

SCIENTIFIC REPORTS



OPEN

QTL mapping using an ultra-high-density SNP map reveals a major locus for grain yield in an elite rice restorer R998

Manshan Zhu^{1,2}, Dilin Liu^{1,2}, Wuge Liu^{1,2}, Dan Li^{1,2}, Yilong Liao^{1,2}, Jinhua Li^{1,2}, Chongyun Fu^{1,2}, Fuhong Fu^{1,2}, Huijun Huang^{1,2}, Xueqin Zeng^{1,2}, Xiaozhi Ma^{1,2} & Feng Wang^{1,2}

To dissect the genetic basis of yield formation in restorer line of hybrid rice, we conducted QTL analysis for 6 yield traits including panicles per plant (PPP), grains per panicle (GPP), grain yield per plant (GY), thousand-grain weight (TGW), above-ground biomass (AGB), and harvest index (HI) using SNP markers in a recombinant inbred lines (RILs) population derived from a cross between a tropical *japonica* inbred Francis and an elite *indica* restorer Guanghui 998 (R998). A total of 26 QTLs were detected using a high density genetic map consisting of 3016 bin markers. Nineteen out of the 26 QTL alleles from R998 had a beneficial effect on yield traits. Most of the QTLs were co-located with previously reported rice QTLs. *qAGB6* and *qHI9*, controlling AGB and HI respectively, were detected as novel QTLs. Four QTLs for GY were repeatedly detected across two years, with all the beneficial alleles from R998. Notably, *qGY8* explained over 20% of the yield variance in both years. Moreover, *qGY8* together with *qTGW8* and *qHI8* formed a QTL cluster. Markers tightly linked with *qGY8* were developed. Cloning of *qGY8* will facilitate its further exploitation in high-yield breeding.

Rice is a staple food crop feeding over half of the population in the world. With the increase of population and decrease of arable land, food production is facing huge challenge over the next 20 years. Hence, developing high-yielding new varieties remains one of the major goals in rice breeding. In the last decades, rice yield was not considerably improved and appeared to reach a plateau. Molecular breeding is widely accepted as an effective solution to improve the breeding efficiency. A few efforts have been made for yield improvement via gene pyramiding and rational design^{1–3}, however, such efforts depends largely on precise genetic dissection of agronomic traits due to the fact that rice yield is a complex trait controlled by both major-effect and minor-effect genetic loci³. Though a large number of reports on QTL mapping and cloning of yield-related loci have been available^{4,5}, information about the QTLs or alleles in many elite parental lines is still lacking, which might prevent efficient use of them in molecular breeding⁶. In addition, yield related traits such as panicles per plant, grains per panicle, and thousand-grain weight could influence each other. The interaction networks underlying the complex traits of yield are still largely unknown.

Due to the rapid development of next generation sequencing technology, genotyping by sequencing (GBS) has gained tremendous popularity as a rapid and cost-effective method in the development of genome-wide markers for genetic studies⁷. High-throughput SNP genotyping has been used in a number of QTL mapping studies in rice^{8–10}. For instance, Duan *et al.*⁹ used MSG sequencing in the QTL mapping of a giant panicle rice accession RI 128 and detected 49 QTLs for five yield traits⁹. Chen *et al.* performed high-resolution QTL mapping for grain appearance traits in *indica* rice¹⁰. As a reduced-representation sequencing approach of multiplexed samples, restriction-site associated DNA sequencing (RAD-seq) is a useful and cost-effective tool for genetic mapping studies¹¹. RAD-seq approach focuses only on short fragments of DNA adjacent to a particular restriction enzyme in the genome and provides significant data complexity reduction and increased throughput, thus allowing efficient and high density SNP discovery and genotyping.

¹Rice Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, 510640, China. ²Guangdong Key Laboratory of New Technology in Rice Breeding, Guangzhou, 510640, China. Manshan Zhu and Dilin Liu contributed equally to this work. Correspondence and requests for materials should be addressed to F.W. (email: fwang1631@163.com)

Chromosome	Number of SNPs in population	Number of bins	Length of genetic distance (cM)
Chr01	13747	389	296.371
Chr02	10533	337	396.433
Chr03	10831	334	227.044
Chr04	8332	278	399.559
Chr05	7650	246	332.199
Chr06	9955	234	269.438
Chr07	7994	228	377.276
Chr08	6062	214	249.178
Chr09	7766	195	220.843
Chr10	5543	179	329.698
Chr11	8924	208	272.52
Chr12	4565	174	275.632
Total	101902	3016	3646.191

Table 1. Number of SNPs and bins per chromosome in the R998 × Francis RILs population.

To better understand the genetic basis of rice yield traits in an elite restorer, we constructed a set of recombinant inbred lines (RILs) population derived from the cross between *japonica* inbred Francis and *indica* restorer R998. The two parents have similar thousand grain weight but different tillering capacity and panicle type. Francis has low tiller numbers and large panicles, while R998 has high tiller numbers and medium-sized panicles. High density bin map was generated after low-coverage RAD-seq of each RIL line and used for QTL mapping of 6 yield related traits including thousand grain weight (TGW), panicles per plant (PPP), grains per panicle (GPP), above-ground biomass (AGB), grain yield per plant (GYP) and harvest index (HI). Due to the fact that R998 is an elite restorer line widely used in commercial hybrid rice production, with 16 hybrid combinations released in the market, genetic dissection of yield formation in such an elite line is useful for breeders and may provide insightful guidance in high-yielding molecular breeding.

Results

Sequence analysis and construction of bin-map. Two parents and a total of 213 RILs were used for RAD-seq and resulted in a total amount of 46.93 Gb raw data, with an average of 218.28 Mb for each individual, among which the raw data for 197 individuals were over 100 Mb. Following the procedure of sliding-window method¹², we used a window size of 15 SNPs. One SNP was sliding each time and the genotype of that window was obtained. Thus we got the genotypes of each individual and eventually obtained a total of 3016 bin markers. The length of bin markers ranged from 20 kb to 4.4 Mb (Supplementary Figure S1), with a mean of 123.8 Kb, indicating the presence of recombination breakpoint every 123.8 kb in average. In total, 87.4% of bin markers were less than 0.25 Mb in length. There were 83 bins larger than 0.5 Mb in size and 4 large bins over 3.0 Mb dispersed on chromosomes 3(chr03_bin178), 4(chr04_bin57), 8(chr08_bin87) and 11 (chr11_bin98). The number of SNPs and bins per chromosome is shown in Table 1. Further, we generated a high-density genetic linkage map with a total distance of 3646.19 cM using the 3016 bin markers (Fig. 1, Supplementary Figure S2). The average distance between two bin markers across the map was 1.21 cM. The number of bin markers on different chromosomes ranged from 174 on chromosome 12 to 389 on chromosome 1 (Table 1). For each chromosome, the average genetic distance between adjacent bins ranged from 0.76 to 1.84 cM, with the maximal distance between 3.44 and 10.61 cM.

Quality and accuracy of the map. To examine the quality of the constructed bin-map, we first compared the linkage map with a map generated from an RIL population of 150 individuals derived from a cross between *indica* cv. 93-11 and *japonica* cv. Nipponbare⁸. The total number of bin markers in Wang's map is 2,334, and that of our map is 3016. The total genetic distance of the 12 chromosomes of these two maps is quite divergent, with 1539.5 cM in their map and 3646.19 cM in our map. It appeared that the total genetic distance of our map is almost twice bigger. As a result, the average genetic distance between adjacent bins with greater than zero distance is 1.21 cM on our map, larger than the average of 0.72 cM in Wang's map. The maximal genetic distance between adjacent markers is similar, with 10.61 and 8.3 cM on our map and Wang's map, respectively. In general, the marker distance appeared to be larger than Wang's map, indicating the algorithm difference between the software MSTMap we used and MAPMAKER for construction of the genetic map.

The accuracy and mapping resolution of the map was examined from mapping of *sd1*, the well-known semi-dwarfing gene, which is also known as the "green revolution gene"¹³. We scored the plant height for the 213 RILs lines and mapped the phenotype with CIM method in WinCartographer 2.5. As a result, seven QTLs associated with plant height were detected. Among them, the peak of *qPH1.2* overlapped with the cloned gene *sd1* at chr01_bin331 with a high LOD value of 15.18. The physical distance between *sd1* gene and chr01_bin331 is only 0.6 Kb, demonstrating a high mapping resolution of the bin-map (Fig. 2A and B). In parallel, we verified the presence of a null allele of *sd1* in R998 and a wild type allele of *Sd1* in Francis using a gel-based functional marker described previously¹⁴ (Fig. 2C).

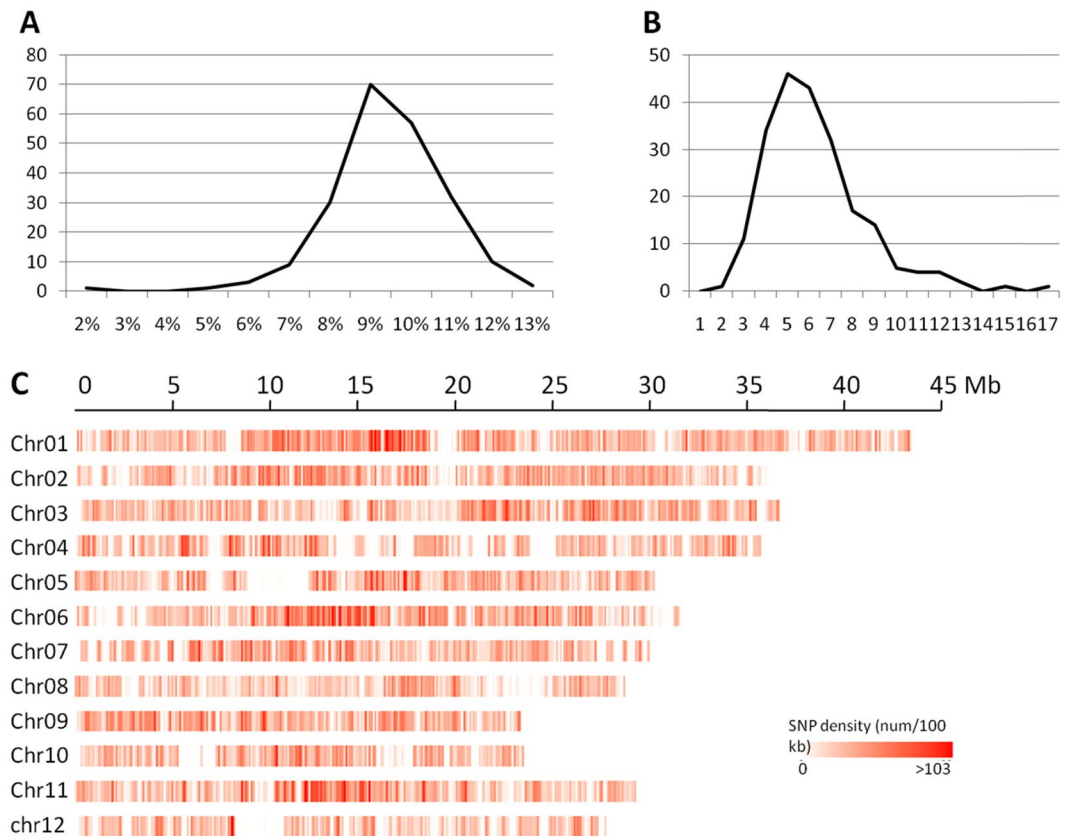


Figure 1. Sequencing results for the RILs population. **(A)** Coverage distribution in the RILs population. The horizontal axis represents coverage and the vertical axis indicates the number of individual lines. **(B)** Distribution of sequencing depth in the RILs population. The horizontal axis represents the sequencing depth and the vertical axis indicates the number of individual lines. **(C)** Heat map distribution of SNPs on each chromosome of the rice genome.

Distribution of yield traits and trait correlations. There are considerable variance between two parental lines regarding the panicles per plant, grains per panicle and yield per plant (Table 2). However, there is only slight difference in thousand-grain weight between two parents. Taken together, Francis is a variety with low tillering capacity, big panicle and relatively low yield, whereas R998 is featured by high tillering capacity, medium-sized panicles and high yield. All the six traits displayed a normal distribution within the RILs population. Most of them showed transgressive segregation except the number of panicles per plant (Fig. 3). Both the degree of skewness and kurtosis were mostly less than 1, indicating their nature as quantitative traits controlled by multiple genes and thus meeting the requirements of QTL analysis.

The correlation between yield traits was evaluated by regressing phenotypic values of one trait on those of another. The correlation coefficients among traits are shown in Table 3. Most of the correlations were in the positive direction, consistent with previous studies¹⁵. For instance, grain yield per plant is positively correlated with panicles per plant, grains per panicle, thousand grain weight, above ground biomass and harvest index. Three pairs of traits have negative correlations, such as panicles per plant and grains per panicle, thousand grain weight and grains per panicle, and above ground biomass and harvest index.

Detection of yield-related QTLs. We used 213 RILs derived from the R998 × Francis cross for mapping QTLs involved in yield formation. Using the composite interval mapping (CIM) in Windows QTL Cartographer 2.5, we detected a total of 26 yield-related QTLs in the population. Significant QTLs were detected for all the six traits as summarized in Table 4 and Fig. 4.

Of the QTLs for yield components, four QTLs were associated with panicles per plant. R998 alleles contributed positively to all the four loci. They were located on the chromosomes 1, 4, 5, and 6 respectively, thus terming *qPPP1*, *qPPP4*, *qPPP5* and *qPPP6*. The phenotypic variation explained by the individual QTLs ranged from 4.63% to 9.07%.

Two QTLs (*qGPP4.1* and *qGPP4.2*) were detected for grains per panicle on chromosome 4, and the positive alleles were attributed to Francis. They explained a phenotypic variation of 6.9% and 7.47%, respectively.

Five QTLs (*qGY1*, *qGY5*, *qGY8*, *qGY10* and *qGY12*) associated with grain yield per plant were detected on chromosome 1, 5, 8, 10 and 12, respectively and the explained phenotypic variance ranged from 5.45% to 21.22%. The R998 allele contributes to the increase of grain yield per plant at all the five loci. One of the QTLs, *qGY8* showed the largest effect and explained 21.22% of the phenotypic variance.

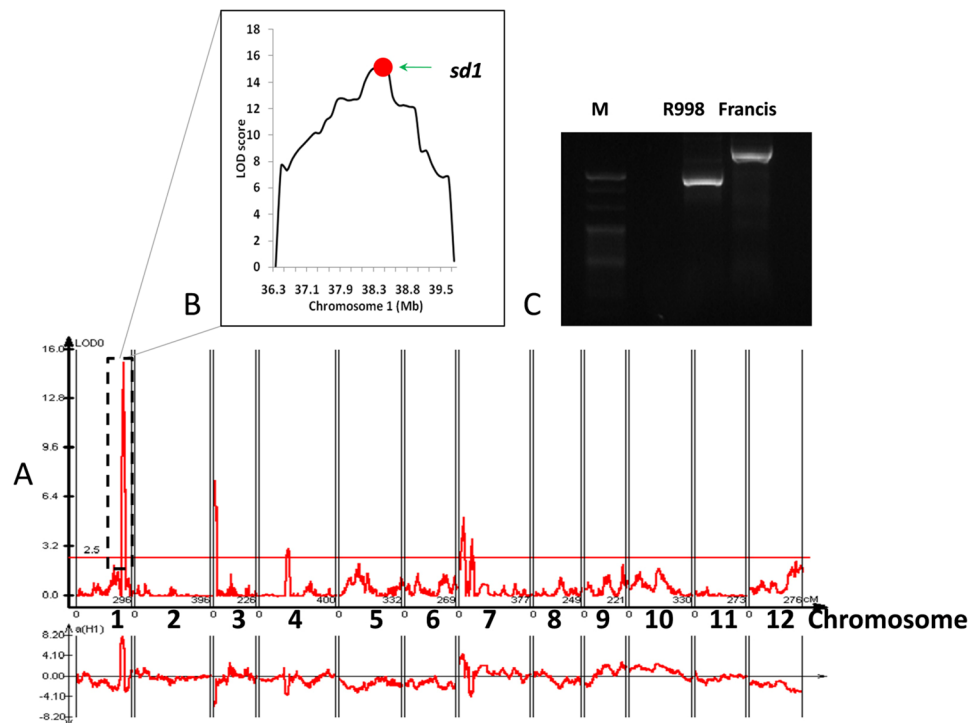


Figure 2. Mapping of QTLs controlling plant height in RILs population and the location of *sd1*. (A) Mapping curve of QTLs controlling plant height on 12 chromosomes. Curves in plot indicate the genetic or physical coordinate (X-axis) and LOD score (Y-axis) of detected QTL. (B) The box inside is the zoom-in image of the peak on chromosome 1. Red dot presents the relative physical position of *sd1* gene. (C) Genotyping of the two parent lines using a functional marker of *sd1*. M, 20 bp ladder (TaKaRa).

Traits	Parent		RILs Population			
	Francis	R998	Mean \pm SD	Range	Kurtosis	Skewness
PPP	3.0	8.0	6.4 \pm 1.14	3.4–10.0	0.52	0.54
GPP	190.0	155.1	153.0 \pm 35.6	35.4–255.1	−0.01	0.49
GY	10.4	23.2	14.8 \pm 5.0	3.0–31.4	0.70	0.02
TGW	20.96	21.23	20.31 \pm 2.80	14.28–31.87	0.88	0.60
AGB	20.6	45	40.05 \pm 16.11	13.8–108	3.35	1.50
HI	0.50	0.52	0.37 \pm 0.12	0.10–0.79	−0.07	0.15

Table 2. Means of yield traits for the parental lines and the RILs population.

Eight QTLs were detected for thousand grain weight, explaining 4.05% to 18.64% of the phenotypic variance in the population. At three of the loci, the Francis alleles conferred a positive effect in increasing the thousand grain weight. While the R998 alleles at the other five QTLs were contributing positively to thousand grain weight. Particularly, *qTGW3* and *qTGW8* are the two QTLs with the largest additive effect on thousand grain weight. And R998 alleles contribute positively at both loci and accounted for 11.52% and 18.64% of the variance for thousand grain weight in the population.

Five QTLs were associated with above ground biomass, locating on chromosome 6, 7, 8, 9 and 10, respectively. They explained a phenotypic variance ranging from 5.87% to 11.66%. Among them, Francis allele conferred a positive effect at *qAGB6* and *qAGB9*. While for the other three loci *qAGB7*, *qAGB8* and *qAGB10*, R998 alleles contributed positively for the formation of above ground biomass.

Two QTLs were detected for harvest index. Both alleles of *qHI8* and *qHI9* from R998 are responsible for increasing the harvest index, explaining a phenotypic variance of 19.92% and 6.89%, respectively. Notably, *qHI8* and *qGY8* coincided at the same location, with a LOD peak at chr08_bin94 (Fig. 5).

Validation of *qGY8* and molecular marker development. To validate the QTLs for grain yield, we repeated the phenotyping of grain yield per plant in 2015 and ran the QTL detection for a second time. As a result, four out of the 5 QTLs associated with grain yield per plant was repeatedly detected, locating on chromosome 1, 5, 8, and 12 respectively. They explained a phenotypic variance of 5.02%, 7.39%, 34.97% and 6.85%, respectively.

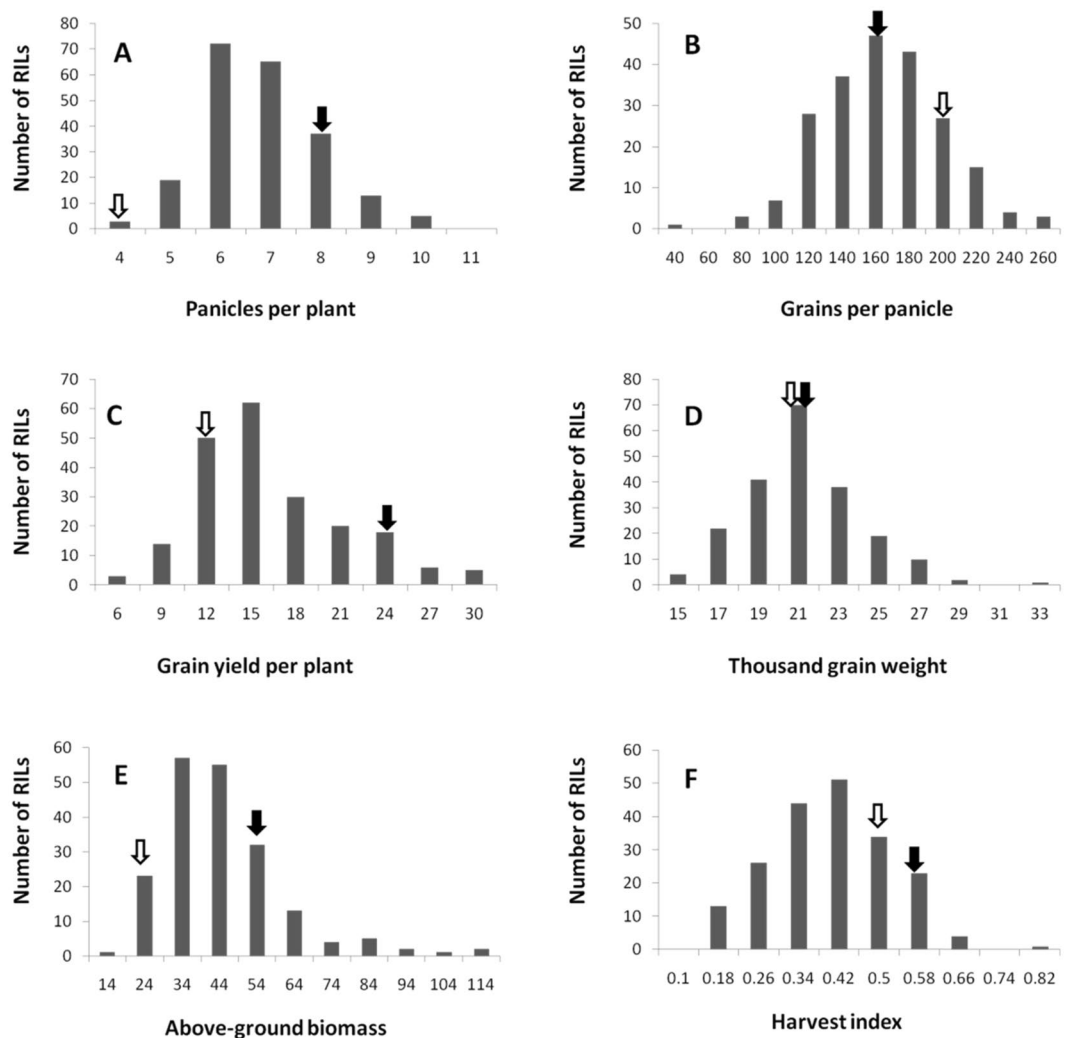


Figure 3. Frequency distribution of phenotypes for yield traits in the RILs population. (A) Frequency distribution of panicles per plant. (B) Frequency distribution of grains per panicle. (C) Frequency distribution of grain yield per plant. (D) Frequency distribution of thousand grain weight. (E) Frequency distribution of above-ground biomass. (F) Frequency distribution of harvest index. Black arrows represent trait means of R998, while white arrows indicate trait means of Francis.

trait	PPP	DPP	GY	TGW	AGB	HI
panicles per plant (PPP)	1					
grains per panicle (GPP)	-0.137	1				
grain yield per plant (GY)	0.469**	0.366**	1			
thousand grain weight (TGW)	0.087	-0.134	0.351**	1		
above-ground biomass (AGB)	0.315**	0.215**	0.544**	0.194**	1	
harvest index (HI)	0.149*	0.272**	0.574**	0.227**	-0.281**	1

Table 3. Correlation coefficients (r) among yield traits in the RILs population derived from the cross of R998 \times Francis. Note: * and ** indicate significant difference at the 0.05 and 0.01 level, respectively.

The QTL locations are exactly the same as those detected in the year 2014. Especially for the major locus *qGY8*, it explained a phenotypic variance of over 20% in both years.

Further, we performed genome re-sequencing of both Francis and R998 in 32x depth. The sequencing data provides useful information to compare genomic variation in the QTL region between the parental lines. Moreover, we developed seven InDel markers within the peak region of *qGY8* according to the polymorphism between the parents (Supplementary Table S1). All of the newly developed markers showed the significant polymorphism between Francis and R998 (Fig. 6), which will be used in marker assisted selection (MAS) and fine mapping of *qGY8* locus.

QTL	Chr	LOD value	peak marker	genetic position (cM)	physical position(bp*)	marker length(bp)	additive	Percentage of variance explained(%)
<i>qPPP1</i>	1	4.45	chr01_bin50	28.3	3857577–3925287	67711	−0.37	7.3
<i>qPPP4</i>	4	5.65	chr04_bin267	389.8	34229022–34276617	47596	−0.41	9.07
<i>qPPP5</i>	5	3.13	chr05_bin23	34.2	2073261–2110577	37317	−0.31	5.54
<i>qPPP6</i>	6	2.72	chr06_bin154	182.4	23728747–23828415	99669	−0.28	4.63
<i>qGPP4.1</i>	4	2.51	chr04_bin238	349.2	31749920–31774872	24953	9.98	6.9
<i>qGPP4.2</i>	4	2.72	chr04_bin248	365.5	32374932–32588914	213983	9.98	7.47
<i>qGY1</i>	1	3.94	chr01_bin29	18.11	2119506–2215651	96146	−1.44	5.45
<i>qGY5</i>	5	3.59	chr05_bin141	229.51	19881099–19940641	59543	−1.50	6.37
<i>qGY8</i>	8	13.46	chr08_bin94	130.21	16465630–16590676	125047	−2.70	21.22
<i>qGY10</i>	10	4.39	chr10_bin146	294.51	20352219–20619333	267115	−1.51	6.84
<i>qGY12</i>	12	4.33	chr12_bin135	209.91	23675639–23803476	127838	−1.54	6.81
<i>qTGW2.1</i>	2	2.56	chr02_bin57	90.11	5657047–5869449	212403	−0.12	4.05
<i>qTGW2.2</i>	2	3.55	chr02_bin169	208.81	22277253–22374393	97141	−0.16	4.97
<i>qTGW2.3</i>	2	3.08	chr02_bin291	334.71	32504342–32555810	51469	0.16	4.27
<i>qTGW3</i>	3	6.7	chr03_bin247	171.81	28494281–28520941	26661	−0.20	11.52
<i>qTGW7</i>	7	2.66	chr07_bin6	14.21	1190683–1237988	47306	0.11	3.12
<i>qTGW8</i>	8	12.53	chr08_bin109	147.11	18841775–19053506	211732	−0.26	18.64
<i>qTGW9</i>	9	3.88	chr09_bin182	207.31	21438495–21495270	56776	−0.14	5.71
<i>qTGW12</i>	12	4.97	chr12_bin173	274.41	27146979–27257344	110366	0.17	8.56
<i>qAGB6</i>	6	2.7	chr06_bin118	142.3	19988889–20195160	206272	4.02	6.09
<i>qAGB7</i>	7	2.54	chr07_bin125	237.4	19443803–19498463	54661	−3.99	5.87
<i>qAGB8</i>	8	3.36	chr08_bin69	103.9	5855674–5897395	41722	−5.11	8.51
<i>qAGB9</i>	9	2.72	chr09_bin30	26.7	6793297–6873993	80697	9.00	11.66
<i>qAGB10</i>	10	3.16	chr10_bin162	314.4	21818948–21843781	24834	−4.15	6.19
<i>qHI8</i>	8	9.31	chr08_bin94	130.2	16465630–16590676	125047	−0.05	19.92
<i>qHI9</i>	9	2.95	chr09_bin55	61.4	9269614–9407427	137814	−0.03	6.89

Table 4. QTLs for yield components identified in the R998 × Francis RILs population. Note: “*” denote physical positions based on the MSU Rice Genome Annotation Project release 7 (<http://rice.plantbiology.msu.edu/>). PPP, panicles per plant; GPP, grains per panicle; GY, grain yield per plant; TGW, thousand grain weight; AGB, above-ground biomass; HI, harvest index.

Discussion

It is widely accepted that mapping population derived from parents divergent in genetic composition and/or phenotype could yield better QTL mapping results¹⁶. We used a recombinant inbred line (RILs) population derived from Francis and R998, which belong to *japonica* and *indica* subspecies respectively and are diversified in several agronomic traits such as tiller number, grain yield and plant architecture. Therefore, as revealed from the distribution pattern of phenotype and polymorphic SNP sites, our RILs population is ideal for QTL mapping. The male parent R998 is an elite *indica* restorer line widely used in three-line hybrid rice seed production in China. It has high combining ability, strong restorability and broad restoring spectrum¹⁷. According to the statistics of ricedata website (www.ricedata.cn), R998 is listed as one of the most-widely used restorer lines during the year of 1983–2010, together with the famous restorer Minghui 63. The R998-series hybrid rice combinations such as Tianyou 998, Boyou 998, Qiyou 998 and Shanyou 998 were widely grown in south China, with an accumulated growing area reaching over 4.5 million hectares up to today. Therefore, QTL mapping of yield traits using such elite line would provide useful insight into the genetic basis of yield formation, as well as into further exploration of its breeding value in the context of rational super rice design through the approach of gene pyramiding.

A total of 101,902 high quality SNPs were identified in the genome between the *indica* line R998 and *japonica* line Francis using low-coverage RAD-seq approach. As a result, a high-density genetic map containing 3016 bin markers was generated, which had higher resolution than traditional gel-based SSR markers or RICE6K SNP¹⁸. As revealed in a previous study by Yu *et al.*⁷, increased density marker will considerably improve the QTL mapping resolution and the mapping accuracy. The high-density genetic map constructed in this study is an ideal map for QTL mapping or map-based cloning.

In this study, we detected in total 26 QTLs controlling 6 yield-related traits and located their physical positions on the corresponding chromosomes. Traditionally, lack of shared markers often prevented a precise comparison of QTLs detected between different studies¹⁹. In our study, however, the available marker information with physical positions in the bin map considerably facilitated the comparison of QTLs with earlier studies. The QTLs were distributed over all the chromosomes except chromosome 11. Among them, four genetic intervals were associated with panicles per plant. *qPPP1* on the short arm of chromosome 1, together with the *qPN1* locus by Tian *et al.*²⁰ and AQDY009 locus by Kobayashi *et al.*²¹ are located in the same interval. *qPPP4* was mapped to the same interval as reported previously by multiple studies in different populations^{15, 22–24}. Additionally, CQK6 mapped by

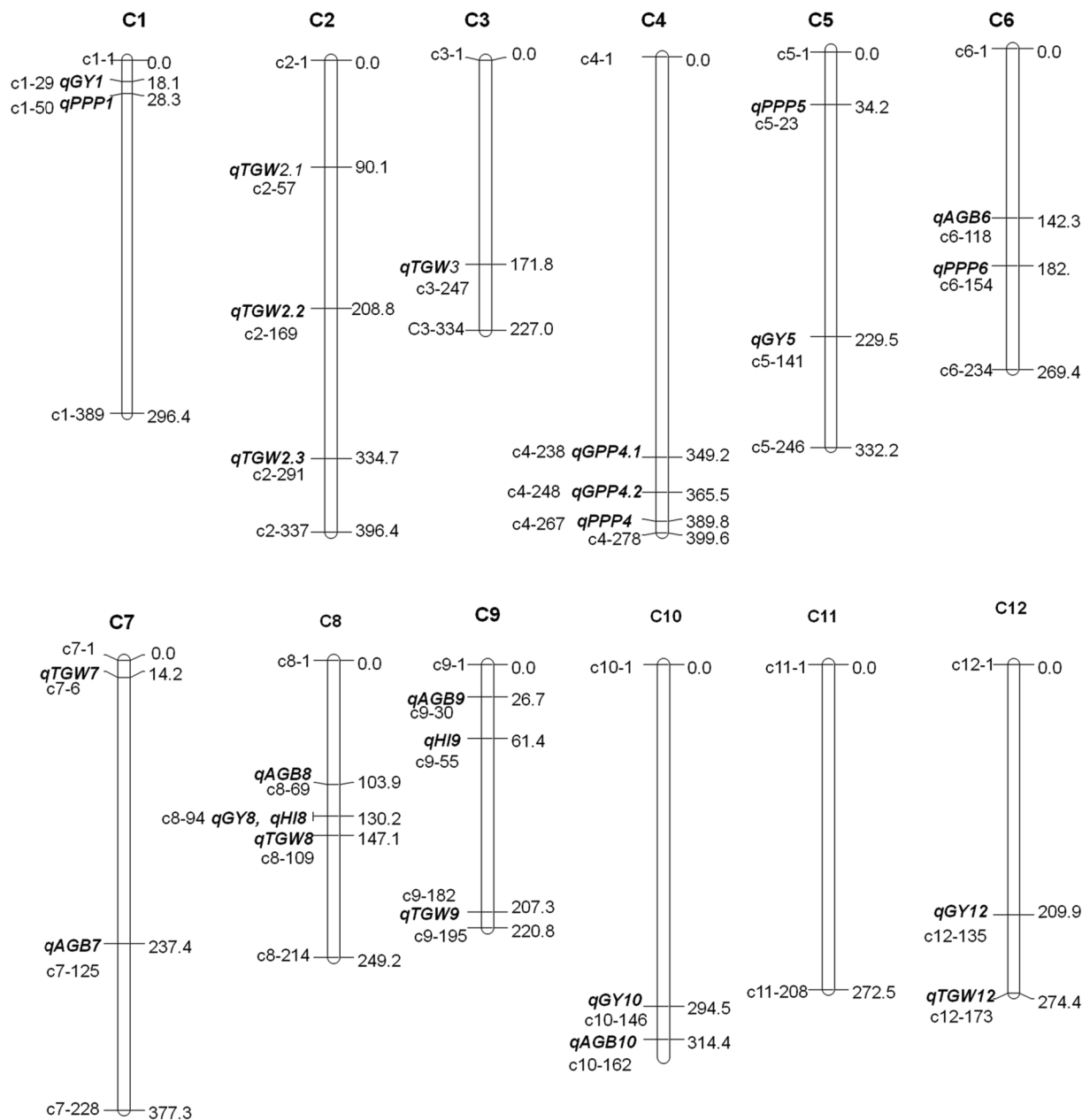


Figure 4. QTL locations of yield-related traits on the bin map. Bin numbers and genetic distance (cM) from the distal end of the short arm of each chromosome are shown. Detected QTLs were marked in bold italic letters. PPP, panicles per plant; GPP, grains per panicle; GY, grain yield per plant; TGW, thousand grain weight; AGB, above-ground biomass; HI, harvest index; C, chromosome. Bin markers were displayed in short form, for instance, c1-1 refers to bin 1 of chromosome 1.

Liao *et al.*²⁵ on the chromosome 5 was in the similar location of *qPPP5* in this study. *qPPP6* on chromosome 6 was co-located with AQDY102²¹ (Gramene QTL accession) and CQJ9²⁵.

Francis has an average panicle size of 190 grains per panicle, whereas that of R998 is only 155. Not surprisingly, both QTLs associated with grain per panicle were contributed by the male parent Francis. Both QTLs *qGPP4.1* and *qGPP4.2*, were found on the end of the long arm of chromosome 4. Previous studies revealed several QTLs at similar position with *qGPP4.1*, such as *gp4*²⁶, *gn-4*²⁷ and AQDR053²⁸. Likewise, *qGPP4.2* appeared to be overlapped with the *qSPP4* locus²⁹.

Grain yield is one everlasting goal in rice breeding. In our study, five loci associated with grain yield per plant were detected and four of them were consistently detected within two years thus they were stably expressed QTLs. They were distributed on chromosome 1, 5, 8, 10 and 12, respectively. All the positive-effect alleles for the trait of grain yield per plant were contributed by R998 alone, which indicated that the elite line R998 is pyramiding all the superior alleles. The *qGY1* locus we detected is within the same interval as *qGY1* reported by Wang *et al.*³⁰

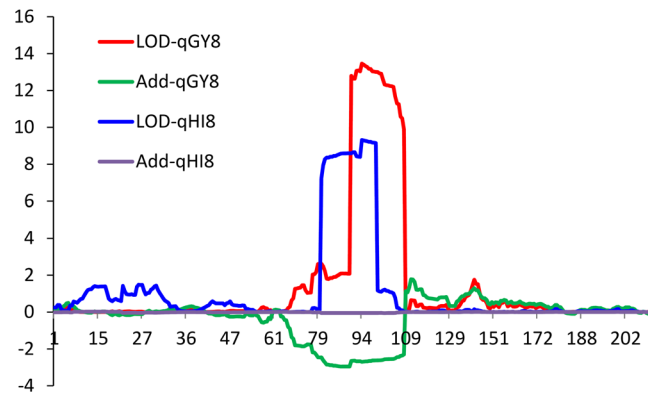


Figure 5. Co-localization of *qGY8* and *qHI8*. The value of LOD and corresponding additive effect for the trait of grain yield per plant (GY) and harvest index (HI) on chromosome 8 were shown. Y axis indicates LOD value or additive effect. X axis indicates the bin numbers on chromosome 8.

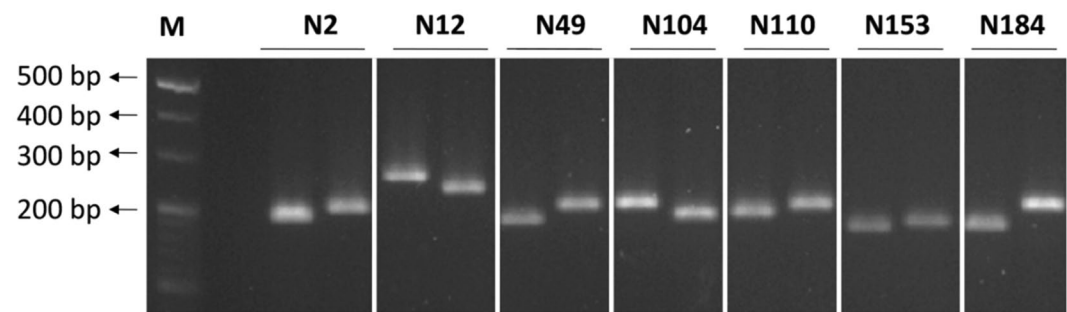


Figure 6. Development of tightly linked InDel markers in the *qGY8* region. M, 20 bp DNA ladder (TaKaRa). For each pair of samples, the left well represents R998 and the right well stands for Francis, respectively. PCR products were separated on 3% agarose gel. The full-length gel is presented in Supplementary Figure 3.

(physical position chr01:1,736,105–2,289,674 bp, IRGSP 1.0). Close to the position of *qGY1*, three other QTLs controlling grain yield were reported including *yld1.1*³¹, *GYP1*³² and *yld1.1*³³.

In the same location of *qGY5* locus, *qHI-5*³⁴, *yd5a*³⁵ and *qgwt5*²³ were reported to be responsible for grain yield per plant. For *qGY10*, no QTL for grain yield per plant was reported at the same location. However, Li *et al.*²⁸ described a QTL controlling grain weight per panicle in its vicinity. *qGY12* has a similar location as the locus *AQCN015* (Gramene QTL accession)³⁶. Similarly, *yd12* for the trait of grain yield was detected 3.1 Mb away from *qGY12* on chromosome 12³⁷. Notably, *qGY8* is a major effect QTL contributing to grain yield per plant and three QTLs for grain yield at the same location were reported, such as *GRYLD*²⁸, *yd8*³⁵ and *yld8.1*³⁸. In two successive years, *qGY8* explained a phenotypic variance of over 20% which most probably was attributed to artificial selection in breeding process. To facilitate further precise selection of *qGY8*, we developed seven tightly linked co-dominant InDel markers according to sequence variation between the two parents at the LOD peak site. These markers could be adopted in marker-assisted selection scheme for gene introgression and gene pyramiding.

Grain weight is an important factor contributing to grain yield. In our study, we detected a total of 8 loci associated with grain weight, among which three were located on chromosome 2. Compared to previous studies, we found QTLs with similar location as *qTGW2.1*. For example, *Gwt1b*³⁹ and *QKw2b*²⁸ were found at the similar site. For *qTGW2.2*, it shared the same location with *QKw2a*²⁸. Two other QTLs linked to grain weight, *gw2.1*¹⁹ and *gw2.1*³¹, were detected at the same position as *qTGW2.3*. On chromosome 3, we detected *qTGW3* at the position of 28,494,281–28,520,941 bp, which is very close to *qGW-3-1*⁴⁰ (Chr3: 29,052,279 bp) and *gw3.2*¹⁹ (Chr3: 33,393,485 bp), but 12 Mb away from the well characterized *GS3* locus (chr03:16,729,501–16,735,109). *qTGW7* (1,190,683–1,237,988 bp) was sharing similar location as *gw7.1* at the end of the short arm of chromosome 7³¹. *qTGW8* (Chr8:18841775–19053506 bp) was overlapping with *qGW8-1*⁴¹ and *HGW8*³². In another study by Xie *et al.*⁴², the major QTL *GW8.1* controlling grain weight was fine mapped to an interval of 306.4 kb between the marker RM23201.CNR151 and RM30000.CNR99. A (Chr8: 21,523,421–21,829,999 bp). Though the physical distance between *qTGW8* and *GW8.1* is over 2.5 Mb (ca. 10 cM in genetic distance), we were not able to completely exclude the possibility that *qTGW8* in the present study is identical to *GW8.1*. Further fine mapping of *qTGW8* is required to verify this point. For *qTGW9* with a LOD peak at 21,438,495–21,495,270 bp, it is located nearby *gw9.1*¹⁹ and in the same interval of CQAS114¹⁵. Interestingly, it exactly coincided with *gw9*^{43,44}, which was flanked by markers RM24718.CNR113–RM30005.CNR142 (Chr9:21,205,446–21,222,285 bp). Considering the small physical distance of 216 kb between *qTGW9* and *gw9*, it is most probably that *qTGW9* is identical to *gw9*. Further,

we detected a QTL for grain weight at the position of 27,146,979–27,257,344 bp on the long arm of chromosome 12, which is in the vicinity of a similar locus detected by Hua *et al.*²⁶ (Chr12: 23,799,065–24,360,652 bp).

Five QTLs associated with above-ground biomass were detected in this study, locating on chromosome 6, 7, 8, 9 and 10, respectively. *qAGB7* is in the same location as AQGI131⁴⁵. *qAGB8* shared a similar location as AQGI263⁴⁵, which was detected as a QTL for dry weight in the seedling stage. *qAGB9* was in the same location as *qTDW9-1* (AQAC016, 6016329 bp). *qAGB10* is coincided with the QTL loci including AQEX020 (*qSDW10*)⁴⁶, AQGI173⁴⁵ and AQGJ029⁴⁷. However, no QTL for biomass was reported around the QTL *qAGB6*. Therefore, *qAGB6* detected here is considered a novel QTL.

Harvest index is important trait in breeding and it represents the ability in the allocation of carbohydrate in the grains rather than in the vegetative part. We detected two loci associated with harvest index on chromosome 8 and 9, respectively. *qHI8* is almost in the same interval as *qGY8* (Fig. 5). In a previous study by Hittalmani *et al.*²⁴, *qHI8-1* and *qHI8-2* (Chr8: 17,437,513–17,438,003 bp) were reported at similar location with *qHI8* (LOD peak position 16,465,630–16,590,676 bp). In the interval of *qHI9*, no QTLs were reported at the same position thus *qHI9* is likely to be a novel QTL for harvest index.

QTLs controlling different traits are sometimes detected at the same or neighboring interval thus forming QTL clusters^{6,19}. QTL clusters for different traits might be explained by tight linkage of multiple causal genes, or by pleiotropic effect of a single gene. We detected a QTL cluster on chromosome 8, with QTLs for grain yield per plant, harvest index and thousand-grain weight at the similar position. Especially for grain yield per plant and harvest index, the interval of *qGY8* and *qHI8* were largely overlapped (Fig. 5). There was positive correlation between harvest index and grain yield per plant, as well as between thousand grain weight and grain yield per plant, therefore the clustering of *qGY8*, *qHI8* and *qTGW8* is reasonable. They showed major effect and the beneficial alleles were from the parent R998. Hence this QTL cluster is a useful target for selection in molecular breeding. Notably, though R998 is pyramiding all the five beneficial alleles for grain yield, there are 7 RILs showing transgressive grain yield over R998, which implies a possible gene interaction between Francis alleles and R998 alleles.

Methods

Mapping population. A recombinant inbred line (RILs) population derived from the cross between *japonica* line Francis and *indica* line R998 was used in genetic map construction and in the measurement of yield-related traits. Francis is a tropical japonica cultivar with fewer panicles, large panicle size and lower grain yield per plant while R998 is an elite *indica* restorer line with more panicles, moderate panicle size and higher grain yield per plant. Two parental lines have similar grain weight. The RILs population was obtained from continuous single seed decent (SSD) method. In this study, both parental lines and 213 randomly selected RILs lines (F₈) were used for genotyping and phenotyping.

Plant cultivation. In the early seasons of 2014 and 2015, 213 RILs lines and two parental lines were grown in Dafeng Experimental Station of Guangdong Academy of Agricultural Sciences in Guangzhou, China. Rice seeds were sown in March and transplanted at the beginning of April. Randomized block design was used for the experiment. Each line contains 4 rows in the block with 7 plants in each row. The plant density was 16.7 cm × 26.7 cm. All plant materials were grown in paddy fields following normal field management practices.

Evaluation of the yield-related traits. After maturity, 10 plants were harvested from the middle rows of each line for analysis of 6 yield-related traits including panicles per plant (PPP), grains per panicle (GPP), grain yield per plant (GY), thousand-grain weight (TGW), above-ground biomass (AGB) and harvest index (HI). Evaluation methods were similar to those described by Moncada *et al.*³³ as follows: (1) Panicles per plant was the average number of panicles on the ten plants, counting the panicles with no less than five seeds. (2) Grains per panicle were measured as the average number of filled grains calculated for the ten plants. (3) Grain yield per plant was the average weight per plant. After removing the empty kernels (grains filled with more than 30% were kept for analysis), bulked harvested filled grains were weighted and divided by 10. (4) Thousand-grain weight was the average weight of 1000 filled grains, measured in grams, averaged over three samples taken from bulk harvested grain from the ten plants. (5) Above-ground biomass was the total dry weight of above-ground part, including the grains and vegetative parts above ground, averaged from ten plants. (6) Harvest index was the dry weight of grains divided by the total dry weight of above-ground biomass evaluated from all ten plants. Correlation among the various traits was calculated in Excel 2007 using the trait averages from the experiments. Skewness and kurtosis were calculated to understand the nature of distribution of yield traits in the RILs population⁴⁸.

DNA isolation, genome re-sequencing and RAD-seq library preparation. Three weeks after transplanting, leaf samples were harvested from the plants. Genomic DNA was extracted from the parents Francis and R998, and the RILs population using CTAB method with minor modifications. Whole genome re-sequencing was carried out for Francis and R998 at 32x depth by Novogene (Beijing, China). Genome-wide SNP development and genotyping for the RIL population were performed by BGI (Shenzhen, China) using RAD-seq approach as described previously¹¹. Bar-coded adapters were designed according to the standard Illumina protocol for paired-end read libraries. One microgram of genomic DNA from each sample was digested with 1 μl FastDigest TaqI (Thermo scientific Fermentas) for 10 min at 65 °C in a 30 μl reaction. Unique barcode adapters (10 μmol) were then added to each sample well. The ligation was done with T4 DNA ligase (Enzymatics) at 22 °C for 1 h and heat inactivated at 65 °C for 20 min. For different samples, twenty-four ligation products were pooled in a single tube and 2 μl chloroform was added to inactivate the restriction enzyme. DNA fragments between 400–600 bp were then selected on a 2% agarose gel and purified using a QIAquick Gel Extraction Kit. All the products were amplified with 10 cycles of PCR (Phusion high-fidelity, Finnzymes) in a volume of 50 μl including 25 μl 2x

Phusion Master Mix, 1 μ l of common primer (10 μ M) and 1 μ l index primer. The amplified library was purified using a QIAquick PCR Purification Kit, quantified on the Agilent 2100 Bioanalyzer and sequenced on Illumina HiSeq 2000 platform.

SNP identification. The sequencing short reads of 100 bp in length of each individual were obtained according to the identified taq sequences. They were aligned with the rice reference genome of Nipponbare IRGSP 1.0 (<http://rapdb.dna.affrc.go.jp/>) using SOAPaligner program⁴⁹. Consensus sequences for each individual were generated by SOAPsnp. Input file was prepared by SAMtools for SNP calling with realSFS. Population SNP calling was performed with realSFS, based on the Bayesian estimation of site frequency at each site. The likelihoods of genotypes for each individual were integrated and sites with a probability of over 0.95 and a population whole depth higher than 40 were accepted as candidate SNPs. Potential SNPs were then filtered using the following criteria: loci with >70% missing data that also showed serious distorted segregation of the two parental genotypes were excluded. All the SNPs were filtered using a PERL script.

Determination of recombinant breakpoints. We converted the SNP data into another format to simplify the genotype calling analysis. The SNP type from Francis was coded as “a”, the R998 alleles were coded as “b”, and the heterozygotes were coded as “h”, while missing data was coded as “-”. A sliding window approach proposed by Huang *et al.*¹² with minor modification was adopted to evaluate a group of consecutive SNPs for genotyping. For each sample, we chose a window size of 15 SNPs without missing data for genotyping calling as described by Duan *et al.*⁹, which covered on average 54 kb or 0.2 cM of rice chromosomes. An a/b ratio of 12:3 or higher was recognized as “a”, while the ratio of 3:12 or lower as “b” and anything in between as “h”. We determined the breakpoints according to a published method for high-throughput genotyping by NGS (next-generation sequencing) with some modifications^{9,50}. Recombination breakpoints were determined by the junction of two different genotypes.

Construction of Bin map and QTL analysis. Bin genotype was obtained for each individual and used for construction of genetic map. Based on filtering of two parents SNP, a total of 101902 polymorphic SNP was obtained on the whole genome one SNP per 3.6 kb on average. Bin map was constructed using MSTMap software⁵¹. Data of the genetic map was imported into MapChart 2.2 software and integrated as linkage map⁵². Genetic distances were recorded for each chromosome. For QTL mapping, composite interval mapping (CIM) in Windows QTL Cartographer 2.5 was used for detection of QTL in the whole genome scale, additive effect (a) and phenotypic variance explained (R^2) was analyzed⁵³. The significance threshold of LOD value was calculated using 1000 permutation test. On the significance level of 5%, the LOD value threshold was used as the presence of a QTL. Nomenclature for QTLs followed the guidelines described by McCouch *et al.*⁵⁴.

References

- Wang, P., Xing, Y., Li, Z. & Yu, S. Improving rice yield and quality by QTL pyramiding. *Mol Breed* **29**, 903–913 (2012).
- Zong, G. *et al.* A pyramid breeding of eight grain-yield related quantitative trait loci based on marker-assistant and phenotype selection in rice (*Oryza sativa* L.). *J Genet Genomics* **39**, 335–350 (2012).
- Zeng, D. *et al.* Rational design of high-yield and superior-quality rice. *Nature Plants* **3**, 17031 (2017).
- Xing, Y. & Zhang, Q. Genetic and molecular bases of rice yield. *Annu Rev Plant Biol* **61**, 421–442 (2010).
- Ikeda, M., Miura, K., Aya, K., Kitano, H. & Matsuoka, M. Genes offering the potential for designing yield-related traits in rice. *Curr Opin Plant Biol* **16**, 213–220 (2013).
- Liu, D. *et al.* Mapping of the genetic determinant for grain size in rice using a recombinant inbred line (RIL) population generated from two elite *indica* parents. *Euphytica* **206**(1), 159–173 (2015).
- Yu, H. *et al.* Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. *PLoS One* **6**, e17595 (2011).
- Wang, L. *et al.* Mapping 49 quantitative trait loci at high resolution through sequencing-based genotyping of rice recombinant inbred lines. *Theor. Appl. Genet.* **122**, 327–340 (2011).
- Duan, M. *et al.* Genetic analysis of an elite super-hybrid rice parent using high-density SNP markers. *Rice* **6**, 21 (2013).
- Chen, L. *et al.* High-resolution QTL mapping for grain appearance traits and co-localization of chalkiness-associated differentially expressed candidate genes in rice. *Rice* **9**, 48 (2016).
- Baird, N. A. *et al.* Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**, e3376 (2008).
- Huang, X. *et al.* High-throughput genotyping by whole-genome resequencing. *Genome Res* **19**, 1068–1076 (2009).
- Monna, L. *et al.* Positional cloning of rice semidwarfing gene, *sd-1*: rice “green revolution gene” encodes a mutant enzyme involved in gibberellin synthesis. *DNA Res* **9**, 11–17 (2002).
- Ellis, M. H. & Spielmeier, W. Perfect markers for the semidwarfing gene *sd1* in rice. *Int. Rice Res. Notes* **27**, 13–14 (2002).
- Xiao, J., Li, J., Yuan, L. & Tanksley, S. D. Identification of QTLs affecting traits of agronomic importance in a recombinant inbred population derived from a subspecific rice cross. *Theor Appl Genet* **92**, 230–244 (1996).
- Bai, X. F., Luo, L. J., Yan, W. H., Kovi, M. R. & Xing, Y. Z. Quantitative trait loci for rice yield-related traits using recombinant inbred lines derived from two diverse cultivars. *J. Genet.* **90**, 209–215 (2011).
- Li, S. *et al.* Characteristics and hybrid seeds production technology of broad-spectrum restorer Guanghui 998. *Guangdong Agri Sci* **2**, 20–21 (in Chinese) (2006).
- Hu, W., Wen, M., Han, Z., Tan, C. & Xing, Y. Scanning QTLs for grain shape using a whole genome SNP array in rice. *J Plant Biochem Physiol* **1**, 104 (2013).
- Thomson, M. J. *et al.* Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. *Theor Appl Genet* **107**, 479–493 (2003).
- Tian, F. *et al.* Fine mapping of a quantitative trait locus for grain number per panicle from wild rice (*Oryza rufipogon* Griff.). *Theor Appl Genet* **113**, 619–629 (2006).
- Kobayashi, S., Fukuta, Y., Sato, T., Osaki, M. & Khush, G. S. Molecular marker dissection of rice (*Oryza sativa* L.) plant architecture under tropical and temperate climates. *Theor Appl Genet* **107**, 1350–1356 (2003).
- Lin, H. X. *et al.* RFLP mapping of QTLs for yield and related characters in rice (*Oryza sativa* L.). *Theor Appl Genet* **92**, 920–927 (1996).
- Zhuang, J. Y. *et al.* Analysis of QTL \times environment interaction for yield components and plant height in rice. *Theor Appl Genet* **95**, 799–808 (1997).

24. Hittalmani, S. *et al.* Identification of QTL for growth- and grain yield-related traits in rice across nine locations of Asia. *Theor Appl Genet* **107**, 679–690 (2003).
25. Liao, C. Y., Wu, P., Hu, B. & Yi, K. K. Effects of genetic background and environment on QTLs and epistasis for rice (*Oryza sativa* L.) panicle number. *Theor Appl Genet* **103**, 104–111 (2001).
26. Hua, J. *et al.* Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. *Proc Natl Acad Sci USA* **100**, 2574–2579 (2003).
27. Lu, C. F. *et al.* Comparative mapping of QTLs for agronomic traits of rice across environments by using a doubled-haploid population. *Theor Appl Genet* **94**, 145–150 (1997).
28. Li, Z., Pinson, S. R. M., Park, W. D., Paterson, A. H. & Stansel, J. W. Epistasis for three grain yield components in rice (*Oryza sativa* L.). *Genetics* **145**, 453–465 (1997).
29. Eizenga, G. C., Neves, P. C. F., Bryant, R. J., Agrama, H. A. & Mackill, D. J. Evaluation of a M-202 × *Oryza nivara* advanced backcross mapping population for seedling vigor, yield components and quality. *Euphytica* **208**, 157–171 (2016).
30. Wang, X. *et al.* Genetic background effects on QTL and QTL × environment interaction for yield and its component traits as revealed by reciprocal introgression lines in rice. *Crop J* **2**, 345–357 (2014).
31. Septiningsih, E. M. *et al.* Identification of quantitative trait loci for yield and yield components in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. *Theor Appl Genet* **107**, 1419–1432 (2003).
32. Brondani, C., Rangel, P. H. N., Brondani, R. P. V. & Ferreira, M. E. QTL mapping and introgression of yield-related traits from *Oryza glumapatula* to cultivated rice (*Oryza sativa*) using microsatellite markers. *Theor Appl Genet* **104**, 1192–1203 (2002).
33. Moncada, P. *et al.* Quantitative trait loci for yield and yield components in an *Oryza sativa* × *Oryza rufipogon* BC₂F₂ population evaluated in an upland environment. *Theor Appl Genet* **102**, 41–52 (2001).
34. Cho, Y. C. *et al.* QTLs analysis of yield and its related traits in wild rice relative *Oryza rufipogon*. *Treat. of Crop Res.* **4**, 19–29 (2003).
35. Yu, S. B. *et al.* Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proc Natl Acad Sci* **94**, 9226–9231 (1997).
36. Albar, L. *et al.* Genetic basis and mapping of the resistance to rice yellow mottle virus. I. QTLs identification and relationship between resistance and plant morphology. *Theor Appl Genet* **97**, 1145–1154 (1998).
37. Hua, J. P. *et al.* Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantageous for performance. *Genetics* **162**, 1885–1895 (2002).
38. Xiao, J. H. *et al.* Identification of trait-improving quantitative trait loci alleles from a wild rice relative. *Oryza rufipogon*. *Genetics* **150**, 899–909 (1998).
39. Tan, Z. *et al.* Identification of QTLs for ratooning ability and grain yield traits of rice and analysis of their genetic effects. *Acta Agronomica Sinica* **23**, 289–295 (1997).
40. Suh, J. P. *et al.* Mapping of QTLs for yield traits using an advanced backcross population from a cross between *Oryza sativa* and *O. glaberrima*. *Korea J Breed Sci* **37**, 214–220 (2005).
41. Cui, K. *et al.* Molecular dissection of the genetic relationships of source, sink and transport tissue with yield traits in rice. *Theor Appl Genet* **106**, 649–658 (2003).
42. Xie, X. *et al.* Fine mapping of a grain weight quantitative trait locus on rice chromosome 8 using near-isogenic lines derived from a cross between *Oryza sativa* and *Oryza rufipogon*. *Theor Appl Genet* **113**, 885–894 (2006).
43. Xie, X. *et al.* Fine mapping of a yield-enhancing QTL cluster associated with transgressive variation in an *Oryza sativa* × *O. rufipogon* cross. *Theor Appl Genet* **116**, 613–622 (2008).
44. Jin, F. X. *et al.* Detection of epistatic interaction of two QTLs, *gw8.1* and *gw9.1*, underlying grain weight using nearly isogenic lines in rice. *Breed Sci* **61**, 69–75 (2011).
45. Lian, X. *et al.* QTLs for low nitrogen tolerance at seedling stage identified using a recombinant inbred line population derived from an elite rice hybrid. *Theor Appl Genet* **112**, 85–96 (2005).
46. Xu, C. G. *et al.* Comparison of quantitative trait loci controlling seedling characteristics at two seedling stages using rice recombinant inbred lines. *Theor Appl Genet* **109**, 640–647 (2004).
47. Talukder, Z., McDonald, A. & Price, A. Loci controlling partial resistance to rice blast do not show marked QTL × environment interaction when plant nitrogen status alters disease severity. *New Phytol* **168**, 455–464 (2005).
48. Snedecor, G. W. & Cochran, W. G. *Statistical Methods*. Iowa University Press, p. 703 (1967).
49. Li, R. *et al.* SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* **25**, 1966–1967 (2009).
50. Davey, J. W. *et al.* Genome-wide genetic marker discovery and genotyping using next generation sequencing. *Nat Rev Genet* **12**, 499–510 (2011).
51. Wu, Y., Bhat, P., Close, T. & Lonardi, S. Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLOS Genet* **4**, e1000212 (2008).
52. Voorrips, R. E. MapChart: Software for the graphical presentation of linkage maps and QTLs. *J Hered* **93**, 77–78 (2002).
53. Wang, S., Basten, C. J. & Zeng, Z. B. *Windows QTL Cartographer 2.5* Department of Statistics, North Carolina State University, Raleigh, NC (2012).
54. McCouch, S. R. Gene nomenclature system for rice. *Rice* **1**, 72–84 (2008).

Acknowledgements

This work was supported by The National Key Research and Development Program of China (2016YFD0101801), Guangdong Provincial Program (2014B020202008, 2014A030310400, 2014A030310234), Earmarked fund for Modern Agro-industry Technology Research System, and National Natural Science Foundation of China (31501281).

Author Contributions

F.W., Z.M. and D.L.L. conceived and designed the experiments. M.Z., W.L., D.L., Y.L. and J.L. performed the experiments. Z.M. and D.L.L. analyzed the data. F.F. and H.H. contributed reagents/materials/analysis tools. D.L.L., C.F. and F.W. wrote the manuscript. M.Z., X.Z. and X.M. prepared figures and tables. All authors reviewed and approved the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-10666-7

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017