
Supplementary information

**A quantitative genomics map of rice
provides genetic insights and guides
breeding**

In the format provided by the
authors and unedited

Supplementary Note

Analysis of G×E interaction

We analyzed the effect of QTNs in two *japonica* populations at nine environments across latitudes ranging from 18° N to 50° N. The result showed that, although the estimates of QTN effects were variable in different environments (Extended Data Fig. 5), the direction of nearly all QTNs was consistent among different environments (except *Hdl* for heading date, plant height and grain length). According to the estimations from the two *japonica* populations, the general pattern is as following: under long day conditions, the effect magnitude of heading date QTNs was decreased by the increase of latitude; in contrast, for QTNs underlying agronomic traits with high heritability (e.g., grain length and grain width, of which the broad sense heritability is 0.94 and 0.91, respectively), the effects were less affected by environments. These results indicate that when using photoperiod-related QTNs in molecular breeding, G×E and trait heritability must be kept in mind.

Optimization of breeding route and selection standards

Besides the QTN map, the breeding route needs to be well designed to reduce the breeding time in a cost-efficient way. However, no benchmarking experiments on breeding design and route optimization had been reported previously. Hence, we performed *in silico* breeding, where different combinations of diverse parameters were used for breeding, using the rice high-density linkage map¹. In each ‘breeding season’, a number of rice individuals from one population were ‘planted’ for genotyping and selecting. Many genetic parameters (e.g., heterozygosity) were extracted from the whole-genome genotype of each individual. To determine which parameters are more important, we designed benchmarking experiments to compare the performances. The experiments suggested that whole-genome heterozygosity and the proportion of homozygous genotype of the recurrent parents are the major factors in the intermediate breeding process while other parameters including the size and the number of introgression segments are secondary (Fig. 5a). Thus, the comparisons provided important information for use in optimization of the selection standards.

Setting the whole genome heterozygosity as a key factor, simulation experiments demonstrated that an increase of population size for each generation could greatly promote the likelihood and breeding speed (that is, shorten the number of breeding generations required) (Fig. 5b). The number of offspring selected for the next generation is also a key factor influencing the likelihood that expected genotypes are obtained in the last generation. We found when a number of samples in one generation could meet basic requirements (i.e. with the targeted loci being heterozygous), selecting fewer samples with least heterozygosity could be an optimal choice for improving the likelihood to obtain the ideal genotype in next generation (Fig. 5c). The influence of the size of the target genomic region on the likelihood was also estimated (Fig. 5d). It was found that when high precision selection for smaller genomic regions was needed, more generations were required to achieve an ideal genotype.

We investigated strategies to improve the probability to reach the final goal when both cost and time were fixed. Breeding costs are largely dependent on the number of total samples for genotyping, and the breeding time is generally dependent on the number of breeding generations required. With a fixed number of the total samples across multiple generations,

the simulation showed the incremental mode with 1:2:3 ratio (e.g., 1000 in BC₁F₁, 2000 in BC₂F₁ and 3000 in BC₃F₁) outperformed the decreasing mode (e.g., 3000 in BC₁F₁, 2000 in BC₂F₁ and 1000 in BC₃F₁), the constant mode (e.g., 2000 in BC₁F₁, 2000 in BC₂F₁ and 2000 in BC₃F₁), and other incremental ratios (e.g. 1:2:7) (Fig. 5e).

Moreover, when faced with multiple choices on QTG targets, the genetic position of QTG had an important role on the breeding efficiency. Typically, the simulation showed that the probability for a breeding selection of a QTG near the centromere was obviously lower than the QTGs residing in other regions (Fig. 5f). This may be explained by the low recombination rate at centromeres. Conversely, the corresponding likelihood for QTGs in telomeres was much elevated, which may be correlated with the higher recombination level around telomeres. In addition, a single cross-over event was enough to facilitate the likelihood of these QTGs being selected.

When selecting multiple QTGs in one breeding program, we found enlarging the population size for each generation proved to be more effective for reducing the generation times, with at least one fewer generations needed when compared to a smaller population size (Fig. 5g). In addition, the likelihood to select QTG₁-QTG_{telomere} shows much higher likelihood than the QTG₁-QTG_{normal} and QTG₁-QTG_{centromere} combinations (Fig. 5h).

Three main modules in RiceNavi system

In this system, given specified agronomic traits and a specific rice line to be improved, RiceNavi-QTNpick provides the basic information on the genotypes of QTGs controlling the traits. The information describes gene functions and qualitative effects, quantitative estimation from GWAS cohorts, candidate donor parents with rice types, the original planting area and common trait records, and potential genetic drag effects around the QTG (Fig. 6a). Screening for suitable donor parents can be performed by using either the built-in QTN library or sequencing data of any new rice lines provided by users. With the guide from RiceNavi-QTNpick, the final donor parent(s) and the final target QTG need to be selected from the list of candidates through an overall consideration of the type of donor parents (e.g., *indica* or *japonica* rice), phenotypic differences (e.g., similar flowering time) and allelic effects of the QTG (e.g., large or minor effects).

Once the target QTGs and the population size of each generation for breeding are set by the user, RiceNavi-Sim simulates the genotypes from BC₁F₁ to BC_nF₁ population (Fig. 6b,c), where the number of backcrossing times (*n*) can be set by the user. RiceNavi-Sim can estimate the likelihood of obtaining the ideal genotype for each backcrossing generation (Fig. 6c). The user can set different population sizes considering the total expense, as well as different target genes, to determine the optimal breeding strategy.

Following the given route, the breeding population in each generation is extensively genotyped by adopting a low-coverage sequencing based method¹ or other similar approaches. RiceNavi-SampleSelect examines whole-genome genotypes of the entire population and screens for a few optimal lines with all target QTN introduced and having the highest proportion of recurrent parent background. As indicated by the computational simulations (Fig. 5a), the parameter of whole-genome heterozygosity could provide the highest selection efficiency. Therefore, RiceNavi-SampleSelect suggests a priority ordering for all lines meeting the basic selection criteria. Furthermore, for each selected line the optimal planting

number of its offspring in the next season can be estimated by RiceNavi-SampleSelect, thus the breeding efficiency can be greatly increased

Implementation in practical breeding

HHZ is known for its good performance on eating quality, resource use efficiency, and adaptation to diverse environments. It has been widely cultivated in southern China during the last decade and was grown in over 4.5 million hectares up to 2017²⁻⁴. HHZ is now very popular with farmers. Nevertheless, the variety can be further improved for several traits, including favorable rice fragrance, better adaptation to condensed planting and shorter growth period. Rice fragrance is a morphologically invisible trait, while condensed planting and growth period are complex traits controlled by multiple QTNs. For these traits, genome-guided breeding is more advantageous than conventional breeding or marker-assisted selection. Thus, we profiled the allelic state in HHZ for all QTNs underlying the three traits and screened the QTN library to search for suitable donor lines using RiceNavi-QTNpick (Fig. 7a). Through the input of whole-genome data, RiceNavi showed HHZ contains inferior alleles for *Badh2* controlling rice fragrance⁵, *TAC1* controlling tiller angle⁶ and *OsSOC1* controlling heading date⁷. Loss-of-function allele of *Badh2* can improve the aroma of rice grains. In addition, based on effect estimations from two *indica* cohorts phenotyped in eastern China, the superior allele of *TAC1* decreased tiller angle with $\sim 5.3^\circ$ and the superior allele of *OsSOC1* shortened growth period with ~ 10.3 days^{8,9}. Thus, these loci were set up as the target QTGs for improvements. Via screening the 404 genotypes in the QTN library, a variety of basmati rice from Pakistan “Basmati Surkh 89-15” (hereafter “Basmati”) was found to contain superior alleles for all the three QTGs (Fig. 7a), with no genetic drag between HHZ and Basmati lines present within 4 Mb regions of the QTG. Therefore, the Basmati line was selected as the donor parent with the aim of improving HHZ through the introduction of the three alleles with minimum changes to other parts of the HHZ genome.

Taking advantage of the benchmarking results by RiceNavi-Sim, individuals with low genome heterozygosity were intentionally selected during each generation, and an incremental population size with a general ratio of 1:2:3 was determined for our breeding design. The population size of BC₁F₁, BC₂F₁ and BC₃F₁ used in our work was 461, 908 and 1190, respectively (Fig. 7b). In the BC₁F₁ generation, individuals with less than 50% heterozygosity were selected. Using RiceNavi-SampleSelect, 10 and 3 optimal lines with a high proportion of HHZ background, low heterozygosity and small introgression segments were selected from BC₂F₁ and BC₃F₁ populations, respectively (Extended Data Fig. 9). RiceNavi-SampleSelect was used for BC₂F₁ population. For the BC₃F₂ population, preliminary screening was done using PCR markers. It was followed with whole-genome resequencing and analysis which finally yielded 21 lines with ideal genotypes. In these lines, the whole-genome data showed the majority of the genomes belonged to HHZ background with the introgression of only three small target segments (4, 4.5 and 1.9 Mb on chromosome 3, 8 and 9, respectively). Furthermore, Sanger sequencing of the three QTN validated that the ideal alleles were introduced (Supplementary Fig. 1).

A quantitative evaluation of QTN-based breeding scheme

We next used 271 introgression lines to quantitatively evaluate the performances of applying

the proposed breeding scheme. A chromosome segment substitution line (CSSL) population had been constructed by using Basmati as a donor parent and HHZ as a recurrent parent. After whole-genome resequencing of the population, 271 introgression lines that contained ≤ 3 introgression segments (Basmati genotype) and $\geq 90\%$ background (HHZ genotype) were selected. For each introgression line, changes of target traits could be predicted according to QTNs introduced from Basmati genotype. The 271 introgression lines were then phenotyped for all the traits related to grain yield and grain quality. Through a comparison between expected changes and observed phenotypic data, we found that 65.7% of samples fully matched the expectation, 15.5 % partially matched the expectation, while 18.8% samples did not match the expectation (Extended Data Fig. 10 and Supplementary Dataset 7). This result indicated that a fraction of QTNs may contain inaccurate information or be affected by epistasis/environmental interactions, but that the majority of predicted effects are fully or partially realized.

Additional tips for RiceNavi

The current pace of genetic improvement is insufficient to meet future demands²⁰. Conventional breeding selection in rice typically requires more than 10 generations of intercrossing, backcrossing and selfing to produce a genetically-stable introgression line. Thus, speed breeding has been developed to shorten crop generation times using extended photoperiods and controlled temperatures¹⁰. The required number of breeding generations could be further shortened through a genomic approach using high-throughput genotyping¹ that allows both background and foreground selections for whole-genome homozygosity and donor alleles, respectively.

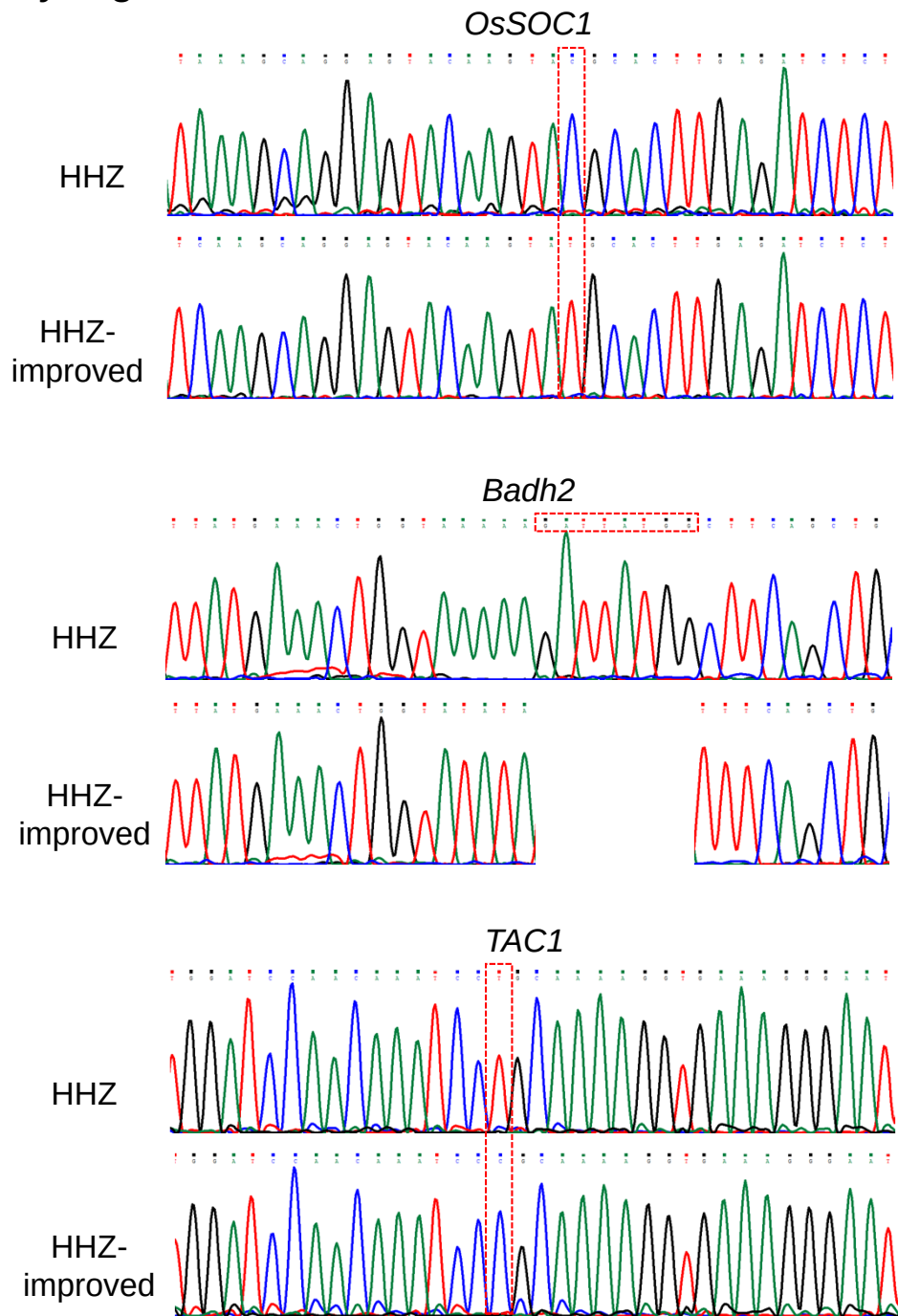
The RiceNavi system is powerful in reducing generation times required for rice breeding. In most instances, rice line improvement requires the introgression of multiple desirable genes into a single genome. If the desired target gene alleles are distributed across different accessions, QTL pyramiding needs to be performed multiple times. Through screening the QTN library or other collections with larger population sizes and higher diversity levels, it is possible to find the accessions containing several or even all of the desirable genes, which can serve as donor parents and save a large amount of time and cost in breeding. Moreover, both our *in silico* computation and real breeding data showed that the level of whole-genome homozygosity and the proportion of both parental genotypes varied widely across the individual offspring, providing a large pool for selection. Evidently, the use of a larger population size permits the occurrence of more ideal genotypes that contain the desirable alleles, exhibit higher homozygosity, contain a lower proportion of the donor genome, resulting in shorter breeding times. Theoretically speaking, the desired genotype could be identified with only two generations when using an extremely large F_2 population. However, breeding practices are subject to a trade-off of more samples for genotyping and higher cost. Thus, practical breeding requires prior estimation by the genome navigation system and an optimal balance between the cost and time.

References in Supplementary Note

1. Huang, X. *et al.* High-throughput genotyping by whole-genome resequencing. *Genome Res* **19**, 1068-1076 (2009).

-
2. Chen, S. *et al.* Genome-wide study of an elite rice pedigree reveals a complex history of genetic architecture for breeding improvement. *Sci Rep* **7**, 45685 (2017).
 3. Zhou, D. *et al.* Pedigree-based analysis of derivation of genome segments of an elite rice reveals key regions during its breeding. *Plant Biotechnol J* **14**, 638-48 (2016).
 4. Zhou, S. *et al.* Breeding and application of Huanghuazhan, a new early, middle or late rice variety with good quality and wide adaptability. *Journal of Agricultural Science and Technology* **12**, 12-17 (2010).
 5. Chen, S. *et al.* *Badh2*, encoding betaine aldehyde dehydrogenase, inhibits the biosynthesis of 2-acetyl-1-pyrroline, a major component in rice fragrance. *Plant Cell* **20**, 1850-1861 (2008).
 6. Yu, B. *et al.* *TAC1*, a major quantitative trait locus controlling tiller angle in rice. *Plant J* **52**, 891-898 (2007).
 7. Lin, H., Ashikari, M., Yamanouchi, U., Sasaki, T. & Yano, M. Identification and characterization of a quantitative trait locus, *Hd9*, controlling heading date in rice. *Breeding Sci* **52**, 35-41 (2002).
 8. Huang, X. *et al.* Genomic architecture of heterosis for yield traits in rice. *Nature* **537**, 629-633 (2016).
 9. Huang, X. *et al.* Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nat Commun* **6**, 6258 (2015).
 10. Watson, A. *et al.* Speed breeding is a powerful tool to accelerate crop research and breeding. *Nat Plants* **4**, 23-29 (2018).

Supplementary Fig. 1



Supplementary Fig. 1 | The sequencing electrophoresis of introgressed QTNs between “Huanghuazhan” and improved-Huanghuazhan. Causative variations in *OsSOC1*, *Badh2* and *TAC1* between Huanghuazhan” and improved-Huanghuazhan are indicated by red frames.

Supplementary Table 1 | Physical and genetic distance of hotspots for genetic drag of QTGs.

QTG pair	Chromosome	Physical			Genetic		
		Gene1 position (Mb)	Gene2 position (Mb)	Distance (Mb)	Gene1 position (cM)	Gene2 position (cM)	Distance (cM)
<i>OsCBL10-NOG1</i>	Chr1	29.6	31.6	2.0	141.2	147.8	6.6
<i>OsCBL10-OsBRI1</i>	Chr1	29.6	29.9	0.4	141.2	142.8	1.6
<i>NOG1-GH3-2</i>	Chr1	31.6	32.2	0.7	147.8	149.7	1.9
<i>NOG1-Pi64</i>	Chr1	31.6	33.1	1.5	147.8	152.8	5
<i>OsPP15-OsAAP6</i>	Chr1	36.3	38.1	1.8	160.5	165	4.5
<i>OsPP15-qSH1</i>	Chr1	36.3	36.4	0.1	160.5	160.6	0.1
<i>qSH1-OsAAP6</i>	Chr1	36.4	38.1	1.7	160.6	165	4.4
<i>LTG1-CAL1</i>	Chr2	24.8	25.2	0.4	92.2	93.5	1.3
<i>qLTG3-1-OsCd1</i>	Chr3	0.2	0.8	0.6	0.5	3.2	2.7
<i>LOX-3-TAC3</i>	Chr3	28.1	29.6	1.5	114.3	123.4	9.1
<i>OsTB1-TAC3</i>	Chr3	28.4	29.6	1.2	115.5	123.4	7.9
<i>Bph3-Bph3</i>	Chr4	6.9	7.0	0.0	25.5	25.6	0.1
<i>Bph3-Bph3</i>	Chr4	6.9	7.0	0.0	25.5	25.6	0.1
<i>COLD1-PiPR1</i>	Chr4	30.3	31.6	1.3	99.5	107.5	8
<i>COLD1-Pi63</i>	Chr4	30.3	31.2	0.9	99.5	106.2	6.7
<i>COLD1-NAL1</i>	Chr4	30.3	31.2	0.9	99.5	106.2	6.7
<i>COLD1-Xa1</i>	Chr4	30.3	31.7	1.3	99.5	107.8	8.3
<i>Pi63-PiPR1</i>	Chr4	31.2	31.6	0.4	106.2	107.5	1.3
<i>Pi63-NAL1</i>	Chr4	31.2	31.2	0.0	106.2	106.2	0
<i>NAL1-PiPR1</i>	Chr4	31.2	31.6	0.4	106.2	107.5	1.3
<i>NAL1-Xa1</i>	Chr4	31.2	31.7	0.5	106.2	107.8	1.6
<i>PiPR1-Xa1</i>	Chr4	31.6	31.7	0.1	107.5	107.8	0.3
<i>Xa1-OsJAZ1</i>	Chr4	31.7	33.3	1.6	107.8	113.6	5.8
<i>PTB1-Chalk5</i>	Chr5	2.6	3.3	0.7	14.6	17.7	3.1
<i>BPH29-Waxy</i>	Chr6	0.5	1.8	1.3	1.9	12.9	11
<i>HL6-APO1</i>	Chr6	27.0	27.5	0.5	100.2	104.9	4.7
<i>OsHMA3-OsNRAMP5</i>	Chr7	7.4	8.9	1.5	47.4	48.7	1.3
<i>eIF4G-OsSAP16</i>	Chr7	22.1	22.9	0.8	65.8	68.1	2.3
<i>eIF4G-Sdr4</i>	Chr7	22.1	23.8	1.7	65.8	71	5.2
<i>OsSAP16-Sdr4</i>	Chr7	22.9	23.8	0.9	68.1	71	2.9
<i>Badh2-TIG1</i>	Chr8	20.4	20.9	0.5	59.6	63.2	3.6
<i>IPA1-OsCERK1</i>	Chr8	25.3	26.9	1.6	91.4	99.3	7.9
<i>Xa13-OsCERK1</i>	Chr8	26.7	26.9	0.2	98.2	99.3	1.1
<i>OsOTUB1-OsCERK1</i>	Chr8	26.9	26.9	0.0	99.3	99.3	0

<i>DEP1-bZIP73</i>	Chr9	16.4	18.1	1.7	51.1	56.3	5.2
<i>DEP1-qRT9</i>	Chr9	16.4	17.1	0.7	51.1	53.2	2.1
<i>OsGluA2-OsbHLH174</i>	Chr10	13.5	13.7	0.2	20.7	22	1.3
<i>RGA5-LHCB5</i>	Chr11	6.6	7.7	1.1	42.5	47	4.5
<i>Nced-PSTOL1</i>	Chr12	14.2	15.5	1.3	43.7	47.1	3.4
<i>PSTOL1-xa25</i>	Chr12	15.5	17.3	1.8	47.1	48.7	1.6
<i>SNORKEL1-TONDI</i>	Chr12	25.5	27.0	1.4	88.6	98.8	10.2
<i>SNORKEL2-TONDI</i>	Chr12	25.6	27.0	1.3	89.1	98.8	9.7

Supplementary Table 2 | Comparisons of phenotypic performances between original and improved HHZ under short-day and long-day conditions.

Trait	Sanya (2019.11-2020.4)			Shanghai (2020.5-2020.9)		
	Original HHZ (<i>n</i> =47)	Improved HHZ (<i>n</i> =21)	<i>P</i> -value (two-tailed <i>t</i> -test)	Original HHZ (<i>n</i> =50)	Improved HHZ (<i>n</i> =50)	<i>P</i> -value (two-tailed <i>t</i> -test)
Heading date (days)	105.55 ± 2.75	102.48 ± 2.89	8.25E-05	82.86 ± 1.25	82.72 ± 1.07	0.55
Plant height (cm)	90.41 ± 3.35	88.27 ± 3	1.43E-02	105.13 ± 3.93	102.70 ± 3.38	1.43E-03
Panicle number	12.6 ± 2.92	12.9 ± 3.53	0.71	11.86 ± 3.06	14.2 ± 3.27	3.34E-04
Tiller angle (°)	15.35 ± 1.72	5.7 ± 2	4.38E-30	7.69 ± 1.73	1.13 ± 1.08	5.05E-57
Flag leaf length (cm)	33.65 ± 4.24	28.86 ± 3.7	3.19E-05	43.22 ± 4.30	35.37 ± 4.22	4.88E-22
Flag leaf width (cm)	1.93 ± 0.09	1.93 ± 0.1	0.95	2.02 ± 0.12	1.84 ± 0.12	1.45E-16
Grain length (mm)	9.06 ± 0.12	9.18 ± 0.22	2.86E-02	9.31 ± 0.13	8.81 ± 0.09	2.59E-36
Grain width (mm)	2.56 ± 0.05	2.59 ± 0.05	1.88E-02	2.24 ± 0.03	2.14 ± 0.02	1.35E-36
Fragrance (score)	1.02 ± 0.06	1.92 ± 0.15	1.82E-18	1.13 ± 0.28	1.85 ± 0.31	4.25E-09
Yield per plant (g, low density)	37.99 ± 12.58	41.27 ± 15.18	0.36	34.58 ± 9.65	34.78 ± 9.11	0.91
Yield per plant (g, high density)	NA	NA	NA	23.8 ± 6.60	26.92 ± 5.20	1.07E-02

Supplementary Table 3 | Data used in analysis of QTGs in open chromatin regions.

Data Type	Information	Technique	Data Origin	SRR ID
Raw Reads	YoungLeaf1	FAIRE-Seq	Zhao et al. 2020	SRR10755116
Raw Reads	YoungLeaf2	FAIRE-Seq	Zhao et al. 2020	SRR10755117
Raw Reads	Panicle1	FAIRE-Seq	Zhao et al. 2020	SRR10755118
Raw Reads	Panicle2	FAIRE-Seq	Zhao et al. 2020	SRR10755119
Raw Reads	Root1	FAIRE-Seq	Zhao et al. 2020	SRR10755120
Raw Reads	Root2	FAIRE-Seq	Zhao et al. 2020	SRR10755121
Raw Reads	MatureLeaf1	FAIRE-Seq	Zhao et al. 2020	SRR10755122
Raw Reads	MatureLeaf2	FAIRE-Seq	Zhao et al. 2020	SRR10755123
Raw Reads	Ad2-1 control	ATAC-Seq	Reynoso et al. 2019	SRR8763790
Raw Reads	ATCON2 control	ATAC-Seq	Reynoso et al. 2019	SRR8763791
Raw Reads	Ad2-5 control	ATAC-Seq	Reynoso et al. 2019	SRR8763792
Raw Reads	Ad2-4 submergence	ATAC-Seq	Reynoso et al. 2019	SRR8763795
Raw Reads	Ad2-6 submergence	ATAC-Seq	Reynoso et al. 2019	SRR8763796
Raw Reads	Ad2-8 submergence	ATAC-Seq	Reynoso et al. 2019	SRR8763797
GFF	Merged open chromatin regions	ATAC-Seq	Wilkins et al. 2016	www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE75794

Supplementary Table 4 | Primers used in Sanger sequencing of three genes.

Gene name	Gene ID	Primer type	Primer sequence
<i>OsSOC1</i>	Os03g0122600	Forward	5' TAAGTCTCCTCGGCTTCC 3'
<i>OsSOC1</i>	Os03g0122600	Reverse	5' TCTGTTTCCTAATGTCGTGTT 3'
<i>Badh2</i>	Os08g0424500	Forward	5' GAACTATCCTCTCCTGATGG 3'
<i>Badh2</i>	Os08g0424500	Reverse	5' GAACTGGCTACTAGAATGATG 3'
<i>TAC1</i>	Os09g0529300	Forward	5' AATTACTGCTTGAGGCTACA 3'
<i>TAC1</i>	Os09g0529300	Reverse	5' CGTCGTGGTATGGACAATA 3'