultural University, Shenyang 110866, China
[3]Chinese National Center for Rice Improvement and State Key Laboratory of Rice Biology, China National Rice Research Institute, Hangzhou, Zhejiang 311401, China

of rice [33]. Similarly, using NILs, researchers were able to locate the micro-effect QTL-*qHd1* controlling rice heading date to a 95.0 kb region on the long arm of chromosome 1. This QTL not only affects heading date but also influences the number of effective panicles, grains per panicle, and yield per plant [3]. The number of grains per panicle, which is influenced by the number of primary and secondary branches, spikelets on branches, and seed setting rate, is a crucial trait affecting rice yield. In a notable study, a research team successfully localized a major QTL, *qGn1a* (*Grain number 1a*), which controls the number of grains per panicle in rice, within a 6.3 kb region on chromosome 1 [2]. Numerous QTLs regulating grain number per spike have been reported both domestically and internationally, including QTL-*gpa7*, QTL-*qSPP7*, QTL-*An-1*, QTL-*GNP1*, and QTL-*LF1* [10, 24, 27, 29, 30].

This study utilized the Indian japonica rice variety M494 as the primary material and performed QTL identification and genetic analysis on yield-related traits of rice. The analysis was conducted using a recombinant inbred line population obtained through hybridization with the indica rice Z9B. Multiple regression analysis and path analysis were employed to construct a regression model for the F_7 population of the recombinant inbred lines, enabling an analysis of the influence of different agronomic traits on individual plant yield. The research employed molecular biotechnology to conduct a comprehensive investigation into the genetic patterns associated with rice yield-related traits, including the identification of molecular markers linked to these traits. The findings lay the groundwork for subsequent fine mapping and cloning efforts, while also providing a theoretical foundation and reference for genetic enhancement of relevant traits and the cultivation of new rice varieties with high yield and quality.

Materials and methods

Construction of RILs

The parental rice varieties utilized in this study were the Indian japonica rice variety M494 and indica rice variety Zhong9B (Z9B). These varieties were obtained from the National Rice Germplasm Bank of China National Rice Research Institute.

Table 1 The agronomic traits, abbreviations, and types of traits involved in this study

Traits	Abbreviation	Traits type
panicle length	PL	quantitative trait
effective panicle number	PN	quantitative trait
the number of grains per panicle	PPG	quantitative trait
seed setting rate	SSR	quantitative trait
yield per plant	YPP	quantitative trait
grain density	GD	quantitative trait

The population utilized in this study consisted of recombinant inbred lines (RILs) derived from the hybridization and subsequent selfing of M494 and Z9B. Specifically, the F_3 and F_7 generations of the RIL population were used. The RIL population consisted of a total of 144 lines.

For the QTL mapping of rice yield-related traits, the F_3 and F_7 generations of the RIL population were employed.

Furthermore, the F_7 generation of the RIL population was used for multiple regression analysis and path analysis of different agronomic traits in relation to individual plant yield.

Rice cultivation

The F_3 population was cultivated at the China National Rice Research Institute Hangzhou Fuyang Experimental Base during the summer of 2018. Subsequently, the F_7 population was planted at the China National Rice Research Institute Hainan Lingshui Experimental Base in the winter of 2021.

A completely randomized block design was employed in this study, with three replicates. Each replicate consisted of six rows, and within each row, there were six individual plants. The spacing between plants was set at 20 cm × 25 cm. The field management followed a conventional management mode.

Measuring item

QTL mapping of yield related traits in rice was conducted using F_3 and F_7 recombinant inbred line populations to classify and statistically analyze six differential traits, namely panicle length (PL), effective panicle number (PN), grains per panicle (PPG), seed setting rate (SSR), grain density (GD), and yield per plant (YPP).

Multiple regression analysis and path analysis were performed using the F_7 recombinant inbred line population to investigate the relationships between different agronomic traits and single plant yield. A total of six differential traits were considered in the analysis, including PL, PN, grains per panicle (PPG, SSR, GD, and YPP. The statistical analysis aimed to assess the impact of these traits on single plant yield. Refer to Table 1 for detailed information on these traits (Table 1).

Determination method

For the investigation of yield-related traits, the following methodology was employed. After rice maturity, the plants were harvested on a per-plant basis, and various agronomic traits related to yield were examined. Prior to threshing, measurements were taken for panicle length, effective panicle number, and grain number per panicle. Three replicates were set, and the average values were recorded. Following threshing, the yield per plant was evaluated, and calculations were made for the seed

setting rate and seed density. The number of grains per panicle was measured using a fully automatic seed counting analyzer (SC-G type, Hangzhou Wanshen Co., Ltd.). The seed setting rate per plant was calculated by dividing the actual number of grains per plant by the total number of grains per plant, while the seed density was calculated by dividing the number of grains per panicle by the length of the panicle. Data statistics were performed using Excel, and correlation analysis, t-tests, and frequency distribution maps between traits were conducted using IBM SPSS Statistics 26 and GraphPad Prism 8.

Selection and identification of molecular markers

In the preliminary work, the researchers of the research group utilized the Gramene website (http://www.gramene.org/) to select a total of 562 SSR markers with good genome coverage for detecting polymorphism between the parents Z9B and M494. Among these markers, 114 showed polymorphism between the two parents. From these, 85 molecular markers that were evenly distributed across the 12 chromosomes were further screened. These 85 SSR markers were then used for genotype identification of the F_3 and F_7 populations, which consisted of 144 strains. The distribution range of polymorphic markers on each chromosome varied from 5.88% to 14.12%, with an average of 8.33%.

DNA isolation, PCR amplification, and electrophoresis.

Genomic DNA from the parental and RIL populations was extracted by CTAB. Take 0.1 g of fresh leaf tissue, grind in liquid nitrogen and add 1 mL CTAB of extraction buffer (2% CTAB, 100 mM Tris–HCl pH 8.0,20 mM EDTA, 1.4 M NaCl) in a 65℃ water bath for 1 h. After centrifugation, the supernatant was removed, an equal volume of chloroform-isoamyl alcohol (24:1) was added, and isopropanol precipitated DNA. The DNA was washed in 70% ethanol and then dissolved in TE buffer.

PCR reaction system (20 μ L): 10 Buffer 2 μL, dNTPs 0.2 mM, 0.5 μ M each of primers, 1 U of Taq enzyme, and 50 ng of DNA template. Amplification procedure: 94℃ pre-denaturation for 5 min; 94℃ 30 s, 55℃ 30 s, 72℃ 1 min, 35 cycles; 72℃ extension for 10 min. The amplified products were visualized by 8% non-denaturing polyacrylamide gel electrophoresis (voltage 120 V for 2 h) and stained with silver nitrate.

Construction of genetic map and QTL analysis

The clear bands observed in the gel map were recorded, and the analyzed genotype data were statistically organized. The Mapmaker/Exp3.0 software was used to construct a molecular marker genetic linkage map based on this data. The recombination rate between markers was converted into genetic map distance (cM) using the Kosambi function. Using the constructed molecular marker genetic map, QTL mapping of different agronomic traits was performed using Windows QTL Cartographer 2.5 software. A LOD score of ≥ 2.5 for the marker interval was considered indicative of potential QTLs controlling the trait in that interval. Additionally, the contribution rate (R^2) and additive effects of each trait were calculated, and QTLs with $R^2 \geq 3$ were selected as the results and given specific names [12].

Multiple regression analysis and path analysis

Using Excel software, statistical analysis was performed on the F7 population to calculate the average, standard deviation, and coefficient of variation for the six phenotypes, including PL, PN, PPG, SSR, GD, and YPP. SPSS Statistics 26 and GraphPad Prism 8 were used for further analysis. Scatter plots were created to visualize the relationships between variables, and Pearson correlation analysis [14] was conducted to examine the correlations between different traits. Univariate linear regression analysis was performed to assess the relationship between individual plant yield and each agronomic trait. Multiple regression analysis [16] was used to analyze the combined effects of multiple traits on yield. The regression model is: $Y_{YPP} = b_0 + b_{PL}X_{PL} + b_{PN}X_{PN} + b_{PPG}X_{PPG} + b_{SSR}X_{SSR}$, where Y is the dependent variable, X is the independent variable, b0 is a constant term, and b is the regression coefficient of each variable. Additionally, path analysis [8] was employed to explore the direct and indirect effects of different agronomic traits on yield.

Results and discussion

Performance of yield correlation data among parents, F_3, and F_7 populations

The six differential traits, namely PL, PN, PPG, SSR, GD, and YPP, were classified and statistically analyzed using the F_3 and F_7 recombinant inbred line populations. The high-value parent for PL, PN, TGW, and YPP was found to be Z9B, while the high-value parent for PPG, SSR, and GD was M494. Significant differences were observed among the six traits of the parents. In general, the average values of these traits in the F_3 population were higher than those in the F_7 population. Furthermore, the variation range of each trait in both populations exceeded the phenotypic values of the parents (Table 2). The frequency distribution of phenotypes indicated that all traits exhibited a super parental advantage (Figs. 1 and 2).

Based on these results, it can be inferred that PL, PN, PPG, YPP, SSR, and GD possess genetic characteristics of quantitative traits, which make them suitable for further QTL analysis.

Table 2 Phenotypic values of yield-related traits in parents, F_3 and F_7 populations

Traits	Generations	F_3 and F_7 RIL Populations				Parents	
		Mean	Range	Skewness	Kurtosis	M494	Z9B
PL(cm)	F_3	22.93	19.2–29.1	0.30	1.18	24.4	26.2**
	F_7	20.35	15.0–28.0	0.55	−0.32		
PN	F_3	9.94	2.0–21.0	0.52	0.55	10	12**
	F_7	8.24	3.0–15.0	0.34	−0.45		
PPG	F_3	171.90	88.0–280.0	0.29	−0.31	206.4**	155.2
	F_7	159.86	75.0–255.0	0.60	−0.18		
SSR	F_3	0.62	0.24–0.94	−0.27	0.35	0.75**	0.68
	F_7	0.66	0.23–0.96	−0.89	0.87		
GD	F_3	7.46	4.3–10.6	0.13	−0.41	8.3**	6.2
	F_7	7.35	3.4–13.9	0.49	−0.04		
YPP(g)	F_3	19.97	3.94–38.64	0.55	0.06	17.33	20.67**
	F_7	16.64	4.42–29.23	−0.19	−0.61		

**indicates a significant at the 0.01 level

QTL Identification of yield related traits in F_3 and F_7 populations

A total of 10 QTLs were detected on chromosomes 1, 2, 3, 4, and 10 of rice using the F_3 population to control the PPG, SSR, PL, TGW, and GD. It includes 1 QTL for PL (*qPL2*), 4 QTLs for PPG (*qPPG1-1*, *qPPG2-1*, *qPPG*, and *qPPG4*), 1 QTL for SSR (*qSSR1*), and 4 QTLs for GD (*qGD1*, *qGD3*, *qGD4*, and *qGD10*) (Fig. 3, Table 3).

A total of 12 QTLs were detected on chromosomes 1, 2, 3, 8, 10, and 11 of rice using the F_7 population to control PL, PN, PPG, GD, and TGW. There are 3 QTLs for PL (*qPL2*, *qPL10*, and *qPL11*), 4 QTLs for PPG (*qPPG1-2*, *qPPG2-2*, *qPPG3*, and *qPPG11*), 3 QTLs for PN (*qPN2*, *qPN8*, and *qPN10*), and 2 QTLs for GD (*qGD2* and *qGD3*) (Fig. 3, Table 3).

In summary, a total of 19 QTLs for yield related traits in rice were identified using F_3 and F_7 populations, among which 3 QTLs were detected repeatedly (*qPL2*, *qPPG3*, and *qGD3*) (Fig. 3, Table 3).

Multiple regression analysis of different agronomic traits and yield per plant in F_7 population

Six differential traits were classified and statistically analyzed using the F7 recombinant inbred line population, including PL, PN, PPG, SSR, GD, and YPP. The variation range of PL is 15.00–28.00 cm, the variation range of PN is 3.00–15.00, the variation range of PPG is 63.00–307.00, the variation range of SSR is 0.23–0.92, the variation range of GD is 3.43–13.90, and the variation range of YPP is 1.42–29.23 g. Among these traits, the CV value of YPP is the highest at 34.66%, while the CV value of PL is the lowest at only 8.84% (Table 4).

In the F_7 population, the YPP was highly significantly positively correlated with PPG, SSR, PN, and PL ($P < 0.01$), but not significantly correlated with GD (Table 5). Excluding GD, the scatter plot and univariate linear regression show a linear correlation trend between the PPG, SSR, PN, PL, and YPP (Fig. 4).

The regression model obtained through multiple linear regression analysis is $Y_{YPP} = -24.515 + 0.694X_{PL} + 1.273X_{PN} + 0.007X_{PPG} + 18.981X_{SSR}$, indicating that the variation in YPP is positively caused by four personality traits: PL, PN, PPG, and SSR. The dependent variable increases with the increase of a certain independent variable value; In the *t*-test results, the significance of the monitoring values for each trait was Sig < 0.05, indicating a relatively stable relationship between the independent and dependent variables expressed in the model; In the collinearity statistical test results, the VIF values are all close to 1, indicating that there is no obvious multicollinearity between the four traits, that is, the regression model obtained is stable and accurate (Tables 6, 7 and 8).

In the residual analysis, the vast majority of residual points are randomly distributed within the (−2, 2) interval, indicating that the residual sequence is randomly distributed, satisfying the assumptions of the regression model (Fig. 5a). Furthermore, the standard residual normal distribution diagram shows that the model residuals follow a roughly normal distribution (Fig. 5b). Based on these results, it can be concluded that the constructed model is stable and accurate. Based on the above results, the stability and accuracy of the model construction results can be demonstrated.The regression model coefficient of determination $R^2 = 0.725$, indicating that 72.5% of YPP variation was explained by PL, PN, PPG, SSR, and the residual effect ($1\text{-}R^2 = 0.275$) suggests that other undetected factors (such as environmental error or microeffect QTL) contributed 27.5% of the variation. Residual analysis showed that the data were fit to a normal distribution (Fig. 5b), DW = 1.672 (near 2), no autocorrelation, and the model was robust.

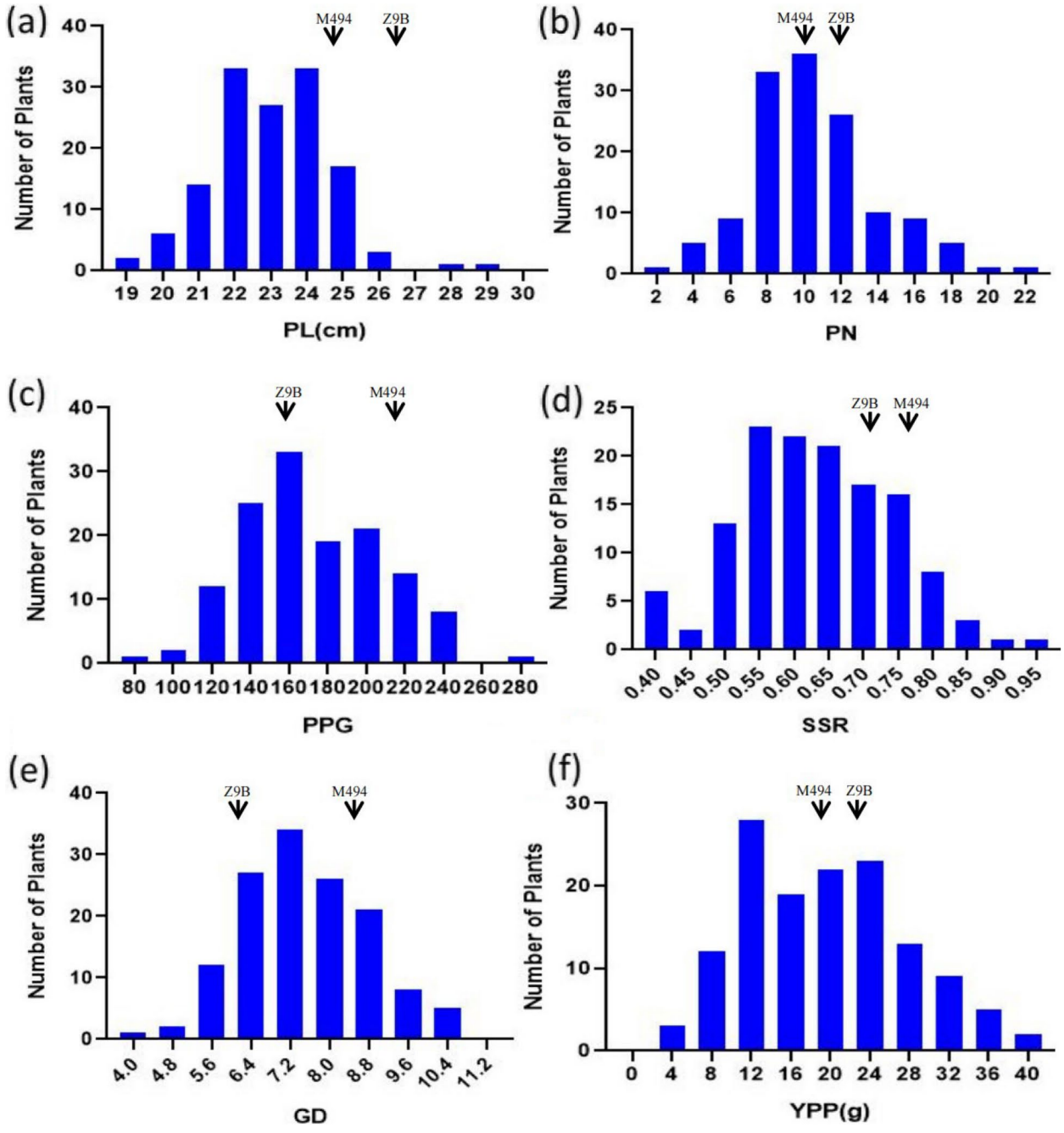

Fig. 1 Phenotype frequency distribution of yield-related traits in F_3 population of Z9B/M494. **a** panicle length **b** panicle number per plant **c** grains per panicle **d** seed setting rate **e** grain density **f** grain yield per plant

Path analysis of different agronomic traits and single plant yield in F_7 population

The analysis of path coefficients provides insights into the direct and indirect effects of independent variables on the dependent variable. The direct path coefficient represents the strength and direction of the relationship between an independent variable and the dependent variable, while the indirect path coefficient reflects the influence of one independent variable on the dependent variable through another independent variable. In the analysis results of path coefficients (Table 9), it is observed that all four traits have a positive impact on YPP, indicating that an increase in these traits leads to an increase in YPP. The largest direct impact on YPP is from SSR, with a path coefficient of 0.629, followed by PN with a path coefficient of 0.380. On the other hand, PPG has the smallest direct impact on YPP, with a path coefficient of 0.050. Additionally, the indirect path coefficients

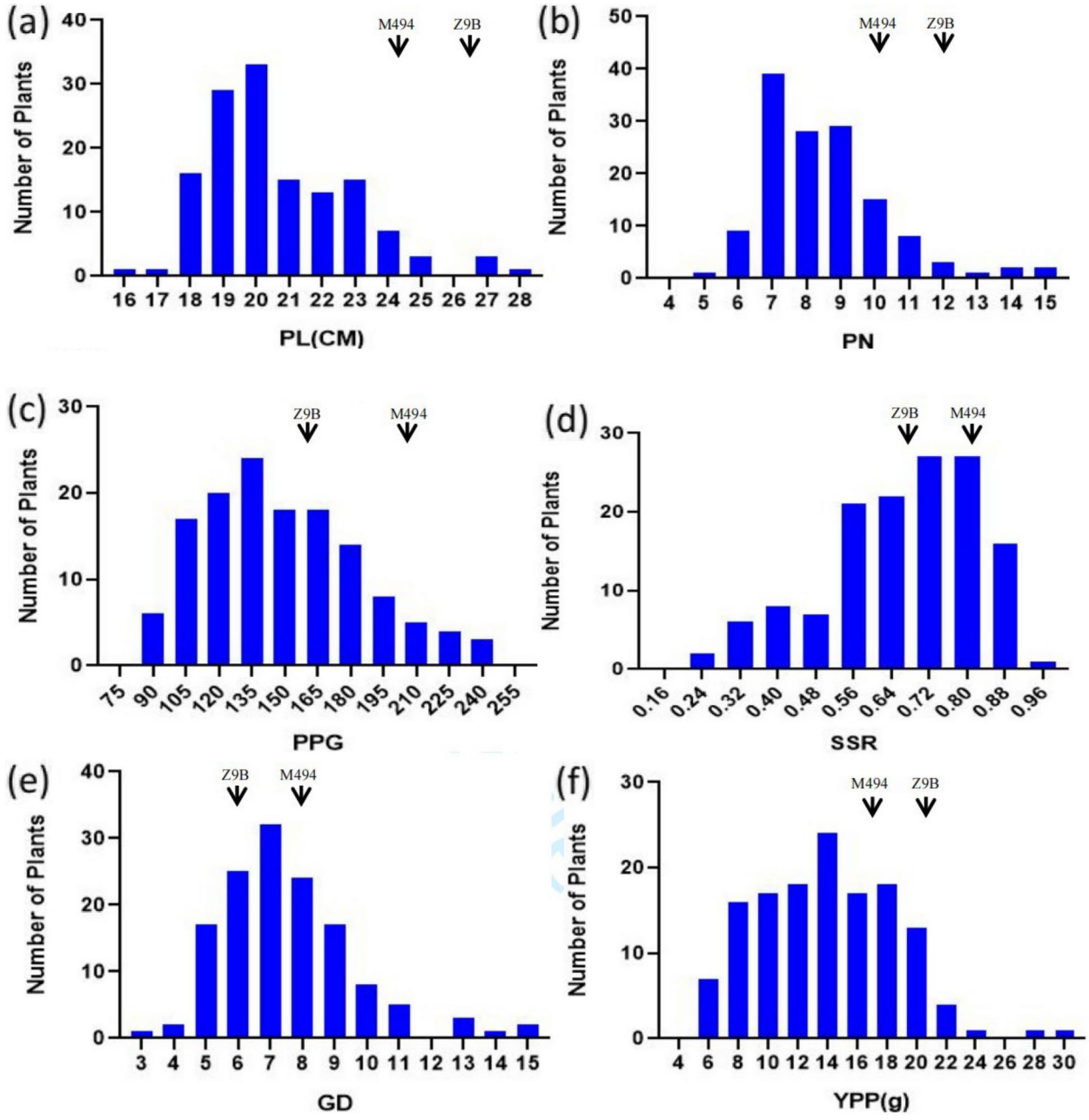

Fig. 2 Phenotype frequency distribution of yield-related traits in F_7 population of Z9B/M494. **a** panicle length **b** panicle number per plant **c** grains per panicle **d** seed setting rate **e** grain density **f** grain yield per plant

reveal the influence of one independent variable on YPP through another independent variable. For example, PL has the greatest positive impact on YPP through SSR, with an indirect path coefficient of 0.193. Conversely, PL has the greatest negative impact on YPP through PN, with an indirect path coefficient of −0.055. Considering the total path coefficients, which are the sum of the direct and indirect path coefficients, it can be concluded that SSR has the highest total path coefficient on YPP (0.740), indicating the strongest overall impact. PL also shows a substantial total path coefficient of 0.420 on YPP.

These findings demonstrate the importance and contributions of each trait to the variation in YPP. SSR and PL emerge as significant factors with strong direct effects on YPP, while PN and PPG also exhibit considerable direct impacts. The path coefficients provide valuable information for understanding the relationships and relative importance of these traits in influencing YPP.

Stability of QTL detection in different populations

The performance of quantitative traits is not only closely related to genes, but also influenced by environmental

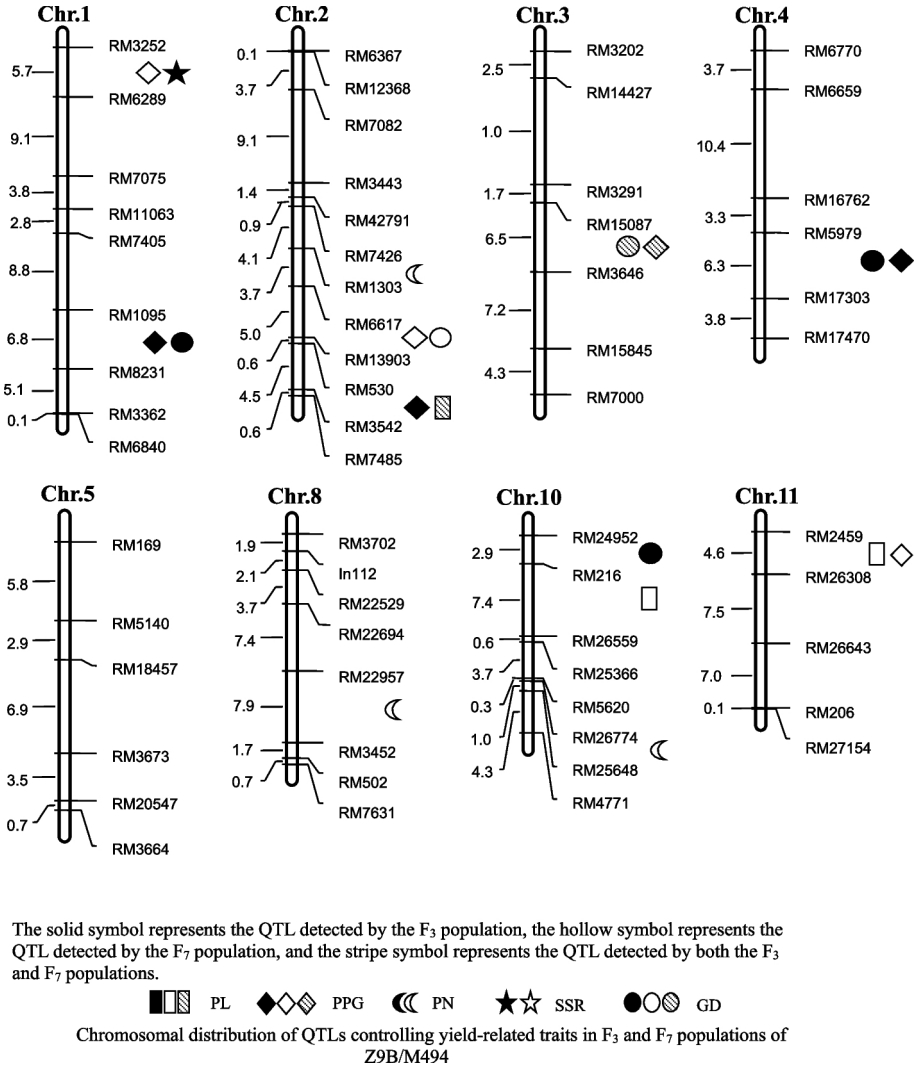

Fig. 3 Chromosomal distribution of QTLs controlling yield-related traits in F_3 and F_7 populations of Z9B/M494

changes, so the detection results of the same trait in different populations may not be consistent, or the detection results of the same population in different years or planting environments may also be inconsistent [1, 21, 31]. This study utilized a recombinant inbred line population derived from M494/Z9B to locate QTLs for six agronomic traits related to yield. A total of 19 QTLs were detected, distributed within 15 marker intervals on chromosomes 1, 2, 3, 4, 8, 10, and 11, respectively, controlling PL, PPG, PN, SSR, and GD. No QTLs were detected that controlled YPP. Among the detected QTLs, 10 QTLs were identified using the F_3 population, while 12 QTLs were identified using the F_7 population. Importantly, 3 QTLs were repeatedly detected in both populations, accounting for 30% of the total QTLs detected in the F_3 population and 25% of the total QTLs detected in the F_7 population. These QTLs were QTL-*qPL2* located on chromosome 2 controlling PL, QTL-*qPPG3* and QTL-*qGD3* located on chromosome 3 controlling PPG and GD (Fig. 3, Table 4). These 3 QTLs controlling different traits were jointly detected by two populations, indicating that they have stable genetic effects. In addition, we found that among the 3 QTLs detected repeatedly, *qPPG3*,and *qGD3* had a contribution rate of over 10.00% in both tests; Only *qPL2* had a contribution rate of 6.79% lower than 10.00% in the F_3 population, and the additive effects of these 3 QTLs measured in both populations were consistent. This can indicate that QTLs with high contribution rates can be stably expressed even in different environments [26, 28, 32]. These QTL loci with high phenotypic contribution rates and stable expression in different environments have greater value for molecular marker assisted selection breeding. The results in this study found that the QTLs controlling SSR and PN had low effect size and contribution rate, which may also be

Table 3 Summary of QTL identified in this study

QTL	Chr	Interval	Distance (cM)	LOD	Additive	Additive source	R^2	Population
qPPG1-1	1	RM1095-RM8231	6.8	4.81	23.496	Z9B	20.91	F_3
qGD1	1	RM1095-RM8231	6.8	4.48	0.681	Z9B	14.42	F_3
qSSR1	1	RM3252-RM6289	9.1	2.51	−0.009	M494	3.73	F_3
qPPG1-2	1	RM3252-RM6289	9.1	2.53	23.686	Z9B	19.87	F_7
qPL2	2	RM530-RM3542	4.5	2.84	−0.596	M494	6.79	F_3
				3.09	−1.041		16.8	F_7
qPPG2-1	2	RM530-RM3543	4.5	2.75	−10.871	M494	4.41	F_3
qPN2	2	RM7426-RM1303	4.1	3.06	−0.420	M494	4.00	F_7
qPPG2-2	2	RM6617-RM13903	5.0	3.28	−12.656	M494	4.93	F_7
qGD2	2	RM6617-RM13903	5.0	3.66	−0.719	M494	8.01	F_7
qPPG3	3	RM15087-RM3646	6.5	10.61	−30.289	M494	35.64	F_3
				3.48	−21.59		18.12	F_7
qGD3	3	RM15087-RM3646	6.5	14.33	−1.189	M494	45.2	F_3
				3.72	−0.913		16.35	F_7
qPPG4	4	RM5979-RM17303	6.3	2.78	−16.067	M494	9.10	F_3
qGD4	4	RM5979-RM17303	6.3	3.5	−0.593	M494	10.12	F_3
qPN8	8	RM22957-RM3452	7.9	2.65	−0.649	M494	9.04	F_7
q*GD10*	10	RM24952-RM216	2.9	3.07	−0.508	M494	7.40	F_3
qPL10	10	RM216-RM26559	7.4	2.67	2.894	Z9B	87.98	F_7
qPN10	10	RM26774-RM25648	1.0	3.40	−0.536	M494	9.57	F_7
qPL11	11	RM2459-RM26308	4.6	2.82	3.268	Z9B	82.12	F_7
qPPG11	11	RM2459-RM26308	4.6	2.6	−8.433	M494	3.56	F_7

Table 4 Statistical data of related quantitative traits in F_7 population of Z9B/M494

Content	PL(cm)	PN	PPG	SSR	GD	YPP/g
Max	28.00	15.00	307.00	0.92	13.90	29.23
Min	15.00	3.00	63.00	0.23	3.43	3.42
Mean	20.34	8.27	149.75	0.66	7.35	13.67
Standard Deviation	1.80	1.41	35.92	0.16	1.60	4.74
Coefficient of Variation	8.84%	17.08%	23.98%	23.71%	21.84%	34.66%

Table 5 Correlation coefficient among traits of Z9B/M494 F_7 population

Traits	PPG	SSR	YPP	PN	PL	GD
PPG	1					
SSR	0.178*	1				
YPP	0.233**	0.741**	1			
PN	−0.092	0.057	0.367**	1		
PL	0.387**	0.307**	0.420**	−0.146	1	
GD	0.921**	0.06	0.074	−0.037	0.013	1

**indicates a significant correlation at the 0.01 level

*indicates significant correlation at the 0.05 level

one of the reasons why stable QTLs were not detected between different populations of these two traits.

The pleiotropy of QTL

The study confirmed the existence of QTL pleiotropy, which refers to the simultaneous influence of multiple traits by QTLs on the same chromosome within the same interval. This phenomenon is common in animals and plants and can be attributed to various genetic factors such as single-cause pleiotropy, gene overlap, or gene linkage [6, 19, 23]. In the current study, out of the 19 QTLs detected on 7 chromosomes of rice, 14 QTLs were found to control two traits within the same chromosome and marker interval. This accounts for 73.7% of the total QTLs detected. These QTLs were distributed on chromosomes 1, 2, 3, 4, and 11 (Fig. 3, Table 4). Notably, significant correlations were observed between certain traits controlled by QTLs within specific intervals. For example, in the F_3 population, highly significant positive correlations were found between PPG and GD controlled

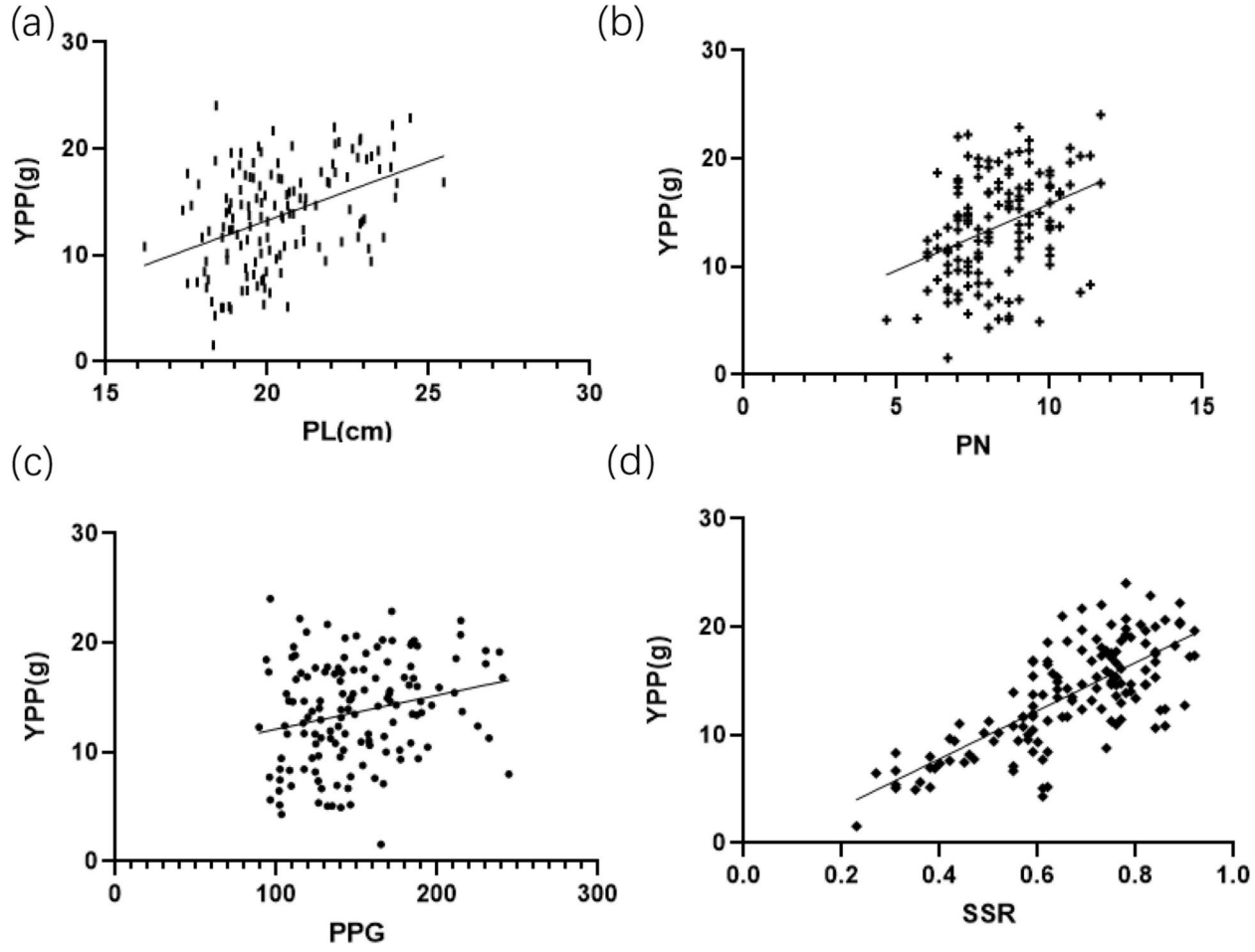

Fig. 4 Scatter plot of yield per plant and other 7 traits in F_7 population of Z9B/M494. X-axis: **a** panicle length **b** panicle number per plant **c** grains per panicle **d** seed setting rate, Y-axis:grain yield per plant

Table 6 Regression model analysis of different traits and yield per plant in rice

Model	Unnormalized coefficient		Normalization coefficient	*t*	Sig	Collinearity statistics	
	B	Standard error	Beta			Tolerance	VIF
	−24.515	2.952		−8.305	0.001		
PL	0.694	0.136	0.263	5.085	0.000	0.778	1.286
PN	1.273	0.156	0.380	8.177	0.000	0.967	1.034
PPG	0.007	0.007	0.050	1.002	0.038	0.843	1.186
SSR	18.981	1.460	0.629	13.005	0.001	0.891	1.122

Table 7 Test of fitting degree of prediction model between yield per plant and different traits

Model	*R*	R^2	Adjusted R^2	Error in standard estimation	Durbin-Watson value
	0.851[a]	0.725	0.717	2.53161	1.672

[a]Dependent variable: YPP

Table 8 F-test for optimal model

Model	Sum of squares	df	Mean square	Sig
Regression	2229.978	4	557.495	0.000[b]
Residual	845.992	132	6.409	
Total		136		

[b]Predictive variables: PL, PN, GPP, SSR, TGW

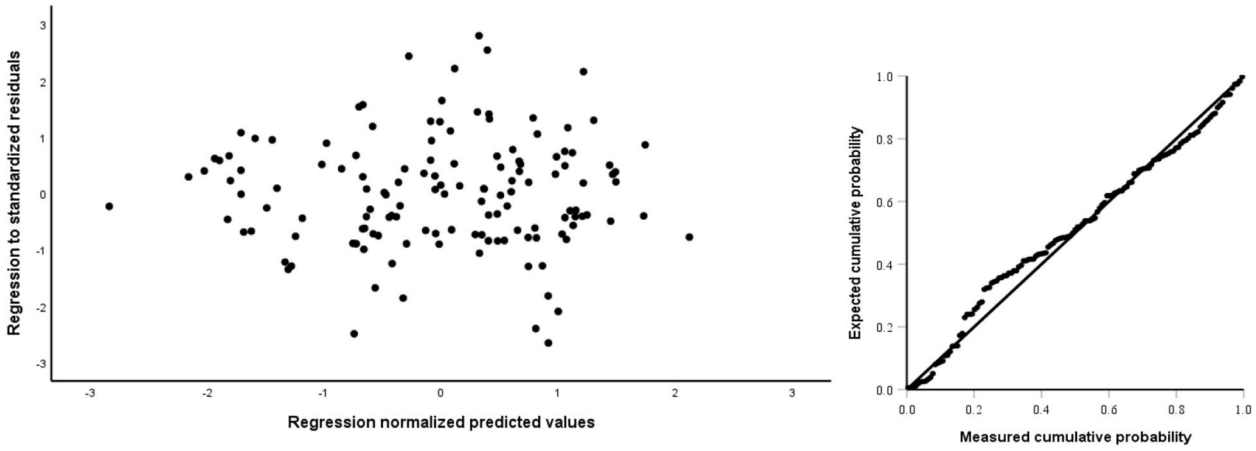

Fig. 5 Residual analysis of the multiple regression model. X-axis: **a** Scatter plot of residuals; **b** Normal p-p plot of regression standardized residual

Table 9 Path coefficient analysis

Traits	Path coefficient	Direct path coefficient	Indirect path coefficient			
			PL-YPP	PN-YPP	PPG-YPP	SSR-YPP
PL	0.420	0.263		−0.055	0.019	0.193
PN	0.373	0.380	−0.038		−0.005	0.036
PPG	0.229	0.050	0.102	−0.035		0.112
SSR	0.740	0.629	0.081	0.022	0.009	

Table 10 Correlation coefficient of yield-related traits in F_3 population of Z9B/M494

Traits	TGW	PPG	SSR	YPP	PN	PL	GD
TGW	1						
PPG	−0.407**	1					
SSR	−0.071	0.168	1				
YPP	0.15	0.386**	0.407**	1			
PN	0.083	0.176*	0.147	0.801**	1		
PL	0.013	0.688**	0.304**	0.588**	0.309**	1	
GD	−0.498**	**0.953****	0.082	0.239**	0.095	0.444**	1

**indicates a significant correlation at the 0.01 level

*indicates significant correlation at the 0.05 level

by two QTLs located within the RM1095-RM8231 interval of chromosome 1. Similarly, highly significant positive correlations were observed between PPG and PL controlled by two QTLs within the RM530-RM3543 interval of chromosome 2. Moreover, a highly significant positive correlation was detected between PN and GD controlled by two QTLs within the RM5979-RM17303 interval of chromosome 4. The F_7 population also exhibited highly significant positive correlations between PPG and GD controlled by two QTLs located within the RM6617-RM13903 interval of chromosome 2, as well as between PPG and PL controlled by two QTLs within the RM2459-RM26308 interval of chromosome 11. Additionally, two QTLs controlled by the RM15087-RM3646 interval on chromosome 3, detected in both F_3 and F_7 populations, showed highly significant correlations between PPG and GD (Tables 10 and 11).

Based on the results of this study, there is a highly significant correlation between different traits controlled by multiple QTLs within the same interval, which fully proves the relationship between the pleiotropy of QTLs and the significance between traits. However, further research is needed to determine whether it is caused by one cause pleiotropy, gene overlap, or gene linkage.

Analysis of QTL additive effects

Additive effect is the accumulation of genotype values of multiple minor genes that affect quantitative traits, and is the main component of trait phenotype values. This study found that among the 4 QTLs affecting PN and SSR, the additive genes all came from the parent M494; Among the 3 QTLs that affect PL, one QTL has an additive gene from M494 and 2 QTLs from Z9B; Among the 7 QTLs that affect PPG, 5 QTLs have additive genes from M494, and 2 QTLs come from Z9B; Among the 5 QTLs

Table 11 Correlation coefficient of yield-related traits in F_7 population of Z9B/M494

Traits	TGW	PPG	SSR	YPP	PN	PL	GD
TGW	1						
PPG	−0.433**	1					
SSR	−0.059	0.178*	1				
YPP	0.233**	0.233**	0.741**	1			
PN	0.117	−0.092	0.057	0.367**	1		
PL	−0.05	0.387**	0.307**	0.420**	−0.146	1	
GD	−0.450**	**0.921****	0.06	0.074	−0.037	0.013	1

**indicates a significant correlation at the 0.01 level

that affect GD, the additive genes of 4 QTLs all come from M494, and 1 QTL comes from Z9B. These results indicate that QTL enhancing genes that control PL, PPG, and GD exhibit a discrete distribution in M494 and Z9B. The results of additive effect analysis showed that among the 19 QTLs controlling yield related traits, M494 provided 14 additive alleles, accounting for 73.7% of the total. This indicates that using the japonica rice variety M494 for related research may provide new breeding or genetic resources for rice.

Comparison of QTL positioning results

This study detected a total of 19 QTLs controlling 5 agronomic traits using 8 agronomic trait data from different populations of RILs. Among them, 3 QTLs controlling effective panicle number were *qPN2*, *qPN8*, and *qPN10*; 7 QTLs controlling PPG, namely *qPPG1-1*, *qPPG1-2*, *qPPG2-1*, *qPPG2-2*, *qGPP3*, *qGPP4*, and *qGPP11*; 3 QTLs controlling PL are *qPL2*, *qPL10*, and *qP11*; 1 QTL for controlling SSR, namely *qSSR1*; 5 QTLs controlling GD, namely *qGD1*, *qGD2*, *qGD3*, *qGD4*, and *qGD10*.

In 2005, a Japanese research team located a QTL-*Gn1* that controls the number of grains per spike on chromosome 1, and further decomposed it into Gn1a and Gn1b loci using a NILs. Gn1a was finely located within a 6.3 kb region between 3A28 and 3A20 markers, with only one ORF, which is the OsCKX2 gene highly homologous to cytokinin oxidase and dehydrogenase [2]. It is located in the same interval as QTL-*qPPG1-2*, which controls the PPG, and QTL-*qSSR-1*, which controls the SSR, detected in this study and located within the RM3252-RM6289 marker on chromosome 1.

In 2002, three research groups successively published articles on the gene sd1, which is located on the first chromosome of rice and has the structure of controlling plant height and panicle type [13, 20, 22], and its location is very close to that of QTL-qPPG1-1 controlling PPG and QTL-qGD1 controlling GD, which are located in the marker RM1095-RM8231 on the 1 chromosome detected in this study.

In 2020, QTL-qTGW2, which controls grain width and grain weight, was mapped within 7.6 kb between P5 and P6 markers on chromosome 2 of rice by using RILs, and this region contains only one ORF [18], which is in the same interval as QTL-qPL2, which controls PL and QTL-qPPG2-1, which controls PPG, detected in this study. The main QTL-GS3, located in the pericentromeric region of chromosome 3, has a full-length cDNA of 956 bp and encodes a transmembrane protein consisting of 232 amino acids, which controls not only grain length and weight but also grain size in rice [5, 11], which is in the same interval as QTL-qPPG3 controlling PPG, and QTL-qGD3 controlling GD detected in this study, which are located in the rm15087-rm3646 marker on chromosome 3.

In 2008, a grain filling defective mutant gif1 was used to fine map the GIF1 gene in the 32 KB region between the CAPS4 and CAPS8 markers on chromosome 4, which controls sucrose transport unloading and grain filling during rice grain development, thereby affecting grain density [25]. The GIF1 gene is close to the QTL-qPPG4 controlling PPG and QTL-qGD4 controlling GD detected in this study, which is located within the RM5979-RM17303 marker on chromosome 4. LTBSG1, a developmental gene regulating panicle and grain in rice, was fine mapped between markers z10-13 and z10-12 on chromosome 10 [17], which is very close to the position of QTL-qPL10 controlling PL identified in this study and located within markers RM216-RM26559 on chromosome 10.

Some researchers have also detected QTLs for related traits on other chromosomes, such as gw2.1, a new allele of GW2 located on chromosome 2, affects rice glume shape by regulating cell proliferation. Compared with wild-type, the near isogenic line NIL-gw2.1 has increased grain length and width, thereby controlling thousand grain weight and effective panicles [7]. According to the results of previous experiments, many QTLs controlling traits can be detected repeatedly in different populations and environments. Most QTLs not only control one trait, but also control different traits of crops together with other genes. Because this study is currently in the initial mapping stage, it is necessary to expand the population to continue the fine mapping research in order to obtain more accurate chromosome QTL regions.

Regression analysis and path analysis

The coefficient of variation of quantitative traits can show the influence of a single quantitative trait on the outcome factors, and the greater the coefficient of variation, the greater the coverage of the trait on the factors affecting the outcome [4]. Referring to the statistics and analysis of crop traits, this study used the statistical data of PL, PN, PPG, SSR, GD, and YPP of RIL-F7 population derived from rice to use the significant relationship between different traits and YPP, excluding the traits (GD) that have no significant relationship with YPP, the multiple linear regression analysis and path analysis of the remaining four agronomic traits on YPP were calculated. The results showed that the variation of YPP was positively caused by four traits, including PL, PN, PPG, and SSR. The positive variation of SSR on YPP was the largest, followed by PL. Therefore, it can be explained that SSR and PL are very potential target traits in rice breeding.

The correlation analysis showed that the simple correlation coefficient between SSR and YPP was the largest, followed by PL. The path analysis also showed that the direct effect of SSR on YPP was the largest, and its indirect effect through other traits was also positive; Among the four traits, the direct path coefficient of PPG on YPP was the smallest, but the main effect of this trait on YPP was indirectly produced through the other three traits. Therefore, it can be seen that the SSR is the most important factor affecting the YPP of rice, followed by the PL. In the process of rice cultivation and breeding, selecting rice varieties with higher SSR and longer PL as the main direction of breeding and cultivation will help to increase the yield of rice.

In biological research, the target traits that are easy to measure or not easy to be damaged can be taken as dependent variables by using multiple regression models and path coefficients. For rice, yield is the top priority in the process of breeding and cultivation. Although many related studies have analyzed the relationship between quantitative traits, there are few studies on the construction of multivariate models of yield and other traits. Based on the construction of multivariate models, this study further used *t*-test, F-test, residual analysis and Multicollinearity diagnosis to test the accuracy and stability of the regression equation. The results showed that PL, PN, PPG, and SSR were statistically significant in the equation, and the effects of different traits on YPP were obtained by using path coefficient. For the regression equation, $R^2 = 0.717$, indicating that the five independent variables in the equation are most of the factors affecting the yield of a single plant, up to 71.7%, but there are also some other traits and factors that will also have a minor impact on the yield of a single plant, which needs further experimental verification.

QTL stability and environment interaction

In this study, 19 QTL were detected in the F3 and F7 populations, and of which 3 (qPL 2, qPPG 3, qGD 3) were repeated in both populations. The stability of these QTL may be related to less environmental interference from their genetic effects. For example, both qPPG 3 and qGD 3 contributed more than 10% and showed high additive effects (from parent M494) in both populations, suggesting that their genetic background may directly affect yield-related traits by regulating spike morphology (e. g., branch number) and grain arrangement density [2]. However, the contribution of qPL 2 was only 6.79% in the F3 population and increased to 16.8% in F7, which may be related to increased homozygosity in the population genetic background or enhanced environmental fitness [31]. In the future, the stability of these QTL should be verified in multiple environments, and their interaction mechanism with climate factors (e. g., light and temperature conditions) should be explored.

Genetic mechanism of pleiotropic QTL

In this study, 73.7% of QTL controlled multiple traits within the same marker interval, such as the RM 530-RM3543 interval of chromosome 2 regulating both PL and PPG. This pleiotropy may result from the following reasons:

Single-gene pleiotropy For example, genes that regulate spike length may indirectly change the grain alignment density (GD), by affecting spike shaft elongation.

Gene cluster linkage There may be multiple functional genes in the RM15087-RM3646 interval of chromosome 3, which regulate PPG and GD respectively, but are inheritance due to the close physical location. These two mechanisms need to be distinguished by fine mapping or transgenic validation. For example, the chromosome 3 region where qPPG 3 is located contains multiple candidate genes known to regulate spike grain number (e. g., OsSPL14), and whether it simultaneously affects GD deserves further investigation.

Possible cause for which no YPP-related QTL was detected

Although YPP was significantly associated with PL, PN, and SSR, a QTL directly controlling YPP was not detected in this study. This may be due to: trait complexity: YPP is regulated by multiple gene microeffects with a single QTL contribution below the detection threshold (LOD 2.5).

Epistatic effects Phenotypic variation in YPP may be driven by multiple QTL interactions (e. g., PLSSR) rather than a single primary QTL. In the future, the genetic network of YPP can be further analyzed by genome-wide

association analysis (GWAS) or machine learning-based multi-trait models.

QTL stability and environment interaction 19 QTL, 3 of which 3 (qPL 2, qPPG 3, qGD 3) were stably expressed in the dual population, with the highest contribution rate of 45.2% (qGD 3).

Collaborative improvement potential of pleiotropic QTL the key interval of chromosomes 2 and 3 can synchronously regulate PL, PG and GD to provide targets for molecular design breeding.

Practical significance of yield prediction model SSR and PL have the largest direct effect on single plant yield, and the model can guide the optimization of high-yield plant type.

Parental genetic contribution and breeding potential The high-frequency additive effect of M494 (73.7%) provides a new resource for indica-japonica hybrid breeding, and its QTL (such as qPPG 3) can break through the intersubspecies yield limit.

Abbreviations

PL	Panicle length
PN	Effective panicle number
PPG	The number of grains per panicle
SSR	Seed setting rate
YPP	Yield per plant
GD	Grain density

Acknowledgements

The authors thank you to the editors and all reviewers for their comments and suggestion. The authors also appreciate the people who patiently participated in the study.

Authors' contributions

Junrong Liu: Conceptualization; formal analysis; investigation; methodology; writing—original draft; writing—review and editing. Xinyi Lou: Conceptualization; data curation; formal analysis; investigation; methodology; writing—review and editing. LIn Zhang: Data curation; investigation; writing—review and editing. Yan Wang: Data curation; investigation; resources; writing— review and editing. Xin Xin: Data curation; investigation; resources; writing—review and editing.Yan WAng: Data curation; investigation; resources; writing—review and editing. Shu Wang: Data curation; investigation; resources; writing—review and editing. Yuancai Huang: Data curation; investigation; resources; writing—review and editing. Chanchan Zhou: Data curation; investigation; resources; writing—review and editing. Baoyan Jia: Data curation; funding acquisition; investigation; project administration; supervision; writing—review and editing. Yue Feng: Data curation; funding acquisition; investigation; project administration; supervision.

Funding

National Key R&D Program of China, 2023YFD2301603 and 2018YFD0300306.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Our study did not require further ethics committee approval as it did not involve animal or human clinical trials and was not unethical. In accordance with the ethical principles outlined in the Declaration of Helsinki, all participants provided informed consent before participating in the study. The anonymity and confidentiality of the participants were guaranteed, and participation was completely voluntary. Not applicable.

Consent for publication

Consent for publication was obtained from the participants.

Competing interests

The authors declare no competing interests.

Received: 26 November 2024 / Accepted: 21 March 2025

Published online: 14 April 2025

References

1. Akond Z, Alam MJ, Hasan MN, Uddin MS, Alam M, Mollah MNH. A Comparison on Some Interval Mapping Approaches for QTL Detection. Bioinformation. 2019;15(2):90–4. https://doi.org/10.6026/97320630015090.
2. Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T, Nishimura A, Matsuoka M. Cytokinin oxidase regulates rice grain production. Science (New York, NY). 2005;309(5735):741–5. https://doi.org/10.1126/science.1113373.
3. Chen JY, Guo L, Ma H, Chen YY, Zhang HW, Ying JZ, Zhuang JY. Fine mapping of qHd1, a minor heading date QTL with pleiotropism for yield traits in rice (Oryza sativa L.). Theor Appl Genet. 2014;127(11):2515–24. https://doi.org/10.1007/s00122-014-2395-7.
4. Ehdaie B, Waines JG. Genetic variation, heritability and path-analysis in landraces of bread wheat from southwestern Iran. Euphytica. 1989;41(3):183–90. https://doi.org/10.1007/BF00021584.
5. Fan C, Xing Y, Mao H, Lu T, Han B, Xu C, Zhang Q. GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor Appl Genet. 2006;112(6):1164–71. https://doi.org/10.1007/s00122-006-0218-1.
6. Hu SD, Luebberstedt T, Zhao GW, Lee M. QTL mapping of low-temperature germination ability in the maize IBM Syn4 RIL population. Plos One. 2016:11(3). https://doi.org/10.1371/journal.pone.0152795.
7. Huang JP, Chen ZM, Lin JJ, Guan BB, Chen JW, Zhang ZS, Huang RY. *gw2.1*, a new allele of *GW2*, improves grain weight and grain yield in rice. Plant Sci. 2002;325. https://doi.org/10.1016/j.plantsci.2022.111495.
8. Li ZX, Liu JL, You J, Li X, Liang ZS, Du JL. Proanthocyanidin Structure-Activity Relationship Analysis by Path Analysis Model. Int J Mol Sci. 2023;24(7). https://doi.org/10.3390/ijms24076379.
9. Lu Y, Chuan ML, Wang HY, Chen RJ, Tao TY, Zhou Y, Yang ZF. Genetic and molecular factors in determining grain number per panicle of rice. Front Plant Sci. 2022;13. https://doi.org/10.3389/fpls.2022.964246.
10. Luo JH, Liu H, Zhou TY, Gu BG, Huang XH, Shangguan YY, Han B. *An-1* Encodes a Basic Helix-Loop-Helix Protein That Regulates Awn Development, Grain Size, and Grain Number in Rice. Plant Cell. 2013;25(9):3360–76. https://doi.org/10.1105/tpc.113.113589.
11. Mao HL, Sun SY, Yao JL, Wang CR, Yu SB, Xu CG, Zhang QF. Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. Proc Natl Acad Sci USA. 2010;107(45):19579–84. https://doi.org/10.1073/pnas.1014419107.
12. McCouch SRJRGN. Report on QTL nomenclature. 1997;14:11–3.
13. Monna L, Kitazawa N, Yoshino R, Suzuki J, Masuda H, Maehara Y, Minobe Y. Positional cloning of rice semidwarfing gene, sd-1: rice "green revolution gene" encodes a mutant enzyme involved in gibberellin synthesis. DNA Res. 2002;9(1):11–7. https://doi.org/10.1093/dnares/9.1.11.
14. Norušis MJ. SPSS 14.0 guide to data analysis. Upper Saddle River: Prentice Hall; 2006.
15. Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature. 1988;335(6192):721–6. https://doi.org/10.1038/335721a0.
16. Petrie A, Bulman JS, Osborn JF. Further statistics in dentistry. Part 6: Multiple linear regression. Br Dental J. 2002;193(12):675–82. https://doi.org/10.1038/sj.bdj.4801659.
17. Qin R, Zeng DD, Yang CC, Akhter D, Alamin M, Jin XL, Shi CH. *LTBSG1*, a New Allele of *BRD2*, regulates panicle and grain development in rice by brassinosteroid biosynthetic pathway. Genes. 2008;9(6). https://doi.org/10.3390/genes9060292.

18. Ruan BP, Shang LG, Zhang B, Hu J, Wang YX, Lin H, Gao ZY. Natural variation in the promoter of *TGW2* determines grain width and weight in rice. New Phytol. 2020;227(2):629–40. https://doi.org/10.1111/nph.16540.
19. Samayoa LF, Malvar RA, Butrón A. QTL for Maize Midparent Heterosis in the Heterotic Pattern American Dent x European Flint under Corn Borer Pressure. Front Plant Sci. 2017;8. https://doi.org/10.3389/fpls.2017.00573.
20. Sasaki A, Ashikari M, Ueguchi-Tanaka M, Itoh H, Nishimura A, Swapan D, Matsuoka M. Green revolution: a mutant gibberellin-synthesis gene in rice. Nature. 2002;416(6882):701–2. https://doi.org/10.1038/416701a.
21. Semeskandi MN, Mazloom P, Arabzadeh B, Moghadam MN, Ahmadi T. Application of correlation coefficients and principal components analysis in stability of quantitative and qualitative traits on rice improvement cultivation. Braz J Biol. 2023;84:e268981. https://doi.org/10.1590/1519-6984.268981.
22. Spielmeyer W, Ellis MH, Chandler PM. Semidwarf (sd-1), "green revolution" rice, contains a defective gibberellin 20-oxidase gene. Proc Natl Acad Sci USA. 2002;99(13):9043–8. https://doi.org/10.1073/pnas.132266399.
23. Sun ZZ, Yin XL, Ding J, Yu D, Hu M, Sun XW, Fang J. QTL analysis and dissection of panicle components in rice using advanced backcross populations derived from *Oryza Sativa* cultivars HR1128 and "Nipponbare." Plos One. 2017;12(4). https://doi.org/10.1371/journal.pone.0175692.
24. Tian F, Zhu Z, Zhang B, Tan L, Fu Y, Wang X, Sun CQ. Fine mapping of a quantitative trait locus for grain number per panicle from wild rice (Oryza rufipogon Griff.). Theor Appl Genet. 2006;113(4):619–29. https://doi.org/10.1007/s00122-006-0326-y.
25. Wang E, Wang J, Zhu X, Hao W, Wang L, Li Q, He Z. Control of rice grain-filling and yield by a gene with a potential signature of domestication. Nat Genet. 2008;40(11):1370–4. https://doi.org/10.1038/ng.220.
26. Wang ZF, Cheng JP, Chen ZW, Huang J, Bao YM, Wang JF, Zhang HS. Identification of QTLs with main, epistatic and QTL x environment interaction effects for salt tolerance in rice seedlings under different salinity conditions. Theor Appl Genet. 2012;125(4):807–15. https://doi.org/10.1007/s00122-012-1873-z.
27. Wu Y, Wang Y, Mi XF, Shan JX, Li XM, Xu JL, Lin HX. The QTL *GNP1* Encodes GA20ox1, Which Increases Grain Number and Yield by Increasing Cytokinin Activity in Rice Panicle Meristems. Plos Genetics. 2016;12(10). https://doi.org/10.1371/journal.pgen.1006386.
28. Xin F, Zhu T, Wei SW, Han YC, Zhao Y, Zhang DZ, Ding Q. QTL Mapping of Kernel Traits and Validation of a Major QTL for Kernel Length-Width Ratio Using SNP and Bulked Segregant Analysis in Wheat. Sci Rep. 2020;10(1). https://doi.org/10.1038/s41598-019-56979-7.
29. Xing YZ, Tang WJ, Xue WY, Xu CG, Zhang Q. Fine mapping of a major quantitative trait loci, qSSP7, controlling the number of spikelets per panicle as a single Mendelian factor in rice. Theor Appl Genet. 2008;116(6):789–96. https://doi.org/10.1007/s00122-008-0711-9.
30. Zhang T, Li YF, Ma L, Sang XC, Ling YH, Wang YT, He GH. *LATERAL FLORET 1* induced the three-florets spikelet in rice. Proc Natl Acad Sci USA. 2017;114(37):9984–9. https://doi.org/10.1073/pnas.1700504114.
31. Zhao XQ, Niu YN, Hossain Z, Shi J, Mao TT, Bai XD. Integrated QTL mapping, meta-analysis, and RNA-sequencing reveal candidate genes for maize deep-sowing tolerance. Int J Mol Sci. 2023;24(7). https://doi.org/10.3390/ijms24076770.
32. Zhu HT, Li Y, Liang JY, Luan X, Xu P, Wang SK, Liu GF. Analysis of QTLs on heading date based on single segment substitution lines in rice (*Oryza Sativa* L.). Sci Rep. 2018;8. https://doi.org/10.1038/s41598-018-31377-7.
33. Zhu JY, Zhou Y, Liu YH, Wang ZD, Tang ZX, Yi CD, Liang GH. Fine mapping of a major QTL controlling panicle number in rice. Mol Breeding. 2011;27(2):171–80. https://doi.org/10.1007/s11032-010-9420-6.

Publisher's Note