Dissecting a heterotic gene through GradedPool-Seq mapping informs a rice-improvement strategy

Changsheng Wang1,2,6, Shican Tang1,2,6, Qilin Zhan1, Qingqing Hou1,2, Yan Zhao1, Qiang Zhao1, Qi Feng1, Congcong Zhou1, Danfeng Lyu1, Lingling Cui1, Yan Li1, Jiashun Miao1,2, Chuanrang Zhu1, Yiqi Lu1, Yongchun Wang1, Ziqun Wang1, Jingjie Zhu1, Yingying Shangguan1, Junyi Gong3, Shihua Yang3, Wuqi Wang4, Jianfu Zhang4, Huaan Xie4, Xuehui Huang5 & Bin Han1

Hybrid rice breeding for exploiting hybrid vigor, heterosis, has greatly increased grain yield. However, the heterosis-related genes associated with rice grain production remain largely unknown, partly because comprehensive mapping of heterosis-related traits is still labor-intensive and time-consuming. Here, we present a quantitative trait locus (QTL) mapping method, GradedPool-Seq, for rapidly mapping QTLs by whole-genome sequencing of graded-pool samples from F2 progeny via bulked-segregant analysis. We implement this method and map-based cloning to dissect the heterotic QTL GW3p6 from the female line. We then generate the near isogenic line NIL-FH676::GW3p6 by introgressing the GW3p6 allele from the female line Guangzhan63-4S into the male inbred line Fuhui676. The NIL-FH676::GW3p6 exhibits grain yield highly increased compared to Fuhui676. This study demonstrates that it may be possible to achieve a high level of grain production in inbred rice lines without the need to construct hybrids.
rice heterosis, or hybrid vigor, refers to the increased yield in a hybrid offspring compared to its inbred parental lines. The rice hybrid varieties typically display a grain yield advantage of 10–30% over their parents. Beginning with the first commercial hybrid maize varieties in the 1930s, and the development of hybrid rice in the early 1970s in China, exploitation of heterosis in crop plants has achieved remarkable yield advantages over traditional breeding of inbred lines. To date, hybrid breeding that combines superior alleles from both parental lines to generate a better F1 variety is still one of the fastest and most efficient approaches in the breeding of rice as well as in many other crops. However, modern hybrid breeding relying on random crosses between diverse varieties and comprehensive phenotypic selection is still labor-intensive and time-consuming.

In the efforts to uncover the genetic basis of heterosis, there are several non-mutually exclusive hypotheses for heterosis, i.e., dominance, overdominance and epistasis, with evidences from molecular genetic experiments that have been performed. In recent work, the genetic basis of heterosis with regard to rice grain yield has been explored by an integrated genomic approach to construct a genome map of 1495 elite hybrid rice varieties and the in-depth genetic analyses on 17 F2 populations. The results showed that a small number of genomic loci from female parents explained a large proportion of the yield advantage of hybrids over their male parents of elite inbred varieties. For most of the heterosis-related loci identified, dominance or incomplete dominance of heterozygous loci plays an important role. Therefore, by optimizing cross design, characterization and dissection of heterotic genes and their allelic distribution in diverse germplasms will greatly enhance genetic improvement of rice. Several important high-yielding or heterosis-related genes have been characterized, such as the Ghd7 gene from the male parent (restore line) MH63 and the Ghd8 gene from the male lines HR5 and 9311. In contrast, yield-associated heterotic genes from female parents (male-sterile lines) remain largely unknown in rice. Genetic mapping for complex agronomic traits through linkage analysis or genome-wide association study (GWAS) showed its great power in the last few decades, which will help to identify the heterotic genes from female parents in hybrid rice.

In this study, to accelerate the genetic mapping processes, we develop a quantitative trait locus (QTL) mapping approach, GradedPool-Seq (GPS), that combines high-throughput sequencing with bulked-segregant analysis (BSA). This method is to score and assign F2 generations derived from a distant cross of parental lines exhibiting contrasting phenotypes into three or more graded groups based on their measured phenotypic values. Compared to previous methods using BSA coupled with whole-genome sequencing, such as MutMap, SHOREmap, next-generation mapping and QTL-seq, the GPS approach has the advantage of performing genetic mapping to simultaneously detect several QTLs at high resolution (~400-kb) by only requiring F2 population. Furthermore, we can assess multiple phenotypic traits using one F2 population. This method also allows us to rapidly identify heterotic genes. Benefitting from the robust GPS method coupled with follow-up experiments, we identified and validated a heterotic gene, GW3p6 (OsMADS1), from the female line (male-sterile line), that contributed greatly to 1000-grain weight and grain yield per plant in an elite hybrid rice variety Guang-Liang-You-676 (GLY-676). Notably, the near-isogenic line (NIL) NIL-FH676:GW3p6 produced by introgressing the GW3p6 allele from the female line (Guangzhan63-4S, hereafter as GZ) into the male line (Fuhu676, hereafter as FH) exhibit grain yield highly increased compared to FH plants. Rice hybrid breeding is currently hindered by bottlenecks of inefficiency and directionlessness, and the results of this study inform that it can open the door to achieving a high level of grain production using inbred lines instead of generating hybrids.

**Results**

**Development of the GPS method.** To rapidly identify QTLs for complex traits, we developed a method to map QTLs by directly sequencing graded-pool samples (GPS) from F2 progeny using modified BSA. The GPS procedure, including data obtaining, the filtering process, and statistical analysis, can be explained using plant height phenotype as an example (Fig. 1, Supplementary Fig. 1). To begin with, the mapping populations were developed from a cross between two genetically distant parental lines, one with a higher and the other with a lower phenotypic value of plant height, followed by subsequent self-pollination. Given that the parental lines display contrasting target phenotypes controlled by multiple genes, we predicted that the phenotypic variations in plant height among the F2 progeny will follow a pattern close to the normal distribution. By assessing plant height in each individual, we classified the F2 population into several graded bulks (three graded bulks in the graphic example) from the highest to lowest bulk according to their frequency distribution, as opposed to the two phenotypic tails used in BSA (Fig. 1a). We selected 100 to 150 individuals from each bulk in equal mixed DNA samples with sufficient genomic coverage (average 0.5×–1× per individual) for sequencing to guarantee sequencing accuracy; selection of 20–30% of the total individuals as a bulk is recommended. Next, the sequence reads of each bulk were aligned against the reference genome sequence to estimate allelic frequencies. We assumed that the depth at each single nucleotide polymorphism (SNP) was 100×. If a variance is not related to the phenotype, it would present 50% reference reads and 50% alternate reads in all three bulks; related SNPs would show great distinction in reference reads to alternate reads between the Grade 1 and Grade 3 bulks (Fig. 1b).

Data analysis was carried out as presented in Fig. 1c. We performed the statistical test, Ridit analysis, which adequately examines and analyses ordinal data, for each variant to compute its p value. Theoretically, if a variant is closely linked to a phenotype-related gene, its p value will be relatively small; however, it is not enough to predict the causative variant due to a large quantity of background noise in SNP calling by the distantly genetic cross. Consequently, background noise reduction must be taken into consideration. The noise-reduction algorithm we implemented in this study was a non-overlap sliding window approach, calculating the ratio of the number of statistically significant variants beyond the set threshold to the total number of variants in a defined genomic interval (~400 kb). Thus, we sought the interval with the largest ratio, i.e., SNPs clusters with significant p values in the highest proportion. The relationship between the ratio and chromosomal intervals reveals those genomic regions where QTLs associated with plant height are most likely to be located.

To assess its efficiency and robustness, we applied the GPS pipeline to identify QTLs underlying four agricultural traits (heading date, plant height, flag leaf angle and tiller angle). Multiple QTLs related to these four traits were mapped by GPS, and we focused on the phenotype of plant height. Three F2 populations of ~400 rice individuals were generated by crossing parental lines exhibiting a distinct plant height phenotype. According to the phenotypic value, we categorized the F2 lines into several ordinal classes from highest to lowest (Supplementary Table 1), followed with sequencing of the pools. After aligning the sequence reads to the reference sequence (IRGSP build 4) via BWA software and variant calling using GATK, we obtained genome-wide SNP information to conduct filtering procedure
**Fig. 1** The procedure of the GPS method. **a** F2 populations were generated by crossing two parent lines with distinct background and showed vastly different in plant height, followed by the self-pollination of F1 generation. Through the observation of plant height, F2 populations were classified from the highest to lowest bulk according to their frequency distribution. The letters M and F indicated the male parents and female parents, respectively. **b** The sequence reads of each bulk were aligned against the reference genome sequence to estimate allelic frequencies. We assumed that the depth at each SNP was 100×. If a variance was not related to the phenotype, it would present 50% reference reads and 50% alternate reads in all three bulks (right), whereas the related SNP would have great distinction of the ratio of reference reads to alternate reads among three bulks (left and middle). **c** Data analysis included Ridit analysis and reducing background noise approach. Several associated genetic intervals were mapped and indicated as red four-pointed star.
In our previous study, we mapped a genomic region containing the QTL GW3p6 contributing to the high grain production of the elite hybrid rice variety Guang-Liang-You 676 (GLY-676) from F2 individuals. To further fine clone the QTL GW3p6, we applied GPS to the F2 population derived from the elite hybrid rice variety GLY-676 (heterozygous first filial (F1)), which was generated from a cross between the varieties FH (male line) and GZ (female line).

First, we selected 1000-grain weight (TGW) as the trait for mapping heterotic genes. We ranked three categories (22.33–29.15 g/1000-grain, 29.16–31.09 g/1000-grain and 31.10–37.3 g/1000-grain) to phenotype TGW according to traits and then created simulated pools with their sequencing reads (individual sequencing reads data from European Nucleotide Archive under the accession number PRJEB13735). We implemented Ridit analysis with allelic frequencies from three bulks to calculate p values for each SNP (Fig. 2a). As numerous background noises complicated locating the QTL at a fine resolution, a noise reducing strategy followed the statistical test. After conducting all analysis, we located a 400-kb candidate interval contributing to grain weight (Fig. 2b). The mapping accuracy and resolution of GPS can reach almost the same level as that of Composite Interval Mapping. Notably, the GPS mapping results show that the method can be used as a faster and more convenient approach than conventional mapping methods in rice breeding. Considering the compatibility of the different versions of the assembled data, we remapped the TGW genes by GPS method based on Os-Nipponbare-Reference-IRGSP-1.0 (ref. 37) and MH63RS2 (ref. 38), and the genetic mapping results from the three reference genome assemblies were almost the same (Supplementary Fig. 6).

Furthermore, we screened the recombinant inbred lines (RILs) from the self-pollination F3 generation, and RIL79, one RIL in which genomic segment of GW3p6 was heterozygous but others were homozygous, was selected as a further segregating population (Supplementary Fig. 7). We used 1,079 plants from the F3 population of RIL79 to fine-scale map GW3p6, and 36 SNP markers were used for genotyping, ultimately narrowed down the interval to a ~5.9-kb region flanked by MP99 and MP100 (Fig. 2c–h). This region contains only the second half of Os03g0215400 (RAP-DB), and further sequencing analysis indicates a 15-bp non-homologous segment at the junction of the seventh intron and eighth exon of Os03g0215400 (Fig. 2g, i), from TCCITGGTGAAGGTA to ATGATATATATCT. The 3’ terminal bases AG of the seventh intron were altered, and we speculated that this might lead to alternative splicing. The cDNA sequencing data showed that the splice site (AG/GT) slipped to the 32nd nucleotide (AG/GC) of the last exon (Supplementary Fig. 8), directly caused a premature stop codon, and the original mature protein was truncated by 32 amino acid residues (Fig. 2j).

Moreover, we evaluated the power of GPS to detect regions responsible for heading date, tiller angle and flag leaf angle. The phenotype categories for the F2 generations are listed in Supplementary Table 1, and ratio plots are presented in Supplementary Fig. 3. The results for these four agronomic traits all suggested robust applications of our method (Table 1). We then performed GPS to analyze Takagi et al.’s data, comparing with another WGS-BSA method. The results showed consistency with their results, and our mapping region of causal genes was narrowed to 0.4 Mb (Supplementary Fig. 4). Additionally, we explored the whole procedure of our approach in depth, especially the influence on results when changing experimental variables (e.g., pool size, coverage, number of bulks, cases of misclassification and different statistical algorithm). The results of our computer simulation experiment are discussed in the Supplementary Note 1 (Supplementary Fig. 5).

### Table 1: Identified genomic intervals for four agronomic traits

| Trait                   | Population | Chromosome | Region (Mb) | Known loci | IRGSP 4.0 |
|-------------------------|------------|------------|-------------|------------|-----------|
| Plant height            | 1          | 6          | 17.2–17.6   | Ghd7a,17   | IRGSP 4.0 |
|                         | 7          | 8.8–9.2    |             |            | IRGSP 4.0 |
|                         | 12         | 22.0–22.4  |             |            | IRGSP 4.0 |
|                         | 2          | 6          | 26.8–27.2   | GW6ab,36   | IRGSP 4.0 |
|                         | 7          | 6.0–6.4    |             |            | IRGSP 4.0 |
|                         | 10         | 17.2–17.6  |             |            | IRGSP 4.0 |
|                         | 3          | 40.0–40.4  | sda,35      |            | IRGSP 4.0 |
|                         | 9          | 18.0–18.4  |             |            | IRGSP 4.0 |
| Heading date            | 1          | 5          | 22.4–22.8   | Ghd7a,17   | IRGSP 4.0 |
|                         | 7          | 8.8–9.2    |             |            | IRGSP 4.0 |
|                         | 4          | 3.6–4.0    | Ghd8b,18    |            | IRGSP 4.0 |
| Tiller angle            | 2          | 9          | 21.2–21.6   | TACTa,60   | IRGSP 4.0 |
|                         | 3          | 1          | 42.4–42.8   |            | IRGSP 4.0 |
|                         | 8          | 24.8–25.2  | IPA1,22     |            | IRGSP 4.0 |
|                         | 9          | 21.2–21.6  | TACTa,60    |            | IRGSP 4.0 |
| Flag leaf angle         | 3          | 1          | 40.8–41.2   | sda,35     | IRGSP 4.0 |
|                         | 8          | 24.8–25.2  | IPA1,22     |            | IRGSP 4.0 |
|                         | 9          | 16.8–17.2  | DEPR,61     |            | IRGSP 4.0 |

*The known genes located in identified genomic intervals via GPS.

**The known genes located closely adjacent to the identified intervals via GPS.**

Cloning and functional analysis of the QTL OsMADS1GW3p6.

In our previous study, we mapped a genomic region containing the QTL GW3p6 contributing to the high grain production of the...
heavier than the non-transgenic control plants (Fig. 3a–c), but constitutive expression of the FH OsMADS1 cDNA driven by the rice Ubiquitin promoter led to abnormalities in lemmas and paleae, as previously described41,42 (Fig. 3a). Moreover, we designed the target primers at the last exon of OsMADS1 using the CRISPR-Cas9 system in FH (CR-FH). Sequencing analysis revealed several insertions and deletions at the last exon in transgenic plants that resulted in loss of function in the C domain of the protein encoded by OsMADS1. The transgenic panicles exhibited phenotypic alterations (Fig. 3a), including elongated leafy paleae and lemmas, as well as low fertility seriously as previously reported40,42. The result was also similar to the
OsMADS1 encodes an MIKCc-type MADS-box transcription factor; it is also an E-class gene involved in rice floral organ development. The MIKC-type MADS-domain proteins consist of four domains: the MADS domain, Intervening domain, Keratin-like domain and C Domain respectively. The C-terminal region has been associated with transcriptional activation, and the non-homologous segment of OsMADS1GW3p6 was located within the C domain. Therefore, we conducted transcriptional activation experiment by yeast one-hybrid system (Supplementary Fig. 12). The results showed that both full-length OsMADS1 and OsMADS1GW3p6 had no transcriptional activity, which may be associated with a full-length inhibition of its transcriptional activity. In addition, the transcriptional activity of the OsMADS1 C-terminal region was higher than that of OsMADS1GW3p6. To further verify our experimental results, we applied a fluorescence report system for transcription activation.
experiments in rice protoplasts. The results were consistent with the data obtained in the yeast system, with OsMADS1 showing approximately 5 times stronger transcriptional activity than OsMADS1GW3p6 (Supplementary Fig. 13). These data were also consistent with previous reports that alternative splicing attenuates the activation of downstream genes, possibly regulating the transcriptional level of downstream auxin-related genes to change the grain size. The analysis of rice young panicles through RT-qPCR indicated that the expression of OsMADS1GW3p6 was indeed higher than OsMADS1 (Supplementary Fig. 14). According to our analysis of the structure of OsMADS1GW3p6, a large number of SNPs and InDels are located upstream of the 5’ UTR, and we evaluated whether the observed changes in expression are associated with nucleotides differences in the promoter. To this end, we used the fluorescence report system to verify the promoter, and found no obvious difference between the promoter of OsMADS1 and OsMADS1GW3p6 (Supplementary Fig. 15). Thus, this change in expression was likely caused by a premature stop codon due to alternative splicing. Taken together, these changes in the C domain may have resulted in the change of grain size.

**Improved grain yield by constructing a NIL containing GW3p6.** To further investigate the genetic function of GW3p6, the near-isogenic line NIL-FH::GW3p6 was generated by introgression of GW3p6 in the FH background. Some RILs with the genetic background of FH accounting for the vast majority were selected as backcrossing materials to generate NILs and were backcrossed twice to FH. With the aid of screening using a large number of molecular markers, and ultimately through sequence-based high-throughput genotyping, we generated a NIL with the FH genetic background and a ~130-kb heterozygous segment.

![Image](Fig. 4) A yield-related traits of FH and NIL-FH: GW3p6. a Plant phenotype of FH and NIL-FH::GW3p6. Scale bar, 10 cm. b Grain morphology of FH, NIL-FH: GW3p6 and FH/ NIL-FH::GW3p6. NIL and FH represent the grain morphology of NIL-FH::GW3p6 and FH, respectively. FH/NIL indicates the grain morphology of plants with heterozygous OsMADS1 (OsMADS1/OsAMDS1GW3p6) genotype. Scale bar, 5 mm. c-l A field-based comparison of the FH and NIL-FH::GW3p6 plants. c Grain length, d grain width, e 1000-grain weight, f grain yield per plant, g panicle length, h panicle number, i seed-setting rate, j grain number per panicle, k plant height, and l heading date. Data are shown as means ± SD (n = 24) with the indicated significance from a two-tailed Student’s t test. For statistical significance, *p < 0.05, **p < 0.01, NS not significant. Source data of Fig. 4c-l are provided as a Source Data file.
grain width, panicle number, panicle length, seed-setting rate, grain number per panicle and plant height (Fig. 4d, g–k), though the grain length and 1000-grain weight of NIL-FH::GW3p6 were ~6–7% higher than those of the FH plants (Fig. 4c, e). In addition, the grain yield per plant of NIL-FH::GW3p6 was increased by more than 8% (Fig. 4f), while the heading date had a 1–2 days delay compared to that of FH (Fig. 4l). Thus, GW3p6 is a useful target gene in breeding. By constructing a NIL, we demonstrated that an introgression line harboring a heterosis gene from the maternal parent could achieve better performance than inbred line. It also proved rice heterosis genes’ incomplete dominance played an important role in hybrid rice (Fig. 4b).

The heterosis effect of OsMADS1GW3p6 in rice breeding. As shown above, the NIL carrying the heterotic gene OsMADS1GW3p6 showed significantly increased grain yield. To further explore the potential of OsMADS1GW3p6 in rice breeding, we pyramided another previously reported heterotic QTL PN3q23 underlying panicle number16. The plants harboring two heterotic genes exhibited ~15% increased grain yield compared to the FH plants (Fig. 5a, Table 2), as well as higher yield than NIL-FH::GW3p6 plants (Table 2). The panicle number of the plants harboring PN3q23 was also significantly increased compared with that of FH (Supplementary Table 2). We measured the yield per plant for FH, NIL-FH::GW3p6 and GLY676 respectively. The results exhibited that GW3p6 explained 27.78% of the heterotic effect (Fig. 5b). The GW3p6 and PN3q23, two major heterotic genes from the female parent, explained over 40% of the heterotic effect (Fig. 5b). These findings implied a few heterotic genes from female parent played important roles in heterosis. We
detected the haplotype of OsMADS1GW3p6 in 1328 varieties of hybrid rice (Supplementary Data 1) simultaneously. OsMADS1GW3p6 was rarely detected among the three-line type hybrids, in which the proportion was ~1.6%. Approximately 11.5% of the two-line type hybrid varieties were found to carry the OsMADS1GW3p6 allele. These data indicate there is a large breeding potential for application of the superior allele of OsMADS1GW3p6 in future hybrid rice breeding.

In conclusion, the heterotic gene GW3p6 can improve the grain yield significantly in hybrid rice. By dissecting the heterotic gene OsMADS1GW3p6, we summarized a rice breeding strategy in hybrid rice and inbred rice (Fig. 5c, d). The F1 progeny derived from different crosses exhibited heterosis in certain agronomic traits, and we could locate superior genes in their F2 generations through the rapid and convenient GPS method, finally achieving fast and precise breeding through the marker-assisted selection approach.

Discussion

Traditional QTL mapping methods depending on the genetic linkage of QTLs to visible makers are laborious and time-consuming. The advent of NGS technologies and BSA has offered new opportunities for rapid identification of QTLs. Several methods have been established to accelerate the works in genetic mapping. We developed GPS, an improved approach combining high-throughput sequencing with modified BSA, for QTL mapping in crop breeding. Our method has several advantages over several previous methods. First, instead of using mutant lines such as Mutmap and Next-generation mapping, we choose parental lines comprising a large quantity of useful alleles directly. Second, our approach only requires F2 generation, reducing a great amount time needed for constructing genetic population. Third, GPS has a high resolution of ~400-kb, whereas the resolution of QTL-seq is ~2 Mb. In this work, we successfully identified the QTLs underlying five target traits (heading date, plant height, tiller angle, flag leaf angle and grain weight) in rice, demonstrating that GPS has a robust and extensive applicability for QTL mapping. In addition, the cost-effectiveness of the entire GPS process is relatively high compared with other methods. With the decline in high-throughput sequencing, applying GPS to identify QTLs would be economical. GPS can map QTLs underlying multiple traits simultaneously in one rice population. For example, a leaf sampled once can be used multiple times. Moreover, GPS does not require genotyping of all individuals, saving both time and laborious effort. Overall, this method significantly enhances the efficiency and cost-effectiveness of mapping candidate genes, enabling rapid identification of heterotic genes for rice breeding. Breeders and researchers would find good trade-offs in cost-effectiveness due to the relatively broad requirements and high efficiency of GPS. Thus, the highly efficient GPS will dramatically accelerate crop improvement in a cost-effective manner.

GPS shows a relatively wide range of applications in fine-mapping and breeding, but some limitations are also existed in GPS approach. First, QTL by environment interaction (QEI) is widely present in crops and other species. GPS with Ridit analysis currently had low power in QEI. Second, the GPS pipeline may not work very well with the overdominant loci. To identify QTLs through the software GPS, the SNPs related to the agronomic traits need to have a great distinction of reference reads to alternative reads. However, for the overdominant loci, there may be a great distinction of heterozygous genotypes to homozygous genotypes, but less great distinctions of reference reads to alternative reads. These limitations of GPS should be accounted for and assessed in plant breeding and fine mapping.

In our simulation experiment, we consider five experimental variables that influence the results of accurate detection. In the first place, we took the possibility that unlinked SNPs were calculated as related one into consideration. The simulation indicated that the number of individuals in each bulk had an impact on it. The more individuals we selected in each group, the less likely errors of false positive occurred. Furthermore, the number of pools is another important variable in our method. Meanwhile, more bulks meant being less influenced by incorrect categorization, thus we needed a balance here. In the simulation of the effects of misclassification, we took the 20% wrong classification of individuals in a certain bulk as an example, and this would be the main element leading to the inaccuracy of results. Three factors, depth, number of selected individuals and number of bulks, together influenced the capability of our method to detect the QTL within the context of misclassification. With the increasing value of these experimental variables (depth, pool size, number of bulks), the power to uncover the QTLs becomes larger.

| (i) Genotype | No. of plants | Grain yield per plant (g)a,b | (j) Genotype | Mean differences (i-J) | Standard error | Significance (p value)c |
|-------------|--------------|-----------------------------|-------------|------------------------|---------------|-----------------------|
| FH          | 70           | 42.644 ± 2.796              | GLY-676     | −14.234                | 1.222         | 4.319 × 10−14         |
| NIL FH::GW3p6| 70           | 46.615 ± 2.853              | FH          | 3.971                  | 0.726         | 7.267 × 10−7          |
| GW3p6 +    | 70           | 48.968 ± 6.330              | GW3p6 +    | −2.352                 | 0.726         | 7.479 × 10−3          |
| PN3q23      | 15           | 56.878 ± 3.888              | NIL FH::GW3p6| 6.323                  | 0.726         | 7.894 × 10−14         |
| GW3p6 +    | 10            | 2.352                       | FH          | 6.232                  | 0.726         | 7.894 × 10−14         |
| NIL FH::GW3p6| 70           | 46.615 ± 2.853              | GW3p6 +    | −7.911                 | 1.222         | 3.662 × 10−14         |
| PN3q23      | 15           | 2.352                       | NIL FH::GW3p6| 10.263                 | 1.222         | 1.082 × 10−13         |
| GW3p6 +    | 10            | 14.234                      | GW3p6 +    | 7.479 × 10−3           | 1.082 × 10−13 |
| NIL FH::GW3p6| 70           | 46.615 ± 2.853              | NIL FH::GW3p6| 3.971                  | 0.726         | 7.479 × 10−3          |
| GW3p6 +    | 10            | 14.234                      | NIL FH::GW3p6| 2.352                  | 0.726         | 7.894 × 10−14         |

aAll data is given as means ± SD
bData for the grain yield per plant are based on a field experiment using a randomized complete block design
cp value produced by the Tukey’s HSD

Table 2 Grain yield per plant of different plants harboring heterosis genes

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-019-11017-y | www.nature.com/naturecommunications
Finally, we also assessed the results of different statistical algo-
rithms and determined the optimal one. It turned out that the
Ridit analysis had the best performance in identifying QTL
among them. As shown above, the near isogenic line NIL-FH::GW3p6
displayed a large increase in yield compared with FH, but it was still
slightly less productive than the F1 plants. The reason for this
phenomenon is mainly there are multiple heterosis genes from
maternal parents. With the benefits of accurate QTL mapping,
GWAS and an improved GPS tool will be able to rapidly identify
heterotic genes for crop breeding. Further studies are needed to
identify and pyramiding more heterotic genes with regard to
heterotic genes for crop breeding. Further studies are needed to
GWAS and an improved GPS tool will be able to rapidly identify
OsMADS1GW3p6
not been exploited in large-scale rice breeding and that the
optimal yield as previously study in tomato48,49.

Methods

Plant materials and trait measurement. In the GPS study, four sets of elite
hybrid rice were used as mapping populations. The F2 generations were obtained
by self-pollinating F1 hybrid rice. The F2 populations with a large range of phe-
notypic variations were used for identifying QTLs for four agronomic traits (plant
height, heading date, flag leaf angle, tiller angle). The four sets of F2 populations
were planted in a rice paddy located in Hainan Province during the standard
growing season, and phenotypic investigations in the field were conducted in the
appropriate growth stage in spring of 2017.

In the fine-mapping study, Fuhui676 (FH), an elite indica restorer line, and the
Photo-Thermo Sensitive Genic Male Sterile Line Guangzhan63-4S (GZ) were used
by self-pollinating F1 hybrid rice. The F2 populations with a large range of phe-
notypic values of F2 individuals were arranged in ascending or descending order.
The F2 individuals corresponding to the phenotypic grade were marked and
assigned to different pools. In the process of identifying heterotic gene GW3p6, we divided the individuals
into three ordinal pools according to 1000-grain weight (22.33–29.15 g/1000-grain,
while the number of total SNPs is less than 10. We
we implemented
the candidate genetic intervals responsible for phenotype via ratio plot.
The location of peak represents a cluster of highly linked variants to a given trait in this
400-kb region.

Genotyping and fine mapping of OsMADS1GW3p6. Plant materials’ genotyping
were mainly based on Applied Biosystems 3730 XL fully automatic DNA sequence
practices/). After applying statistical tests to mapping populations, all the identified SNPs
were used to confirm the background genotype34. The high resolution melting curve
analysis by Roche LightCycler 480 and other genotyping methods (SNPs geno-
typing by Sanger Sequencing and InDels genotyping by gel electrophoresis) were
used to genotype for constructing NIL-FH::GW3p6.

The fine-mapping study focused on 1.079 F1 individuals of NIL-R79. 36 SNPs markers were used to screen the recombinants. Finally, the
OsMADS1GW3p6 locus was narrowed down to a 5.9-kb region between markers
30 × 16 cm during the standard growing season at experimental fields located in
Sanya (Hainan Province, China).

Seedling length and width were measured using an automatic digital grain size
scanner, and fully filled grains were used for measuring 1000-grain weight. Ten
plants in the middle of each row were harvested individually and used to
investigate yield-related traits, such as panicle number and grain weight per plant.
The phenotypic data for plant height, panicle length, grain number per panicle
and seed-setting rate were obtained using the main culm of each plant. The phenotype
of NIL-FH::GW3p6 plants was investigated in spring of 2018 in Hainan, China.
The phenotype of plants containing GW3p6 and PN3q23 was investigated in
autumn of 2018 in Shanghai, China.

Bulked-segregant analysis. To validate our method, we chose four complex traits
to conduct proof-of-principle experiments. As illustrated in Supplementary
Table 1, four agronomic traits were observed in the four sets of F2 populations. For
plant height, according to phenotypic values, we classified F2 populations into
several bulks from high to low plant height. The same strategy was carried out with
the other three traits using their F2 progeny. The number of pools depends on the
grade of the phenotypic difference and the number of F2 individuals. The phe-
notypic values of F2 individuals were arranged in ascending or descending order.
The F2 individuals corresponding to the phenotypic grade were marked and
assigned to different pools. The incomplete dominance effect of OsGW3p6 may be due to an
allelic dosage effect as in maize47. Uncovering more alleles and fine-tuning the dosage of OsMADS1GW3p6 may produce an optimal yield as previously study in tomato18,49.
MP99 and MP100. The phenotype of grain shape and grain weight in selected recombinants was confirmed using the self-progeny test. The genomic DNA of candidate OsMADS1 genes from OZ, FH and other landrace varieties were sequenced and analyzed by Sanger Sequencing System. A list of the markers used for fine-scale mapping and NIL construction were given in Supplementary Table 3 and Supplementary Table 4.

**Plasmid construction and plant transformation.** To generate the overexpression vector, full-length coding sequence without the stop codon of OsMADS1 and OsMADS1GW3p6 were cloned into the vector pNCGR-OX fused with Huis-tag. And to construct sgCRISPR-Cas9 vector, one CRISPR/Cas-mediated target in C- Domain was cloned into CRISPR/Cas9 vector55. All the constructs were transformed to japonica cv. Nipponbare and indica cv. FH by Agrobacterium-mediated transformation. The relevant PCR primer sequences are given in Supplementary Table 5.

**RNA extraction and quantitative RT-PCR analysis.** The total RNA was extracted from rice young panicles using TRIzol regent (Invitrogen), and 0.5 micrograms of total RNA was used to synthesize first-strand cDNA by the ReverTra Ace qPCR RT Master Mix with gRNA Remover (Code NO.FSQ-301, TOYOBIO). Quantitative real-time PCR was performed from cDNA using THUNDERBIRD SYBR qPCR Mix (Q201, TOYOBIO) according to the manufacturer’s instructions on Applied Biosystems Q5. The rice ubiquitin 5 gene was used as an internal control, and each qRT-PCR assay was performed at least three times in biological and technological replicates. The relevant PCR primer sequences are given in Supplementary Table 5.

**Transcriptional activation assay in yeast.** The full-length and different truncations or deletions coding sequence of OsMADS1 and OsMADS1GW3p6 were introduced into the pGBK7 (Clontech) vector to fuse the GAL4 DNA-binding domain. The empty pGBK7 vector was as negative control. Then the vectors were transformed into yeast strain AH109 and the clones were diluted to an absorbance of 1.0 (1/10, 1/100, 1/1000) at OD600. 15 μl of liquid culture was plated on the control medium and SD/-Leu-His-Ade/- Trp medium for 3d at 30 °C. The transcription activation assays were conducted according to the Matchmaker GAL4 transcriptional activation assay in yeast. The full-length and different truncations or deletions coding sequence of OsMADS1GW3p6 were marked as an asterisk by iTOL.

**Phylogenetic analysis.** The 1439 indica-indica phylogenetic trees were constructed by the published data32. The haplotypes of OsMADS1 were classified into two categories: OsMADS1 and OsMADS1GW3p6, the iTOL (version 4.2.4)39 was used to display and annotate the phylogenetic tree, the red annotation indicated the haplotype of OsMADS1GW3p6. Similarly, the neighbor-joining tree of the 66 rice accessions constructed by PHYLIP and the package MEGA5 was as known data. According to the haplotype of OsMADS1, the haplotype of OsMADS1GW3p6 was marked as an asterisk by iTOl in NJ-tree of 66 rice accessions.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability** Data supporting the findings of this work are available within the paper and its Supplementary Information files. A reporting summary for this Article is available as a Supplementary Information file. All other relevant data are available from the corresponding author upon request. All sequencing data that support the findings of this study have been deposited in European Nucleotide Archive (ENA) with the accession code PRJEB30329 [https://www.ebi.ac.uk/ena/data/view/PRJEB30329]. The source data underlying Figs. 2f, h, 3b, c, 4e-1, and 5b, as well as Supplementary Figs. 10a, b, 11d, e, 13, 14, and 15 are provided as Source Data file.

**Code availability** All software codes of the GPS approach are available in GitHub [https://github.com/sctang1991/GPS-pipeline].

Received: 2 January 2019 Accepted: 4 June 2019
Published online: 05 July 2019

**References**
1. Luo, D. et al. A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. *Nat. Genet.* 45, 573–577 (2013).
2. Hochholdinger, F. & Baldau, J. A. Heterosis in plants. *Curr. Biol.* 28, R1889–R1902 (2018).
3. Cheng, S. H., Zhuang, J. Y., Fan, Y. Y., Du, J. H. & Cao, L. Y. Progress in research and development on hybrid rice: a super-domesticate in China. *Ann. Bot.* 100, 959–966 (2007).
4. Chen, Z. J. Genomic and epigenetic insights into the molecular basis of heterosis. *Nat. Rev. Genet.* 14, 471–482 (2013).
5. Hochholdinger, F. & Hoecker, N. Towards the molecular basis of heterosis. *Trends Plant Sci.* 12, 427–432 (2007).
6. Birchler, J. A., Auger, D. L. & Riddle, N. C. In search of the molecular basis of heterosis. *Plant Cell* 15, 2236 (2003).
7. Schnable, P. S. & Springer, N. M. Progress toward understanding heterosis in crop plants. *Annu. Rev. Plant Biol.* 64, 71–88 (2013).
8. Xiao, J., Li, J., Yuan, L. & Tankoey, S. D. Dominance is the major genetic basis of heterosis in rice as revealed by QT analysis using molecular markers. *Genetics* 140, 743 (1995).
9. 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).
10. Zhou, G. et al. Genetic composition of yield heterosis in an elite rice hybrid. *Proc. Natl Acad. Sci. USA* 109, 15847–15852 (2012).
11. Melchinger, A. E., Utz, H. F., Piepho, H. P., Zeng, Z. B. & Schon, C. C. The role of epistasis in the manifestation of heterosis: a systems-oriented approach. *Genetics* 177, 1815–1825 (2007).
12. Garcia, A. A., Wang, S., Melchinger, A. E. & Zeng, Z. B. Quantitative trait loci mapping and the genetic basis of heterosis in maize and rice. *Genetics* 180, 1707–1724 (2008).
13. Seymour, D. K. et al. Genetic architecture of nonadditive inheritance in Arabidopsis thaliana hybrids. *Proc. Natl Acad. Sci. USA* 113, E7317 (2016).
14. Li, L. et al. Dominance, overdominance and epistasis condition the heterosis in two heterotic rice hybrids. *Genetics* 180, 1725–1742 (2008).
15. Huang, X. et al. Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nat. Commun.* 6, 6258 (2015).
16. Huang, X. et al. Genomic architecture of heterosis for yield traits in rice. *Nature* 537, 629–633 (2016).
17. Xue, W. et al. Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. *Nat. Genet.* 40, 761–767 (2008).
18. Yan, W. H. et al. A major QTL, Ghd8, plays pleiotropic roles in regulating date and yield potential in rice. *Genetics* 180, 767–767 (2008).
19. Zhang, H. et al. Heterosis and dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. *Proc. Natl Acad. Sci. USA* 100, 2574–2579 (2003).
20. Takeda, S. & Matsuoka, M. Genetic approaches to crop improvement: responding to environmental and population changes. *Nat. Rev. Genet.* 9, 444–457 (2008).
21. Mackay, T. F., Stone, E. A. & Ayroles, J. F. The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* **10**, 565–577 (2009).

22. Jiao, Y. et al. Regulation of OsNSP14 by OsMask156 defines ideal plant architecture in rice. *Nat. Genet.* **42**, 541–544 (2010).

23. Huang, X. et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **42**, 961–967 (2010).

24. Huang, X. et al. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat. Genet.* **44**, 32–39 (2011).

25. Si, L. et al. OsSPL13 controls grain size in cultivated rice. *Nat. Genet.* **48**, 447–456 (2016).

26. Abe, A. et al. Genome sequencing reveals agronomically important loci in rice using MapMatch. *Nat. Biotechnol.* **30**, 174–178 (2012).

27. Schneeberger, K. et al. SORTEmap: simultaneous mapping and mutation identification by deep sequencing. *Nat. Methods* **6**, 550–551 (2009).

28. Austin, R. S. et al. Next-generation mapping of Arabidopsis genes. *Plant J.* **67**, 715–725 (2011).

29. Takagi, H. et al. QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* **74**, 174–183 (2013).

30. Mansfeld, B. N. & Grumet, R. QTLSeqr: an R package for bulk segregant analysis with next-generation sequencing. *Plant Genome* **20**, 1297–1303 (2010).

31. Lonjou, C. F. et al. Hybrid breeding in autogamous cereals. *TAG* **125**, 1087–1096 (2012).

32. Bross, I. D. J. How to use Redit analysis. *Biometrika* **14**, 18–38 (1958).

33. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

34. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

35. Sasaki, A. et al. A mutant gibberellin-synthesis gene in rice. *Nature* **416**, 701 (2002).

36. Song, X. J. et al. Rare allele of a previously unidentified histone H4 acetyltransferase enhances grain weight, yield, and plant biomass in rice. *Proc. Natl Acad. Sci. USA* **112**, 76–81 (2015).

37. Kawahara, Y. et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6, 4 (2013).

38. Zhang, J. et al. Extensive sequence divergence between the reference genomes of two elite indica rice varieties Zhenshan 97 and Minghui 63. *Proc. Natl Acad. Sci. USA* **113**, E5163 (2016).

39. Liu, Q. et al. G-protein β subunits determine grain size through interaction with MADS-domain transcription factors in rice. *Nat. Commun.* **9**, 852 (2018).

40. Yu, J. et al. Alternative splicing of OsLG3b controls grain length and yield in whole genome resequencing of two bulked populations. *Plant J.* **74**, 174–183 (2013).

41. Jeon, J.-S. et al. *leafy hull sterile1* is a homeotic mutation in a rice MADS box gene. *Plant Physiol*. **166**, 1009–1018 (2014).

42. Prasad, K., Parameswaran, S. & Vijayraghavan, U. OsMADS1, a rice MADS-box gene, controls flowering time and grain yield in rice by affecting flowering time and grain yield in rice. *Proc. Natl Acad. Sci. USA* **112**, 76–81 (2015).

43. Arora, R. et al. MADS-box gene family in rice: genome-wide identification by deep sequencing of two elite indica rice varieties Zhenshan 97 and Minghui 63. *Proc. Natl Acad. Sci. USA* **113**, E5163 (2016).

44. Liu, Q. et al. Natural variation at the DEP1 locus enhances grain yield in rice. *Nat. Genet.* **41**, 494 (2009).

Acknowledgements

We thank the China National Rice Research Institute and Fujian Academy of Agriculture for providing the hybrid rice varieties. We thank Yaoyang Liu (South China Agricultural University) for providing the CRISPR/Cas9 vector. This work was funded by the National Natural Science Foundation of China (31630055/317788103), and the Chinese Academy of Sciences (XDB7010301).

Author contributions

B.H. and X.H. designed studies and contributed to the original concept of the project; C.W. performed cloning and functional experiment work; S.T. performed computational study on development of the mapping method; Q.F., C. Zhou and Y. Lu performed the genome sequencing; Q. Zhan, Y.W. and Z.W. carried out field planting; Q.H., D.L. and L.C. conducted the genetic and field phenotype analysis; J. Zhu, Y.S. performed rice transformation; Y.Z., Q. Zhao, Y. Li, J.M. and C. Zhu performed the genome data analysis; J.G., S.Y., J. Zhang, W.W. and H.X. performed genetic analysis of hybrid rice varieties; C.W., S.T., X.H. and B.H. analyzed whole data and wrote the paper. All of the authors discussed the results and commented on the manuscript.

Additional information

**Supplementary Information** accompanies this paper at https://doi.org/10.1038/s41467-019-11017-y.

**Competition interests:** The authors declare no competing interests.

**Peer review information:** Nature Communications thanks Xiangdong Fu, and the other anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

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