

Genetic Basis of Relationship Between Yield and its Component Traits in Rice Revealed by Genome-Wide Association and Mendelian Randomization Study

Jing Su (✉ sujing0703@126.com)

Hunan Agricultural University <https://orcid.org/0000-0003-4272-4974>

Kai Xu

Hunan Agricultural University

Chao Wu

Hunan Agricultural University

Zirong Li

Hunan Agricultural University

Zhongli Hu

Wuhan University

Xingfei Zheng

Hubei Academy of Agricultural Sciences

Sufeng Song

Hunan Agricultural University

Wenbang Tang

Hunan Agricultural University

Zhonghai Tang

Hunan Agricultural University

Lanzhi Li (✉ lancy0829@163.com)

Hunan Agricultural University

Original article

Keywords: Rice yield, Component traits, Meta-GWAS, Mendelian randomization

Posted Date: May 13th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-28492/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Rice yield has a complex genetic architecture, which mainly determined by its three component traits: the number of grains per panicle (GPP), kilo-grain weight (KGW) and tillers per plant (TP). Exploring ideotype breeding based on selection for genetically less complex component traits is an alternative route for further improving rice production. Thus, it is important that studying the genetic basis of relationship between rice yield and component traits and clarifying the effects of each component trait on yield.

Main text

In this study, we carried out meta-analyses of genome-wide association study (Meta-GWAS) with two population (575 + 1495 F₁) in different environment for yield and its three component traits in rice. Totally, 3589 significant loci for three components traits were detected, while only 3 significant loci for yield were detected. It indicated that rice yield is mainly controlled by minor-effect loci and hardly to be identified. Selecting quantitative trait locus (QTL)/gene affected component traits to further enhance yield is recommended. A Mendelian randomization (MR) design was adopted to further estimate the causal relationship between rice yield and its component traits. Both GPP (Beta=0.086, 95% CI: 0.030~0.141, $P=0.003$) and TP (Beta=1.865, 95% CI: 1.035~2.694, $P<0.0001$) has a positive causal relationship with yield, but no significant relationship between KGW and yield (Beta=0.456, 95% CI: -0.119~1.031, $P=0.120$) was observed. Additionally, TP (Beta=1.865) has a greater effect on yield than GPP (Beta=0.086). Four significant loci for TP and GPP with indirect effect on yield were identified. Pyramiding superior alleles of the four loci revealed improved yield. A combination of direct and indirect effects may better contribute to the yield potential of rice.

Conclusions

Our results suggested rice production would improve by ideotype breeding based on selection for GPP and TP. By studying the nature and strength of the relationship between yield and its components, provide genetic insights for further improving rice yield potential.

Background

Rice is a staple food crop for about half of world people. Improving rice productivity has been the main goal of rice breeding research since the growth of population and the loss of arable land. However, rice yield has a complex genetic architecture, which determined by various physiological processes changing temporally during the growing period. These processes often matched the yield component traits that are genetically less complex than yield (Kadam et al. 2018). Therefore, selecting the component traits of yield was proposed as a complementary route for further improving the rice production, which also has been emphasized by national and international rice breeding programs (Huang et al. 2009). Studying the

causal relationship between rice yield and component traits, and clarifying the effects of each component trait on yield, will provide new clues for enhancing rice yield potential.

Rice yield is a very complex agronomic trait mainly determined by its three component traits: the number of grains per panicle (GPP), kilo-grain weight (KGW) and tillers per plant (TP), which are typical quantitative traits that are affected by multiple genes and the environment, with a low heritability (Xing et al. 2010). With the development of high-throughput technology, a large number of genes/quantitative trait loci (QTLs) of the three component traits were identified using QTL mapping and genome-wide association study (GWAS) methods (Le et al. 2019; Wang et al. 2019). At the end of 2019, 209, 223, and 239 genes/QTLs for GPP (TO: 0000445), KGW (TO: 0000382) and TP (TO: 0000152) were identified respectively (<http://www.gramene.org/>), which densely distributing across the 12 chromosomes. Some of them have been applied in rational design of super rice by marker-assisted selection (MAS) breeding, in which multiple defined genes with superior alleles pyramided to increase rice yield (Qian et al. 2016). Liu et al. (2012) introduced the *DEP1* and *Gn1* genes introduced into the restorer line *93-11*, then the yield of the *DEP1 / Gn1-9311* line was significantly improved, due to resource allocation improved. In 2020, Wang et al. (2020) compared the transgenic lines with *GNP1* or *NAL1* to the transgenic lines with both genes. They found the latter had a significantly higher yield, it indicated the two genes combinations may enhance the source-sink relationship. In above researches, only a small number of genes combined for super rice breeding, if more genes are selected for pyramiding, the trade-offs between different traits need to carefully consider (Zeng et al. 2017). Therefore, understanding the nature and strength of the relationship between yield and its components will be helpful for efficient gene selection in MAS breeding (Li et al. 2019).

The relationship between rice yield and its components investigated by various researchers with different materials and methods, but they were inconsistent. In Huang et al.'s study, the superior alleles of grain number generally had a positive effect on yield, while the superior alleles of grain weight generally had a negative effect on yield (Huang et al. 2015). Path analyses were performed by Oladosu et al. (2018) on rice yield and component traits revealed that three component traits possessed positive effect with yield. Xu et al. (2015) conducted a correlation analysis between yield and its components of 300 rice germplasms. Their result indicated that yield was significantly correlated with GPP or KGW, but non-significant correlations of yield were found with TP. One possible explanation for the conflicting results is that the bias caused by small sample size and lack of proper control for potential unmeasured confounders. For allowing the synthesis of results from different studies to estimate a common summary effect, meta-analysis was recognized as the appropriate method to achieve adequate sample sizes and optimal power (Panagiotou et al. 2013). In 2019, Zhao et al. (2019) reported the meta-analysis of GWAS from three GWAS panels, discovered 305 significant loci associated with tomato flavor, and demonstrated the benefits obtained from meta-analysis.

Recently, Mendelian randomization (MR) approach is a popular technique to assess causal relationship between disease and environmental risk factors within a meta-analysis framework in epidemiology (Bowden et al. 2019). MR method was used to investigate the role of ATP citrate lyase inhibitors in

cardiovascular disease (FERENCE et al. 2019), in which the potential unmeasured confounders could be well protected from observed association. In MR approach, genetic variants were used as instrumental variables to avoid the possibility of confounding, because the genetic variants are randomly allocated at meiosis (Mokry et al. 2015). Thus, combine meta-analysis and MR for complex traits will help researchers to obtain more reliable conclusion of their genetic relationship.

GWAS has been proved to be a new strategy for explaining the genetic basis of complex traits, which has the advantage of improving the efficiency of detecting natural variations (Li et al. 2018). Most GWAS studies focused on dissecting the genetic basis of single yield traits (Ta et al. 2018; Jiang et al. 2019), but the study on clarifying the genetic basis of relationship between the yield and component traits of rice is few. Here we carried out meta-analyses of GWAS results from two population (575 + 1495 F₁) in different environment, and adopted an MR design to further estimated the causal relationship between rice yield and component traits of rice. Our aim was to detect significant single-nucleotide polymorphisms (SNPs) associated with yield or component traits, to analyze the genetic bases contributing to relationship between them, and to investigate possible utilization pattern for selecting the component traits of yield in breeding practice so as to improve the rice production. The study will provide theoretical guidelines for enhancing rice yield potential.

Methods

Materials and phenotyping

Two populations of rice hybrid varieties were used in our study. One of population consists of 575 F₁ hybrid rice lines, which produced by 115 varieties (restorer lines of 29 three-line wild-deficient hybrid rice and 86 accessions of micro-core germplasm) as male parents were crossed with 5 sterile lines (4 two-line sterile lines and 1 three-line sterile line) as female parents. The 575 hybrid lines were grown both in Huazhong Agricultural University and Wuhan University in 2012. The other population consists of 1495 F₁ hybrid rice lines, including 1,170 lines were bred from the three-line system and 325 lines were generated from the two-line system. The 1495 hybrid lines were grown in Hangzhou and Sanya respectively. This population was obtained from the national center for gene research of Chinese academy of sciences (Huang et al. 2015). A total of four agronomic traits including GPP, KGW, TP and yield (YD) were recorded in both populations. These agronomic traits were measured for at least three samples of each accession, and the average measurement was taken as the phenotypic value for GWAS analysis.

Resequencing And Genotyping

The population of 575 hybrid rice lines was sequenced on the Illumina HiSeq2500 platform at 11× genome coverage on average. By quality control, we obtained 1,894,012 high quality SNPs with minor allele frequency (MAF) > 5% and missing rate < 20% across the 575 accessions. The high diversity SNP

maps of 1,495 hybrid rice varieties are publicly available (<http://www.ncgr.ac.cn/RiceHap4>) (Huang et al. 2015). The genomes of 1,495 hybrid lines were sequenced on the Illumina HiSeq2000 at twofold genome coverage, and 1,531,463 SNPs passing quality control (MAF < 1%).

Genotype Imputation And Gwas Analysis

3000 rice genomes project (<https://snp-seek.irri.org/download.zul>) as the reference panel to perform SNP imputation in the genotype data of 575 and 1495 hybrid rice lines by using beagle software (version 5.0) (Browning et al. 2018), and all imputed SNPs with MAF < 1% were filtered. Then conducting separate GWAS using mixed-linear-model association (MLMA) in GCTA software (Yang et al. 2014) and collecting the summary statistics to run a meta-GWAS. At last, a total of 1,838,525 SNPs from four GWAS datasets were used for meta-analysis.

Meta-gwas Analyses

We used the fixed-effect model in Metal as the primary approach to conduct the meta-analyses (Willer et al. 2010), and the Cochran's Q-test was performed to heterogeneity test (Cochran et al. 1954). For those SNPs where heterogeneity occurs ($I^2 \geq 50\%$), the random effect model in METASOFT was adopted (Han et al. 2011). The genome-wide significant P -value for meta-GWAS was set as $P < 1E-06$ ($-\log_{10}P = 6$).

Mr Analysis

For the causal effect of rice yield and each component trait to be consistently estimated, the genetic variants were selected according to the three assumptions in MR analysis (Burgess et al. 2013), (i) the genetic variants were obtained from the results of meta-GWAS that associated with the single component trait at a genome-wide significant level ($P < 1E-06$); (ii) the genetic variants are not associated with any confounders; (iii) the genetic variants only effect yield through the single component trait, not through other component traits ($P > 0.05$). Since the selected SNPs in linkage disequilibrium (LD) may result in confounding (Mokry et al. 2015), we calculated the r^2 (LD) of all pairs between all selected SNPs using plink (version 1.90) (Chang et al. 2015) and discarded SNPs in LD ($r^2 > 0.01$) based on larger P -value.

The inverse-variance weighting (IVW) method was conducted for MR analysis to assess the effect of component traits on yield, which by summarizing the effects of multiple independent SNPs (Burgess et al. 2013). In sensitivity analyses, the weighted median method (Bowden et al. 2016) and MR-Egger method (Burgess et al. 2017) are used for MR analysis, which are more robust due to pleiotropic or invalid instruments involved.

Analysis Of Superior Alleles Of Significant Associate Loci

Calculated the average phenotypic measurement corresponding to genotypes of each significant SNP, and the least significant difference method was used for multiple comparison. Following Huang et al.'s method (Huang et al. 2015), the genotype of SNP with the highest-level yield or component trait was set to be the superior allele (for example, the allele corresponding to the largest number of grains per panicle was set to be the superior allele). Calculated the number of superior alleles in each hybrid rice line and recorded their corresponding average yield measurements. Omitted the number of superior alleles with less than 3 hybrid lines.

Results

Meta-GWAS analyses

Meta-analyses of GWAS were performed based on four datasets' (two locations for each population) GWAS results (Additional file 1: Figures S1-S4). Manhattan plots and quantile–quantile plots of meta-GWAS are shown in Fig. 1. A total of 3592 significant loci were identified (Additional file 2: Table S1), including 2450, 1116, 23 and 3 significant associated loci were separately detected for GPP, KGW, TP and YD, which were distributed on all of the rice chromosomes except for chromosome 10. According to the information of RAP-DB (<http://rapdb.dna.affrc.go.jp/>), candidate genes were searched in a genomic region of 200 KB around the associated SNPs (Additional file 1: Table S2). We discovered 7, 7, and 3 cloned genes separately associated with GPP, KGW and TP. A total of three candidate genes associated with different traits, among which *OsBZR1* (Zhu et al. 2015) and *OsSPL14* (Jiao et al. 2010) have been reported previously and *Os02g0106966* was novelty discovered. Both *OsBZR1* and *OsSPL14* were detected in KGW and GPP, *Os02g0106966* was detected in KGW and TP. In this study, only 3 significant loci for YD were detected, but 3589 significant loci for the component traits were detected. It may be because rice yield has a low heritability which mainly affected by many minor-effect genes, the low heritability of rice yield is also showed in our previous study (Xu et al. 2018). These results suggested that selecting the component traits of yield as a complementary route to improve the rice production is recommended.

Causal Relationship Between Yield And Its Components

MR analyses were performed to estimate the causal relationship between rice yield and its component traits in the study. The results of MR analyses were shown in Table 1. The IVW method results showed both GPP and TP have a positive causal relationship with YD ($P < 0.05$), and there was no directional causal relationship between KGW and YD. In sensitivity analyses, the results of the weighted median method are also confirmed the results of IVW method. The intercept term obtained by MR-Egger regression analysis indicated no evidence of directional pleiotropic ($P > 0.05$). The MR analyses provided some evidence that rice yield probably enhanced by pyramiding the superior alleles of genes controlling GPP or TP, and the superior alleles of genes controlling TP should be priority to pyramid, because TP has a greater effect (Beta = 1.865) on yield than KGW (Beta = 0.456) and GPP (Beta = 0.086).

Table 1
MR results of a causal relationship between yield and its component traits

Trait	Methods	Beta	95% CI	<i>P</i>
GPP	IVW	0.086	0.030 ~ 0.141	0.003
	Weighted median	0.081	0.009 ~ 0.152	0.028
	MR-Egger	-0.029	-0.160 ~ 0.103	0.668
	MR-Egger(intercept)	1.387	-0.063 ~ 2.836	0.061
KGW	IVW	0.456	-0.119 ~ 1.031	0.120
	Weighted median	0.035	-0.758 ~ 0.828	0.931
	MR-Egger	-1.002	-2.858 ~ 0.855	0.290
	MR-Egger(intercept)	1.115	-0.236 ~ 2.466	0.106
TP	IVW	1.865	1.035 ~ 2.694	< 0.0001
	Weighted median	1.540	0.353 ~ 2.727	0.011
	MR-Egger	1.797	-1.633 ~ 5.228	0.304
	MR-Egger(intercept)	0.046	-2.165 ~ 2.256	0.968

Note: *CI*/confidence intervals, *P* statistically significant associations with a $P < 0.05$.

Causal Relationship Between Gpp And Yd

As required for MR analysis, a total of 2450 SNPs reached genome-wide significance for GPP ($P < 1E-06$) in meta-analyses of GWAS, among which six SNPs were selected as instrumental variables to estimate the causal relationship between GPP and YD (Table 2). The six SNPs were not associated with KGW or TP ($P > 0.05$), and no evidence of LD between them (all pairwise $r^2 < 0.01$). In MR analysis, a positive causal relationship between GPP and YD were observed with the IVW method (Table 1, Fig. 2a). One standard deviation (SD) genetic higher GPP was associated with a 0.086 SD higher YD (Beta = 0.086, 95% CI: 0.030 ~ 0.141, $P = 0.003$). In sensitivity analyses, the Cochran's Q-test illustrated no obvious heterogeneity ($I^2 = 5\%$, $P = 0.38$). The weighted median method also showed GPP had a positive effect on YD (Beta = 0.081, 95% CI: 0.009 ~ 0.152, $P = 0.028$). MR-Egger regression indicated no evidence of directional pleiotropy for the associations of GPP with YD (intercept = 1.387, $P = 0.061$). It is worth noting that some cloned genes were detected in the meta-GWAS on GPP, the phenotype of transgenic plants with these genes had a similar phenomenon. For example, the *OsSPL14* mutant produced more grain number per panicle, enhanced rice yield (Jiao et al. 2010). Compared with the control non-transgenic plants, the over-expression of *OsBZR1* plants showed the 1000-grain weigh was increased about 3.4% and the spikelet number per panicle was increased 21.9%, that resulting in enhanced yield (Zhu et al. 2015). The

cd1 mutant exhibited a variety of phenotypic traits, such as reduction in grain number and panicle length, the biomass was lower than that of the wild type (Luan et al. 2011).

Table 2
Information about instrumental variables

SNP	Chromosome	Position	GPP		YD	
			Beta	P-value	Beta	P-value
chr03_29979498	3	29979498	-18.724	1.12E-07	-0.633	0.543
chr03_898774	3	898774	18.953	2.92E-07	1.254	0.299
chr05_7226049	5	7226049	-7.242	2.25E-08	-1.329	0.005
chr08_25257522	8	25257522	-16.559	8.58E-07	1.178	0.553
chr09_12464309	9	12464309	-7.538	4.49E-07	-0.974	0.066
chr12_22633431	12	22633431	15.738	3.01E-07	1.439	0.158

Note: All the SNP markers are named after the chromosome _ position

Causal Relationship Between Kgw And Yd

As required for MR analysis, a total of 1116 SNPs reached genome-wide significance for KGW ($P < 1E-06$) in meta-analyses of GWAS, among which eleven SNPs were selected as instrumental variables to estimate the causal relationship between KGW and YD (Table 3). These SNPs were not associated with GPP or TP ($P > 0.05$), and no evidence of LD between them (all pairwise $r^2 < 0.01$). In MR analysis, no significant associations between KGW and YD were observed with the IVW method (Table 1, Fig. 2b) (Beta = 0.456, 95% CI: -0.119 ~ 1.031, $P = 0.120$). In sensitivity analyses, the Cochran's Q-test illustrated no obvious heterogeneity ($I^2 = 0\%$, $P = 0.67$). The weighted median method also showed no evidence of causal relationship between KGW and YD (Beta = 0.035, 95% CI: -0.758 ~ 0.828, $P = 0.931$). MR-Egger regression indicated no evidence of directional pleiotropy for the associations of KGW with YD (intercept = 1.115, $P = 0.106$). Some regulate KGW genes were detected in the meta-GWAS on KGW. Relationship between KGW and yield is inconsistent in the transgenic plants with these genes. In Chen et al.'s study, the KGW of *FUWA* plants was 16.7% higher than that of wild type, but their yield was comparable (Chen et al. 2015). Song et al. (2007) reported loss of *GW2* function accelerated the grain milk filling rate, resulting in increased grain weight and yield.

Table 3
Information about instrumental variables

SNP	Chromosome	Position	KGW		YD	
			Beta	P-value	Beta	P-value
chr01_3547491	1	3547491	0.706	7.14E-08	-0.036	0.948
chr01_5524333	1	5524333	1.018	6.56E-07	0.310	0.840
chr02_1118809	2	1118809	0.786	1.56E-09	1.030	0.082
chr02_334316	2	334316	-1.474	2.38E-07	0.948	0.505
chr02_7792121	2	7792121	-1.256	9.21E-08	0.029	0.980
chr03_17810847	3	17810847	0.651	4.74E-07	0.711	0.394
chr03_33060865	3	33060865	-0.505	2.03E-08	-0.978	0.027
chr04_13785932	4	13785932	0.739	3.16E-07	0.623	0.406
chr05_16393143	5	16393143	0.466	4.71E-07	0.029	0.955
chr07_23215227	7	23215227	0.952	1.09E-07	-0.308	0.715
chr08_25464238	8	25464238	-0.875	2.01E-09	-0.004	0.997

Causal Relationship Between Tp And Yd

As required for MR analysis, a total of 23 SNPs reached genome-wide significance for TP ($P < 1E-06$) in meta-analyses of GWAS, among which three SNPs were selected as instrumental variables to estimate the causal relationship between TP and YD (Table 4). These SNPs were not associated with KGW or GPP ($P > 0.05$), and no evidence of LD between them (all pairwise $r^2 < 0.01$). In MR analysis, a positive causal relationship between TP and YD were observed with the IVW method (Table 1, Fig. 2c), 1 SD genetic higher TP was associated with a 1.865 SD higher YD (Beta = 1.865, 95% CI: 1.035 ~ 2.694, $P < 0.0001$). In sensitivity analyses, Cochran's Q-test illustrated no obvious heterogeneity ($I^2 = 0\%$, $P = 0.43$), The weighted median method also showed TP had a positive effect on YD (Beta = 1.54, 95% CI: 0.353 ~ 2.727, $P = 0.011$). MR-Egger regression indicated no evidence of directional pleiotropy for the associations of TP with YD (intercept = 0.046, $P = 0.968$). The cloned gene *OsPIN2* were detected in the meta-GWAS on TP. Chen et al. (2012) found that the *OsPIN2* transgenic plants had more effective tiller number, lower 1000-grain weight and higher yield.

Table 4
Information about instrumental variables

SNP	Chromosome	Position	TP		YD	
			Beta	P-value	Beta	P-value
chr02_21604477	2	21604477	-1.279	8.25E-07	-1.977	0.043
chr06_1578700	6	1578700	-0.639	4.52E-07	-1.691	3.21E-04
chr11_26492375	11	26492375	-0.507	7.04E-08	-0.720	0.043

Loci for component traits had an indirect effect on yield

We identified four significant loci that had an indirect effect on yield by MR analyses (Fig. 2, Additional file 1: Table S3). Among them, the SNP chr05_7226049 (Fig. 2a) for GPP had an indirect effect on yield, and located nearby the cloned gene *OsPYL11*. Kim et al. (2014) reported that compared with the control plants, the transgenic plants over expressing *OsPYL11* showed no significant difference in tiller number, but the yield was severely reduced. Our study indicated the yield severely reduced may be caused by the number of grains decreased. The SNP chr06_1578700 (Fig. 2c) for TP closed to the *D62* (a gene regulating tillers). Li et al. (2010) found that the tiller number of *D62* mutant rice was less than that of wild type. The SNPs chr02_21604477 and chr11_26492375 for TP also had indirect effects on yield (Fig. 2c, Table S3), which were first detected in our research. These finding provided new information for further improve rice yield potential.

Pyramiding Superior Alleles Of Significant Loci

The average yield performance of F_1 lines with different superior allele number of significant loci with direct effect, indirect effect and direct plus indirect effect were showed in Table 5 and Fig. 3. Three loci had an direct effect on yield were detected in the meta-GWAS on YD (Additional file 1: Table S3), among them, the average yield of the lines without superior alleles was 41.29 g, and the average yield of the lines with one superior alleles was 44.26 g (Fig. 3a, Table 5). The superior alleles of four loci that had an indirect effect on yield were also pyramided in the study. The results showed that the average yield of F_1 lines with 0 to 4 superior alleles was: 42.75 g, 42.52 g, 43.32 g, 45.34 g, 52.60 g, respectively. In general, pyramiding superior alleles of loci revealed enhanced yield (Fig. 3b, Table 5). A similar phenomenon also found in pyramiding the direct plus indirect loci, and the yield probably further increased (Fig. 3c, Table 5). Other research reported that the phenotype performance improved by pyramiding the superior alleles of loci associated with agronomic traits (Huang et al. 2015), our results suggested the yield enhanced also by pyramiding the superior alleles of loci that had indirect effect on yield. A combination of direct and indirect effects may better contribute to the yield potential of rice.

Table 5
The average yield performance with different number of superior alleles

Effect	No. of superior alleles				
	0	1	2	3	4
Direct	41.29	44.26			
Indirect	42.75	42.52	43.32	45.34	52.60
Direct plus indirect	40.71	43.53	43.16	44.98	52.91

Discussion

In this study, a total of 3592 significant SNPs were detected in meta-GWAS on yield and its component traits, which provide more information for rice agronomic traits breeding. It is worth noting that only 3 loci were detected in meta-GWAS on yield, this maybe result from that rice yield has a low heritability and hardly to be detected. For a low-heritability trait (such as yield), highly correlated auxiliary traits (such as GPP) will help improve the selection of traits with low heritability since they reflecting a shared biological basis (Wang et al. 2017). Then MR model was used to explore the causal relationship between yield and its component traits. The results showed that both GPP and TP has a positive causal relationship with YD, which are consistent with the improvement of rice production achieved by increasing GPP and TP in previous study conducted by Yan et al. (2011). However, no causal relationship between KGW and YD were observed in the study. Some genes that regulate KGW have been reported have the potential to enhance rice yield, such as *GW2* (Song et al. 2007) and *GW5* (Liu et al. 2017), we also detected a locus for KGW that had indirect positive effect on yield, but the summary effect of multiple loci showed no directional causal relationship KGW and YD (Fig. 2b). It probably because KGW as a agronomic trait is closely related to appearance quality traits (Zhu et al. 2019), while quality traits are often negatively correlated with yield (Zeng et al. 2017). The results of MR provided rationale for the improvement yield by ideotype breeding based on selection for GPP and TP. Finally, four loci were identified with indirect effect on yield by MR analysis, providing new information for enhancing the yield potential of rice. Previous study indicated that pyramiding the superior alleles of significant associated loci increased yield (Huang et al. 2015). Our results suggested the improvement of yield also by pyramiding the superior alleles of loci with indirect effect on yield, and a combination of direct and indirect effects may better contribute to the yield potential of rice.

The strengths of the study are: (i) a meta-analysis of GWAS data from multiple population and environments to estimate a summary effect provided greater statistical power (Panagiotou et al. 2013); (ii) The MR approach could less prone to confounders since the genetic variants were used as instrumental variables (Mokry et al. 2015); (iii) MR method were used to analysis the causal relationship between quantitative traits in this study, which weighted the effects of multiple independent SNPs into a summary effect, for quantitative traits, most of them are affected by multiple genes or the interaction of

genes, while the individual SNP only explain a small fraction of the variation in the quantitative traits. However the MR analysis may be biased by the possibility of invalid instrumental variables, it is difficult to completely exclude type I error and the potential influence of pleiotropy since the instrumental variables derived from meta-analysis of GWAS in the study. Thus, we conducted weighted median method and MR-Egger method to do sensitivity analysis. Compared the IVW method, the weighted median method showed have better finite-sample Type I error rates, the estimator is consistent even if up to 50% of the information comes from invalid instrumental variables (Bowden et al. 2016). The results of MR-Egger and heterogeneity test indicated the genetic variants had no pleiotropic effects on yield to some extent (Burgess et al. 2017). These results strengthened our confidence in the validity of assumptions.

In conclusion, we analyzed the genetic basis of the relationship between yield and its component traits by GWAS and MR methods, providing genetic insights for further improving rice yield potential. Our results suggested the improvement of rice production by pyramiding the superior alleles of genes regulating GPP and TP, and a combination of direct and indirect effects may better contribute to the yield potential of rice in breeding practice. These findings will provide theoretical guidelines for the rational design of rice by MAS breeding.

Abbreviations

GPP: the number of grains per panicle; KGW:kilo-grain weight; TP:tillers per plant; GWAS:genome-wide association study; QTL:quantitative trait locus; MAS:marker-assisted selection; MR:Mendelian randomization; SNP:single-nucleotide polymorphism; YD:yield; MAF:minor allele frequency; IVW:inverse-variance weighting; LD:linkage disequilibrium; CI:confidence intervals; SD:standard deviation

Declarations

Ethics approval and consent to participate

No applicable.

Consent for publication

No applicable.

Availability of data and materials

The datasets supporting the conclusions of this article are provided within the article and its additional files.

Competing Interests

The authors declare that they have no competing interests

Funding

Funding for this study was provided by the National key technology research and development program (2016YFD0100101), Open Research Fund of State Key Laboratory of Hybrid Rice (Hunan Hybrid Rice Research Center (2019KF05); Wuhan University (KF201912)), Research Foundation of Education Bureau of Hunan Province (19A244), Double first-class construction project of Hunan Agricultural University (SYL2019028), Hubei Provincial cooperative Innovation Center, Hubei Science and Education letter (2016) No.3.

Authors' contributions

LL and ZH designed and supervised the research; JS, ZL, XZ, SS, WT and ZT performed the research; JS, KX and CW analyzed the data; JS and LL wrote the paper. All authors read and approved the final manuscript.

Acknowledgements

We thank Prof. Sibin Yu and Tongmin Mou from Huazhong Agricultural University, Prof. Shuangcheng Li from Sichuan Agricultural University, Prof. Shuzhu Tang from Yangzhou University, Prof. Guanghua He from Southwest University, Prof. Huaxiong Qi from Hubei Academy of Agricultural Sciences for providing material to us. We also gratefully thank Zhaoming Chen from Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences for supercomputing technology support.

References

1. Bowden J, Davey Smith G, Haycock PC, Burgess S (2016) Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* 40:304–314
2. Bowden J, Del Greco MF, Minelli C, Zhao Q, Lawlor DA, Sheehan NA, Davey Smith G (2019) Improving the accuracy of two-sample summary-data Mendelian randomization: moving beyond the NOME assumption. *Int J Epidemiol* 48:728–742
3. Browning BL, Zhou Y, Browning SR (2018) A one-penny imputed genome from next-generation reference panels. *The American Journal of Human Genetics* 103:338–348
4. Burgess S, Butterworth A, Thompson SG (2013) Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 37:658–665

5. Burgess S, Thompson SG (2017) Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol* 32:377–389
6. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:s13742–015
7. Chen J, Gao H, Zheng XM, Jin M, Weng JF, Ma J, Wang JL (2015) An evolutionarily conserved gene, FUWA, plays a role in determining panicle architecture, grain shape and grain weight in rice. *Plant J* 83:427–438
8. Chen Y, Fan X, Song W, Zhang Y, Xu G (2012) Over-expression of OsPIN2 leads to increased tiller numbers, angle and shorter plant height through suppression of OsLAZY1. *Plant biotechnology journal* 10:139–149
9. Cochran WG (1954) The combination of estimates from different experiments. *Biometrics* 10:101–129
10. Ference BA, Ray KK, Catapano AL, Ference TB, Burgess S, Neff DR, Danesh J (2019) Mendelian randomization study of ACLY and cardiovascular disease. *N Engl J Med* 380:1033–1042
11. Han B, Eskin E (2011) Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *The American Journal of Human Genetics* 88:586–598
12. Huang X, Qian Q, Liu Z, Sun H, He S, Luo D, Fu X (2009) Natural variation at the DEP1 locus enhances grain yield in rice. *Nat Genet* 41:494
13. Huang X, Yang S, Gong J, Zhao Y, Feng Q, Gong H, Chen N (2015) Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nature communications* 6:1–9
14. Jiang S, Wang D, Yan S, Liu S, Liu B, Kang H, Wang GL (2019) Dissection of the Genetic Architecture of Rice Tillering using a Genome-wide Association Study. *Rice* 12:43
15. Jiao Y, Wang Y, Xue D, Wang J, Yan M, Liu G, Qian Q (2010) Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. *Nat Genet* 42:541
16. Kadam NN, Struik PC, Rebolledo MC, Yin X, Jagadish SK (2018) Genome-wide association reveals novel genomic loci controlling rice grain yield and its component traits under water-deficit stress during the reproductive stage. *J Exp Bot* 69:4017–4032
17. Kim H, Lee K, Hwang H, Bhatnagar N, Kim DY, Yoon IS, Kim BG (2014) Overexpression of PYL5 in rice enhances drought tolerance, inhibits growth, and modulates gene expression. *J Exp Bot* 65:453–464
18. Le Nguyen K, Grondin A, Courtois B, Gantet P (2019) Next-generation sequencing accelerates crop gene discovery. *Trends in plant science* 24:263–274
19. Li F, Xie J, Zhu X, Wang X, Zhao Y, Ma X, Zhang S (2018) Genetic basis underlying correlations among growth duration and yield traits revealed by GWAS in rice (*Oryza sativa* L). *Frontiers in plant science* 9:650
20. Li R, Li M, Ashraf U, Liu S, Zhang J (2019) Exploring the relationships between yield and yield-related traits for rice varieties released in China from 1978 to 2017. *Frontiers in plant science* 10:543

21. Li W, Wu J, Weng S, Zhang Y, Zhang D, Shi C (2010) Identification and characterization of dwarf 62, a loss-of-function mutation in DLT/OsGRAS-32 affecting gibberellin metabolism in rice. *Planta* 232:1383–1396
22. Liu J, Chen J, Zheng X, Wu F, Lin Q, Heng Y, Zhang X (2017) GW5 acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. *Nature Plants* 3:1–7
23. Liu J, Tao H, Shi S, Ye W, Qian Q, Guo L (2012) Genetics and breeding improvement for panicle type in rice. *Chinese Journal of Rice Science* 26:227–234
24. Luan W, Liu Y, Zhang F, Song Y, Wang Z, Peng Y, Sun Z (2011) OsCD1 encodes a putative member of the cellulose synthase-like D sub-family and is essential for rice plant architecture and growth. *Plant biotechnology journal* 9:513–524
25. Mokry LE, Ross S, Ahmad OS, Forgetta V, Smith GD, Leong A, Richards JB (2015) Vitamin D and risk of multiple sclerosis: a Mendelian randomization study. *PLoS medicine* 12(8)
26. Oladosu Y, Rafii MY, Magaji U, Abdullah N, Miah G, Chukwu SC, Kareem I (2018) Genotypic and phenotypic relationship among yield components in rice under tropical conditions. *BioMed research international* 2018:1–10
27. Panagiotou OA, Willer CJ, Hirschhorn JN, Ioannidis JP (2013) The power of meta-analysis in genome-wide association studies. *Annu Rev Genom Hum Genet* 14:441–465
28. Qian Q, Guo L, Smith SM, Li J (2016) Breeding high-yield superior quality hybrid super rice by rational design. *National Science Review* 3:283–294
29. Song XJ, Huang W, Shi M, Zhu MZ, Lin HX (2007) A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat Genet* 39:623–630
30. Ta KN, Khong NG, Ha TL, Nguyen DT, Mai DC, Hoang TG, Dinh BY (2018) A genome-wide association study using a Vietnamese landrace panel of rice (*Oryza sativa*) reveals new QTLs controlling panicle morphological traits. *BMC plant biology* 18:282
31. Wang Q, Tang J, Han B, Huang X (2019) Advances in genome-wide association studies of complex traits in rice. *Theoretical and Applied Genetics* 1–11
32. Wang X, Li L, Yang Z, Zheng X, Yu S, Xu C, Hu Z (2017) Predicting rice hybrid performance using univariate and multivariate GBLUP models based on North Carolina mating design II. *Heredity* 118:302–310
33. Wang Y, Zhai L, Chen K, Shen C, Liang Y, Wang C, Xu J (2020) Natural Sequence Variations and Combinations of GNP1 and NAL1 Determine the Grain Number per Panicle in Rice. *Rice* 13:1–15
34. Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26:2190–2191
35. Xing Y, Zhang Q (2010) Genetic and molecular bases of rice yield. *Annu Rev Plant Biol* 61:421–442
36. Xu Q, Chen W, Xu Z (2015) Relationship between grain yield and quality in rice germplasms grown across different growing areas. *Breeding science* 65:226–232

37. Xu Y, Wang X, Ding X, Zheng X, Yang Z, Xu C, Hu Z (2018) Genomic selection of agronomic traits in hybrid rice using an NCI population. *Rice* 11:1–10
38. Yan WH, Wang P, Chen HX, Zhou HJ, Li QP, Wang R, Zhang QF (2011) A major QTL, *Ghd8*, plays pleiotropic roles in regulating grain productivity, plant height, and heading date in rice. *Molecular plant* 4:319–330
39. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL (2014) Mixed model association methods: advantages and pitfalls. *Nat Genet* 46:100–106
40. Zeng D, Tian Z, Rao Y, Dong G, Yang Y, Huang L, Hu J (2017) Rational design of high-yield and superior-quality rice. *Nature plants* 3:1–5
41. Zhao J, Sauvage C, Zhao J, Bitton F, Bauchet G, Liu D, Causse M (2019) Meta-analysis of genome-wide association studies provides insights into genetic control of tomato flavor. *Nature communications* 10:1–12
42. Zhu X, Liang W, Cui X, Chen M, Yin C, Luo Z, Zhang D (2015) Brassinosteroids promote development of rice pollen grains and seeds by triggering expression of Carbon Starved Anther, a MYB domain protein. *Plant J* 82:570–581
43. Zhu YJ, Sun ZC, Niu XJ, Ying JZ, Fan YY, Mou TM, Zhuang JY (2019) Dissection of three quantitative trait loci for grain size on the long arm of chromosome 10 in rice (*Oryza sativa* L.). *PeerJ* 7:e6966

Figures

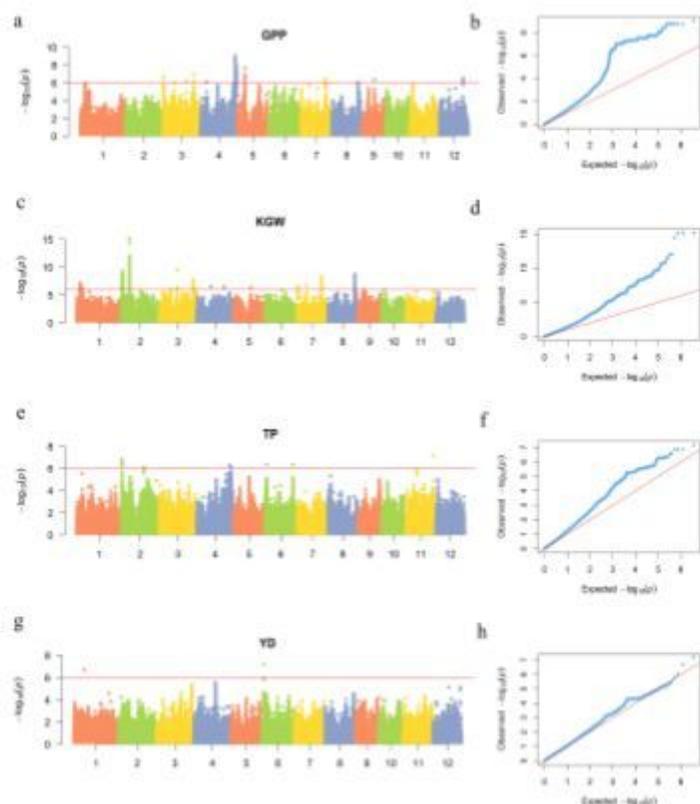


Figure 1

Meta-analyses results for GWAS. (a, b) Manhattan plots and quantile–quantile plots of GPP. (c, d) Manhattan plots and quantile–quantile plots of KGW. (e, f) Manhattan plots and quantile–quantile plots of TP. (g, h) Manhattan plots and quantile–quantile plots of YD. The genome-wide significant P value threshold $P < 10^{-6}$ is indicated by a horizontal line.

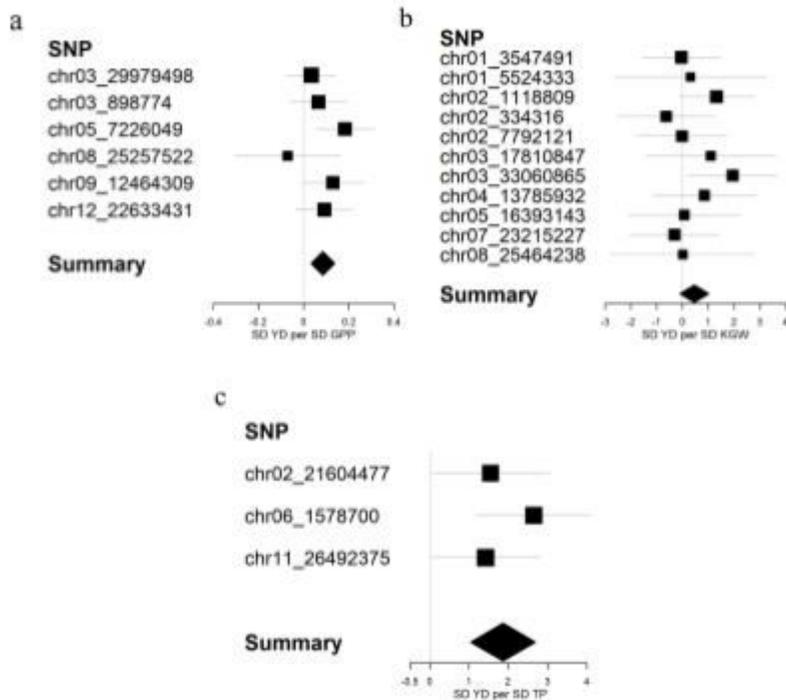


Figure 2

Causal effect estimates of yield and its components. Estimates are derived from IVW method of MR analyses. a Causal effect estimates between GPP and YD. b Causal effect estimates between KGW and YD. c Causal effect estimates between TP and YD.

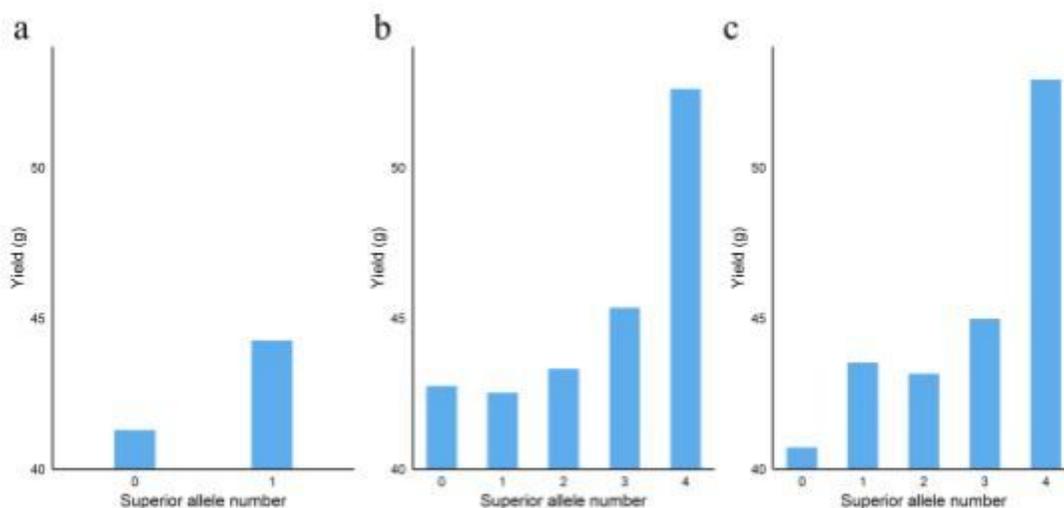


Figure 3

The average yield performance of F1 lines with different superior allele number of significant loci. a Direct loci, b Indirect loci, c Direct plus indirect loci

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile2.xlsx](#)
- [Additionalfile1.pdf](#)