Supplementary Information for

Identifying a large number of high-yield genes in rice by pedigree analysis, whole genome sequencing and CRISPR-Cas9 gene knockout

Ju Huang[a,1], Jing Li[a,1], Jun Zhou[b], Long Wang[a], Sihai Yang[a], Laurence D. Hurst[c,2], Wen-Hsiung Li[d,e,2], Dacheng Tian[a,2]

[a]State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing 210023, China; [b]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, USA; [c]Department of Biology and Biochemistry, University of Bath, Bath, U.K, BA2 7AY ; [d]Biodiversity Research Center, Academia Sinica, Taipei 115, Taiwan; [e]Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637

[1]J.H.and J. L. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: whli@uchicago.edu (W-H. L.) dtian@nju.edu.cn (D. T.) or bssldh@bath.ac.uk (L.D.H.)

This PDF file includes:

Supplementary text
Figs. S1 to S8
Tables S1 to S17
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Datasets S1 to S2

Supplementary Information Text

SI Materials and Methods

Read mapping and variant calling

The paired-end reads of all the rice accessions were aligned against the Nipponbare reference (IRGSP-1.0) using Stampy (1) and BWA (2) to speed up mapping. The mapping results were then processed according to the GATK best practice (3, 4) to cleanup over-sequenced DNA molecules and false positive variants due to mapping artifacts.

Joint SNP calling and genotyping of all sequenced samples was performed using the GATK UnifiedGenotyper package, with a minimum requirement of 20 for the score of mapping quality and 50 for the minimum threshold to make a confident call. To ensure the accuracy of genotyping for each sample, we used the following criteria: i) the sample should be covered by at least 5 reads but no more than 60 reads; ii) the genotype quality should be ≥20; iii) excluding heterozygous genotypes (in a selfed rice plant the heterozygous alleles are likely to have resulted from non-allelic sequence alignments, caused by copy number variations or translocated segments). At the end, a total of 5,206,873 SNPs across all sequenced samples were called and used to estimate diversities among samples or groups.

On the other hand, 677,992 and 541,778 SNPs were called in the two parent-offspring trios, i.e., the starting trio of DGWG-Peta-IR8 and the terminal trio of IR30-Gui630-MH63 (Fig. 1), so that we could ensure the accuracy of the pedigree analysis (see *SI Appendix*, Fig. S4 for details). After that, more stringent filtering steps were implemented to obtain the quality markers used in pedigree analysis: i) clear bi-allelic SNPs in at least half of all the samples; ii) excluding SNPs in annotated transposable element regions; and iii) excluding SNPs that only appeared in a single sample. Finally, 592,603 and 481,385 high-quality SNP markers in the trios of

DGWG-Peta-IR8 and IR30-Gui630-MH63, respectively, were used for block inference and inheritance analysis (*SI Appendix*, Fig. S4 and Dataset S1).

Block inference and inheritance analysis

The ancestral status of the SNP markers in IR8 and MH63 can be determined by analyzing the two sets of trios (5). Therefore, merging consecutive markers with the same ancestral genotype allowed the identification of initial blocks. Briefly, blocks with at least 15 markers and longer than 10 kb were used as seeds, each of which was used to extend itself to the nearest block of another ancestral type. However, in a few inferred blocks, regions with limited reliable markers (defined as <1 marker in 10-kb region) or regions with a high density of markers (e.g., ≥5 markers clustered within a 100-bp segment) between its two ancestral genotypes were excluded because such blocks could result in potential errors in inferring blocks possibly owing to misalignment. Some paradoxical markers were observed in a few inferred blocks, which might be caused by gene conversion (*SI Appendix*, Fig. S9) or heterozygous sites in the ancestors. In such cases, only those blocks that have at least 75% of markers inferred to one of its parental genotypes (the ratio of markers in one genotype to the other was at least 3:1) were retained. Finally, all borders in those blocks were manually checked and confirmed.

Subsequently, the candidate blocks in MH63 that were inferred to be inherited from IR30 were further compared with the ancestors IR24, IR8, DGWG and Peta to identify their inheritance status. The blocks of MH63 inherited from DGWG or Peta are expected to have the same genotypes as those of its pedigree ancestors (e.g., IR8 and IR24) and at least 80% ancestral markers that can be detected in each generation.

Identification of recombination events

Crossover events were identified through exchange of blocks of different ancestral types. Shorter blocks showed a higher risk of non-allelic sequence alignment caused by copy number variations, sequence rearrangements or translocated segments.

Therefore, the distribution of the block size was surveyed between two trio-sets and two pair-sets (IR8-IR24 and IR24-IR30). The comparisons revealed no statistical differences among large blocks (e.g., ≥400 kb) (two sample *t*-test, P-value = 0.828), but significant differences among blocks smaller than 400 kb in size (two sample *t*-test, P-value < 0.001). These results suggest that small blocks could be strongly affected by the methods used in block identification. Thus, when detecting crossover events, we only considered blocks larger than 400 kb (*SI Appendix*, Table S17).

Block inheritance probability calculation.

MH63 has a proportion of 13.4% inherited from Peta (Fig. 1). So, for a single block in Peta, its probability to be inherited in MH63 is:

$$P(inherited) = 13.4\%$$

Let p' be the probability that a block in MH63 was passed to all of the 8 descendants of MH63. Only those parts of MH63 derived from IR30 could contain the ancestor blocks of Peta or DGWG. p' could be calculated by multiplying the proportions of MH63 in the 8 descendants, which were inherited from IR30. Let $p_1 \sim p_8$ be the proportions that MH63 descendants inherited from IR30. For example, p_1 equals to 24.5% and so MH63 descendant CDR22 has 24.5% inherited from IR30. The details of other proportions are provided in *SI Appendix*, Table S4. p'could be calculated as:

$$p' = \prod_{n=1}^{8} p_n$$

$$=24.5\%\times24.5\%\times24.5\%\times24.5\%\times24.5\%\times12.2\%\times18.4\%\times6.1\%$$

$$=1.21\times10^{-6}$$

Thus, the probability that a block in Peta was shared by MH63 and all of its 8 descendants is:

$$P(shared) = P(inherited) \times p' = 13.4\%\times1.21\times10^{-6} = 1.62\times10^{-7}$$

In fact, the total length of blocks from Peta shared by all 8 MH63 progenies is 8.76M, which constitute 2.4% of the 370M genome (*SI Appendix*, Table. S8). This is much higher than *P(shared)*, providing evidence of selection.

For a single block in DGWG, $P(inherited)$ =3.9%,

$$P(shared) = P(inherited) \times p' = 4.72\times10^{-8}$$

The total length of blocks from DGWG shared by all 8 MH63 progenies is 6.72M, 1.82% of the genome (*SI Appendix*, Table. S8), which is also much higher than the corresponding $P(shared)$.

The extreme low probability for a block from DGWG or Peta to be shared by all 8 MH63 progenies could be proved in another way. Assume that one chromosome has n blocks derived from a certain ancestor (DGWG or Peta), each with a probability p to be present in MH63. Then, the probability of at least one block to be present in MH63 is

$$P(inherited) = 1 - \sum_{i=0}^{n} C_n^i (p)^i \ (1-p)^{(n-i)}$$

Let p' be the probability a block in MH63 be passed to all of the 8 descendants of MH63:

$$p' = 1.21\times10^{-6}$$

Then P(shared), the probability that at least one block in IR8 derived from either DGWG or Peta is retained in all 8 MH63 descendants is

$$P(shared) = P(inherited) \times p'$$

For example, chromosome 3 contains the largest number of 17 (n) Peta blocks, each of which has p = 13.4% to be present in MH63 and $p' = 1.21\times10^{-6}$ to be in all of the 8 MH63 descendants (Fig. 1A and *SI Appendix*, Table. S4). Thus, the largest probability that chromosome 3 has one or more Peta blocks in MH63 and in all of its 8 descendants is:

$$P(shared) = (1- C_{17}^0 (p)^0 (1-p)^{(17-0)}) \times 1.21\times10^{-6} = 1.11\times10^{-6}$$

For a DGWG block, p=3.9%.

$$P(shared) = (1- C_{17}^0 (p)^0 (1-p)^{(17-0)}) \times 1.21\times10^{-6} = 5.95\times10^{-7}$$

These probabilities indicate that there should be virtually no chance to have one or more DGWG or Peta blocks in any chromosome in OLP in the absence of selection.

Identification of selectively targeted gene loci in ancestral blocks of DGWG and Peta

As shown in Fig. 1B, we first identified gene loci stably inherited from DGWG and Peta to MH63 and all of its eight descendants. Second, π (polymorphic sites/informative sites) was calculated for each 10-kb window to compare the shared regions to the corresponding regions in the other five collateral varieties, namely, 9311, IR1544, SH527, IR36 and IR26. When the corresponding regions are also derived from Peta or DGWG, the diversity in these regions is expected to be zero or very low. (When a cultivar experienced many crosses with different parents, gene conversion could occur; or different cultivars from the same parents could have a few sequence differences, e.g., the IR26 and a sister line of UPR221, a parent of IR30 in Fig. 1A, were derived from the same F3 plant but 1/8 of the time a heterozygous site are expected to be different in the descendants). To identify the gene loci that were subjected to strong directional selection, which was also applied to the collateral varieties, we required a diversity value <0.0001 (about 1/50 of the average diversity among landraces (6)) and that at least 3 collaterals had identical alleles.

Third, 8 tall rice landraces and 1 wild rice variety in addition to 10 wild rice varieties from a previous study (6) (a detailed list in *SI Appendix*, Table S1) were used to exclude the non-targeted genes. With the genes sorted by π values, we first removed the bottom 50% of genes with lower π values compared to wild cultivars and then filtered out the bottom 50% with lower π values compared to tall cultivars. Only those genes with low divergence values within groups (e.g., high-yielding and similar height), but high divergence compared to the tall landraces or wild rice, were retained.

The genes identified by our criteria above were under strong artificial selection during the breeding process. However, if two or more genes overlap completely or partly within the same locus, our criteria cannot determine which one was the target gene. For this reason, genes are defined as gene loci, in which the two members of an overlapping pair are counted as one locus. As a consequence, the final list of genes

identified in Fig. 1B is 159 genes from the 129 gene loci, which include 28 loci with two or more overlapping genes.

Construction of CRISPR/dCas9 Vectors and target gene loci knockdown

We cloned dCas9:BRD block from pEGB 35s:dCas:BRD:tNos (addgene No.75401)(7) and replace Cas9 on pRGEB31 to construct permanent dCas9 knockdown plasmid. For each target locus, gRNAs were designed to target specific sites and a pair of DNA oligonucleotides with appropriate cloning linkers were synthesized (BGI, Inc). Each pair of oligonucleotides were phosphorylated, annealed, and then ligated into BsaI-digested dCas9 knockdown plasmid. Then knock-down plasmids were transformed in the same way as knockout plasmids. Japonica cultivar, Taipei, was used as the background.

RNA extraction and QPCR

Plant RNA extraction and reverse transcription were carried out with minibest universal RNA extraction kit (TAKARA) following the manual. QPCR was done with SYBR Green qPCR Master Mix(TAKARA) on Step-One Plus Real-time system(ABI)

Yield-related genes enrichment calculation

The information of 1051 reported functional genes was downloaded from Q-TARO (http://qtaro.abr.affrc.go.jp). Under the definition that morphological traits and physiological traits are yield-related, there are 554 reported yield related genes among the 37869 genes in the rice genome. Based on this ratio (1.5%), 2.33 reported genes are expected among the 159 genes we identified (contained in 129 gene loci). However, we find 6 reported genes ($p=0.077$, $\chi2=3.11$, $df=1$, Chi-squared test with Yate's correction). In addition, we also find 15 genes in knockout experiments and 10 in knockdown experiments that control phenotypic changes, adding up to 31 genes in total. This means that there are significantly more yield related genes in the genes we identified than expected ($p<0.001$, $\chi2=334$, $df=1$, Chi-squared test with Yate's correction.

SI References

1. Lunter G, Goodson M (2010) Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res* 21:936–939.

2. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.

3. DePristo MA, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5):491–498.

4. McKenna A, et al. (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297–1303.

5. Conrad DF, et al. (2011) Variation in genome-wide mutation rates within and between human families. *Nat Genet* 43(7):712–714.

6. Xu X, et al. (2011) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat Biotechnol* 30(1):105–111.

7. Vazquez-Vilar M, et al. (2016) A modular toolbox for gRNA–Cas9 genome engineering in plants based on the GoldenBraid standard. *Plant Methods* 12. doi:10.1186/s13007-016-0101-2.

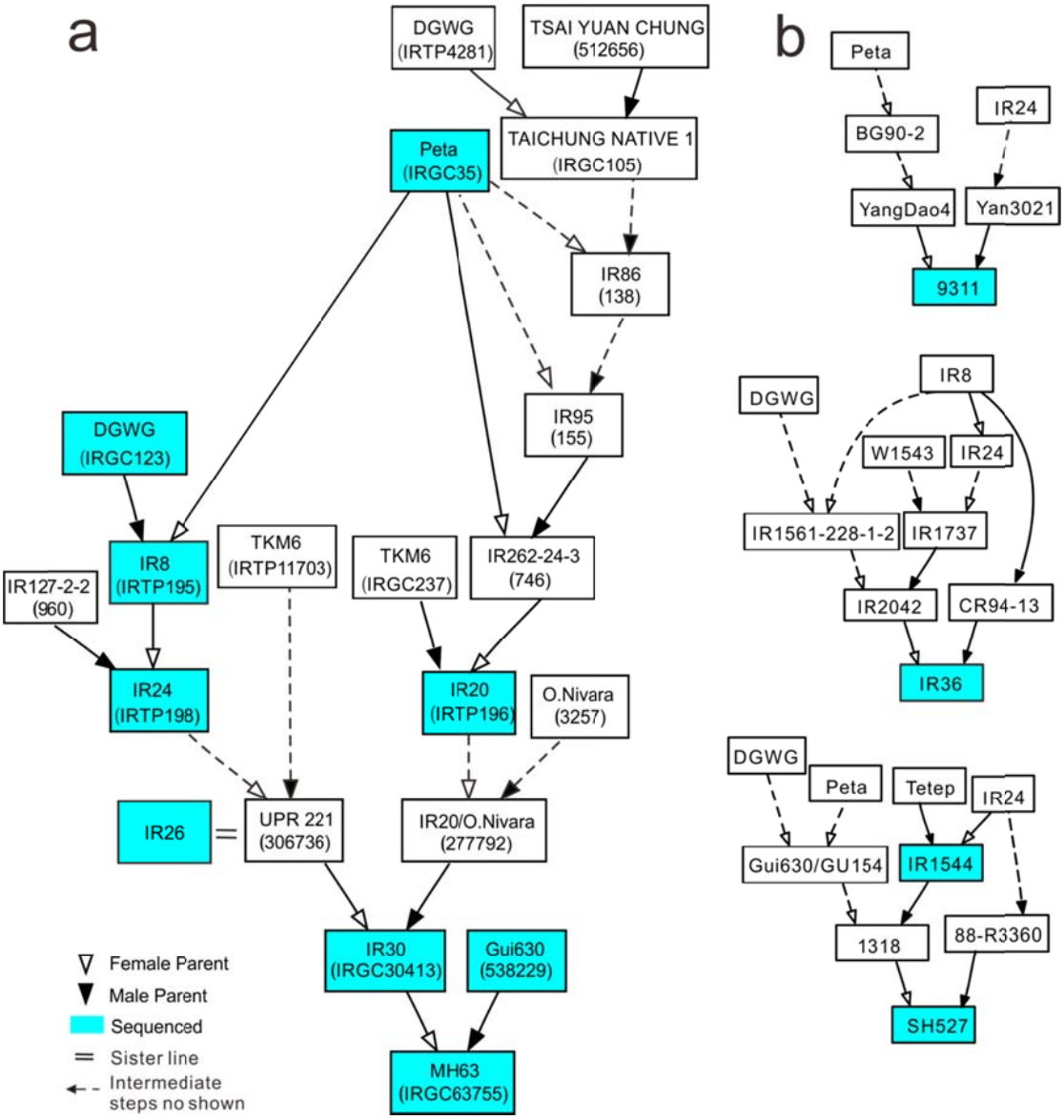


Fig. S1. Pedigree maps of the cultivars used in this study. a) The full pedigree of the major cultivars used. **b)** Brief pedigrees of 4 collateral cultivars. Data from IRRI germplasm pedigree (http://www.iris.irri.org/germplasm/) and CNRRI (http://www.ricedata.cn/variety/index.htm). The solid arrow denotes a direct parent and the dotted arrow indicates an indirect ancestor.

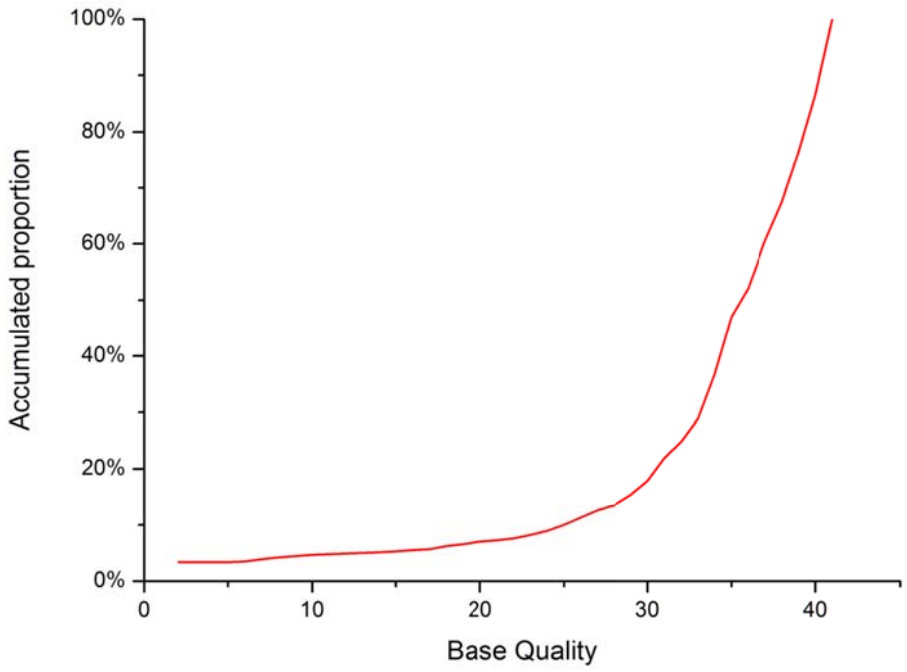


Fig. S2. Quality score distribution of all bases in sequencing reads. The y-axis is the accumulated proportion of all bases. The red line indicates the proportion of bases with phred quality lower than a given value (x-axis). More than 90% bases have phred quality score greater than 25, indicating high sequencing quality of our re-sequencing data.

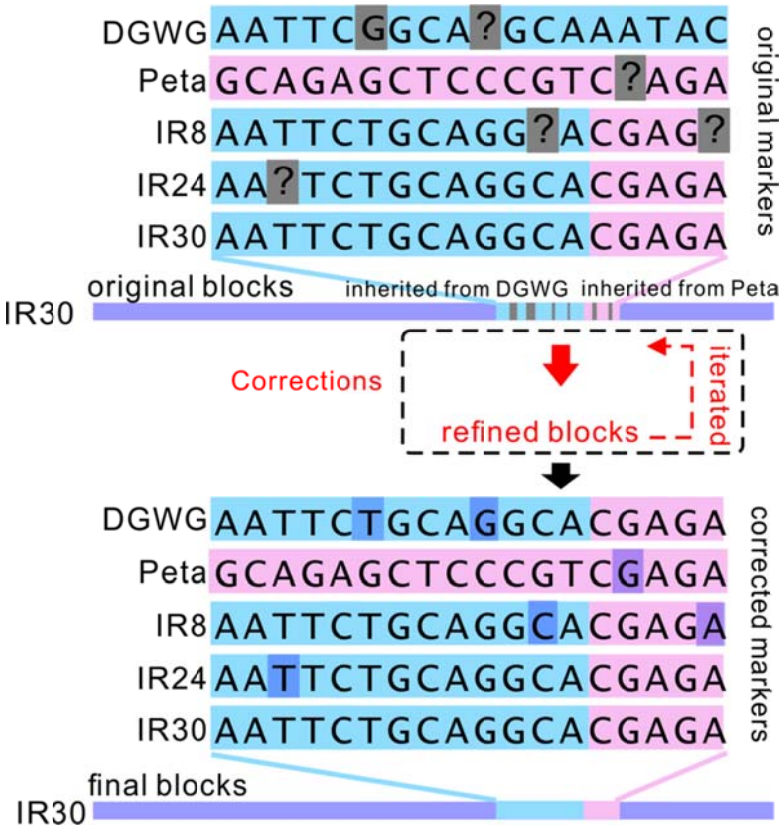


Fig. S3. Brief workflow of ancestral markers correction and ancestral blocks refinement using pedigree information. The gray question marks show that those SNPs failed in read calling or have low quality which could not pass the previous filtering steps. The gray shaded SNPs are the ones that might have a difference between the original cultivar and the current sequenced one. Through these refinement steps, 5714 markers can be inferred or corrected and added to the DGWG-Peta-IR8 trio (Dataset 2); these markers were previously not identified as markers due to their low quality, uncalled, heterozygous genotypes or difference from the original cultivar used. Iteration of these correction steps could further identify reliable and useful markers, which could be used in other analysis. For example, we could verify the gray question marks and infer the missing SNPs by comparing the sequences of the ancestor (X) and descendants IR8, IR24, and IR30. More specifically, the gray SNP "G" in the DGWG block represents an unknown genotype that is possibly a within-cultivar polymorphism between the original cultivar and the current sequenced one. However, it could be corrected to a "T" based on the allelic sequences of Peta and its descendants. Similarly, we could also infer more accurate SNPs or missing SNPs in descendant blocks based on the sequences of the ancestor and other descendants. In addition, we could improve the accuracy of the SNPs or other variant [e.g., insertions/deletions (indels)] calls, and more accurately discriminated blocks from either ancestor.

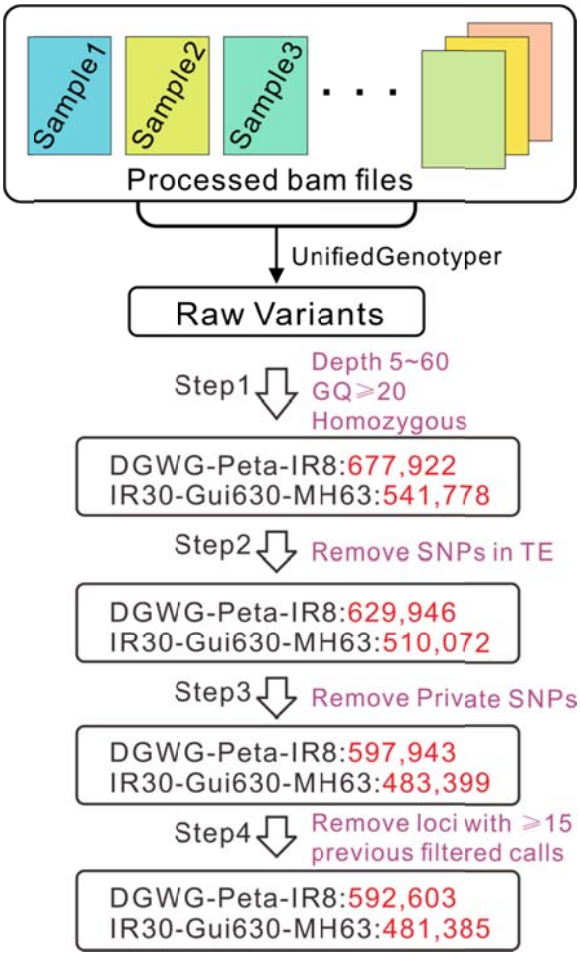


Fig. S4. Flowchart of variant filtering and marker preparation for two trios.

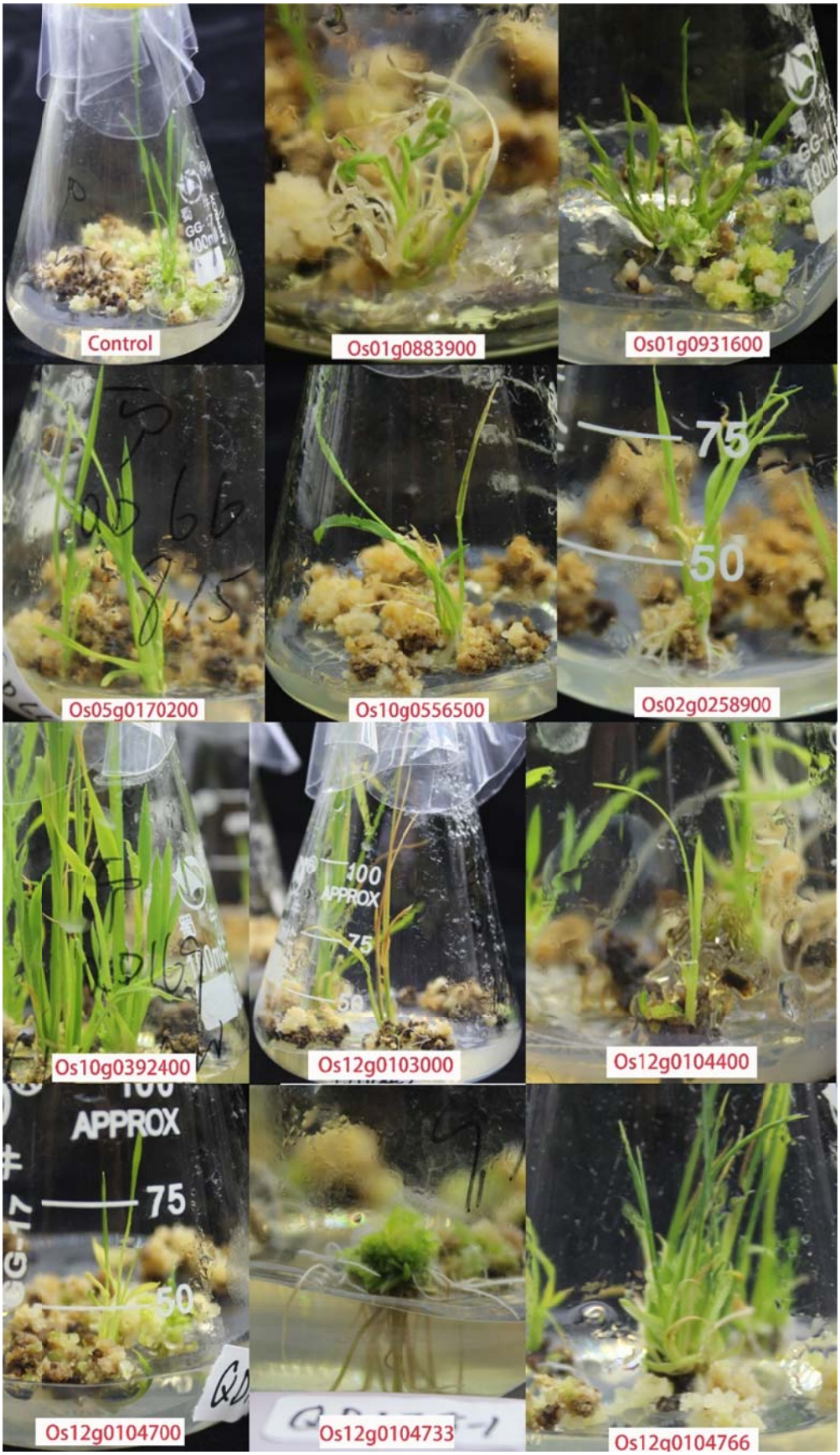


Fig. S5 Phenotypic changes of knockdown transformants

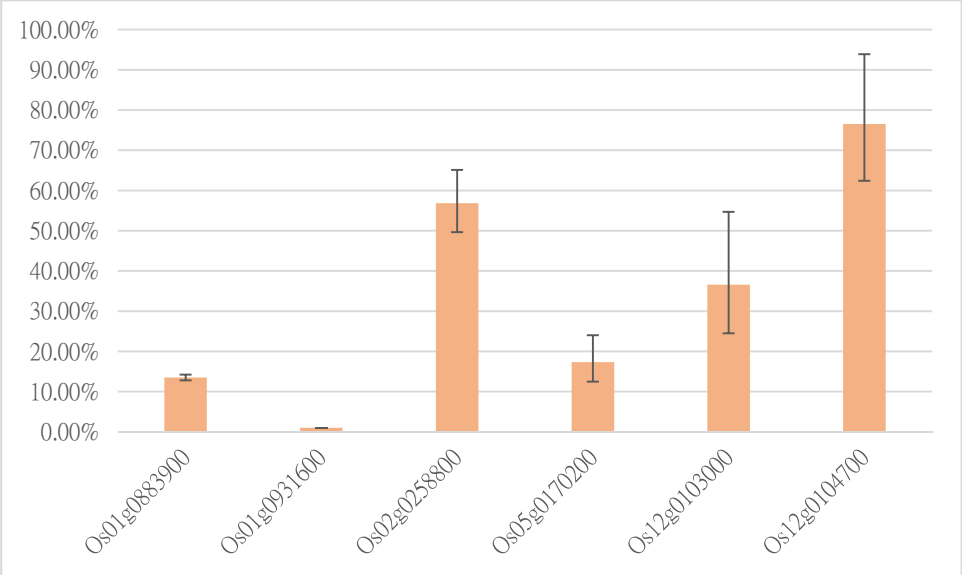


Fig. S6. Relative expression levels of 6 knockdown loci

Among the 10 knockdown genes, which exhibited distinctly phenotypic changes, only 6 had enough leaves for RNA extraction to test their relative expression levels compared with their respective wild type. Each of them showed a relative lower expression level than in the wild type, ranging from 0% to 80%.

We used pEGB 35s:dCas:BRD:tNos(GB1172) to knockdown expression, which was a gift from Diego Orzaez (Addgene plasmid # 75401). They found expression levels ranging from 20% to 100% in transformants. But we did not test the knockdown efficiency on many different genes. One of our knocked-down loci, Os01g0931600, showed an extremely low relative expression level, 3×10^{-7}% of the wild type. The knockdown transformants showed severe phenotypic changes.

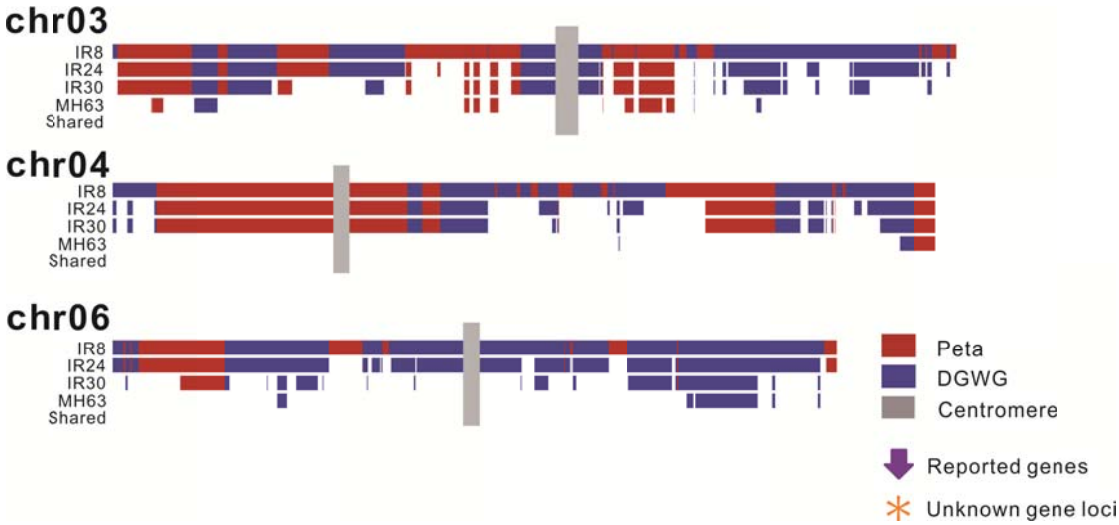
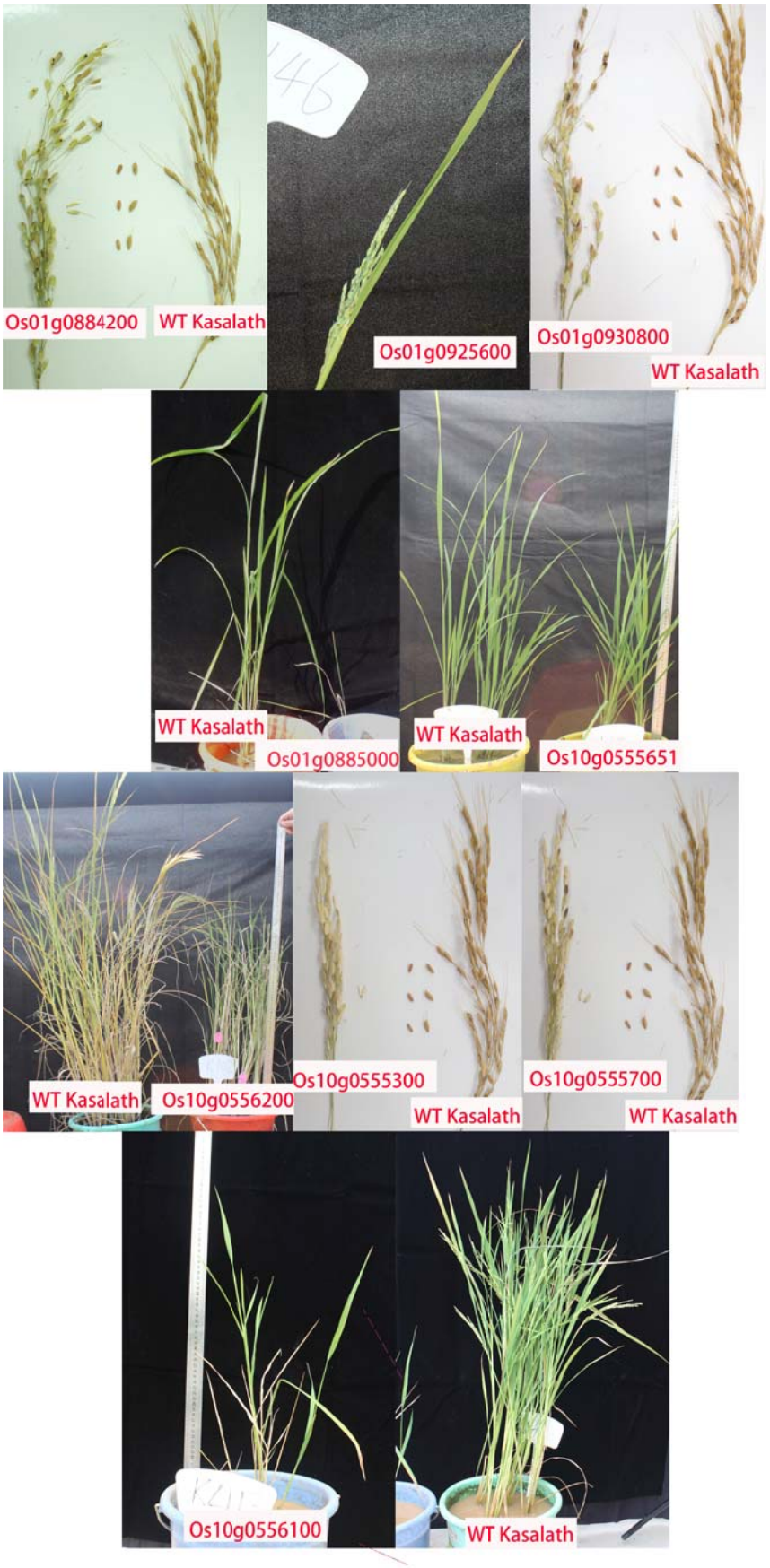


Fig. S7. Blocks inherited from DGWG and Peta in IR8, IR24, IR30, MH63 and the 8 descendants of MH63 on Chromosomes 3, 4 and 6.


Os01g0884200
WT Kasalath
Os01g0925600
Os01g0930800
WT Kasalath
WT Kasalath
Os01g0885000
WT Kasalath
Os10g0555651
WT Kasalath
Os10g0556200
Os10g0555300
WT Kasalath
Os10g0555700
WT Kasalath
Os10g0556100
WT Kasalath

b

C

Fig. S8. Additional mutant pictures of target gene loci in positive and negative controls. A) 9 knockout mutants of target gene loci; B) 4 positive controls showed mutant phenotypes similar to those reported previously; C) negative controls showed no significant changes in knockout mutants compared with the wild type, and so only one example is shown here.

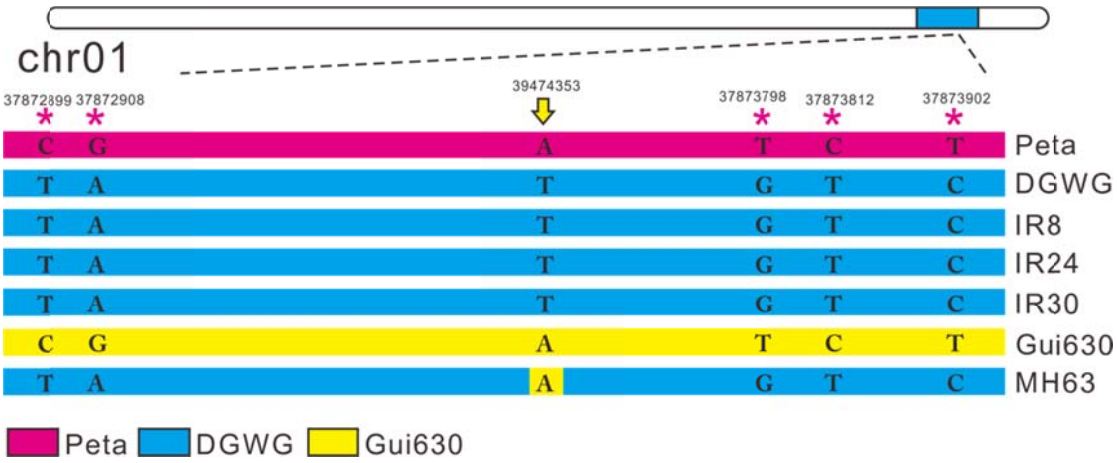


Fig. S9. Graphic representation of a gene conversion event in which a Gui630 block was introduced into MH63 within a background of DGWG in chromosome 1. A total of 2,358 markers from the IR30-Gui630-MH63 trio were identified in this background block (2,646,745bp). Sequences were colored by the inherited source of its ancestral parent. The position of the converted marker was marked by a yellow arrow. This converted SNP A was maintained in 7 descendants of MH63. Polymorphism markers used to denote the DGWG background were marked with pink asterisks.

Table S1. Summary of rice cultivars sequenced.

Sample No.	Sample Name	Sample ID	Source	Group	Total reads	Raw read depth [a]	Genome			Gene Regions		
							Mean Depth	Coverage (All mapped reads)	Coverage (Reliable mapped reads [b])	Mean Depth	Coverage (All mapped reads)	Coverage (Reliable mapped reads [b])
1	Pitai	Peta	IRRI	Major	87,258,474	23.4	21.5	88.4	75.8	19.6	90.2	86.6
2	Dee geo woo gen	DGWG	IRRI	Major	90,508,600	24.2	22.4	89.2	76.7	19.0	90.9	87.4
3	IR8	IR8	CNRRI	Major	87,726,432	23.5	21.6	88.3	75.9	19.1	90.0	86.6
4	IR24	IR24	CNRRI	Major	93,843,410	25.1	23.5	89.6	77.9	23.1	92.2	88.9
5	IR20	IR20	IRRI	Major	92,335,680	24.7	22.3	89.2	76.4	20.2	91.6	87.8
6	IR26	IR26	CNRRI	Major	93,155,880	25.0	23.2	90.0	78.5	22.1	92.7	89.4
7	IR30	IR30	IRRI	Major	85,963,262	23.0	21.3	88.2	75.7	19.8	89.9	86.4
8	GUI630	Gui630	CNRRI	Major	90,148,266	24.2	22.1	89.1	76.7	19.7	90.7	87.2
9	Minghui63	MH63	CNRRI	Major	86,495,846	23.2	21.3	88.6	76.2	19.5	90.6	87.0
10	CDR22	CDR22	CNRRI	Descendants	91,089,560	24.4	22.6	90.2	78.6	21.6	92.7	89.4
11	Chenghui047	CH047	CNRRI	Descendants	90,229,126	24.2	22.5	90.3	78.9	21.5	92.6	89.4
12	Fuhui838	FH838	CNRRI	Descendants	83,219,546	22.3	20.8	89.7	77.8	20.3	92.1	88.7
13	Hanghui570	HH570	CNRRI	Descendants	89,043,808	23.9	21.9	89.9	77.9	20.5	92.1	88.8
14	Minghui70	MH70	CNRRI	Descendants	86,022,882	23.0	21.5	89.6	77.3	21.0	92.4	89.0

15	Mianhui725	MH725	CNRRI	Descendants	91,077,382	24.4	22.3	86.1	74.7	22.4	87.1	84.0
16	Minghui86	MH86	CNRRI	Descendants	90,646,724	24.3	22.4	90.1	78.4	21.6	92.5	89.2
17	Wanhui88	WH88	CNRRI	Descendants	90,286,724	24.2	22.7	91.5	79.6	22.6	95.2	92.0
18	9311	9311	CNRRI	Collateral	95,787,192	25.7	23.7	92.5	79.9	22.3	96.5	93.2
19	IR1544	IR1544	CNRRI	Collateral	90,450,672	24.2	22.7	91.0	79.2	22.4	95.0	91.8
20	Shuhui527	SH527	CNRRI	Collateral	79,846,444	21.4	19.8	88.0	76.0	19.0	90.1	86.8
21	IR36	IR36	IRRI	Collateral	95,179,540	25.5	23.7	88.9	76.8	23.1	91.5	88.2
22	Chuannong422	CN422	CNRRI	Tall	83,573,006	22.4	21.0	90.7	78.1	22.2	94.7	91.3
23	Guangchang13	GC13	CNRRI	Tall	92,595,026	24.8	23.1	90.3	78.4	21.3	92.4	89.2
24	Luchang-3	LC3	CNRRI	Tall	90,073,330	24.1	22.8	90.9	79.4	22.5	93.7	90.4
25	Nante	NT	CNRRI	Tall	85,366,722	22.9	21.6	91.6	79.4	19.7	95.5	92.1
26	Shenglixian	SLX	CNRRI	Tall	92,941,868	24.9	23.6	90.3	78.4	23.1	92.8	89.5
27	Tetep	Tetep	CNRRI	Tall	93,480,914	25.0	22.8	88.8	77.2	22.3	91.2	87.9
28	Wanlixian	WLX	CNRRI	Tall	89,342,572	23.9	22.5	90.9	78.6	22.5	93.0	89.6
29	Zhongnong-4	ZN4	CNRRI	Tall	89,338,824	23.9	22.5	90.8	78.8	21.9	94.7	91.4
30	W1543	W1543	IRRI	Wild	90,551,852	24.3	21.7	88.9	76.9	21.2	91.9	88.6

[a] Raw read depth: depth estimated by all sequenced reads

[b] Reliable mapped reads: aligned reads with mapping quality ≥ 30

Table S2. Expected and Observed proportions of Peta and DGWG blocks in IR8, IR24, IR30 and MH63 .

Cultivar	Expected		Observed	
	Peta	**DGWG**	**Peta**	**DGWG**
IR8	50.0%	50.0%	38.83%	61.17%
IR24	25.0%	25.0%	28.12%	42.31%
IR30	26.8%	7.7%	20.98%	26.88%
MH63	13.4%	3.9%	6.30%	7.39%

Table S3. **The parents of 8 MH63 descendants.** The information of cultivars was retrieved from CHINA NATIONAL RESEARCH INSTITUTE (CNRRI, http://www.ricedata.cn/variety/index.htm)

Cultivar	Female parent	Male parent
CDR22	IR50	MH63
CH047	IR2588-5-1-2	MH63
MH70	IR54	MH63
MH86	P18(IR54/MH63//IR60/Gui630)	GK148(Jing187/IR30//MH63)
MH725	PA64	MianHui501(MH63/(TaiYing1Hao/IR26))
HH570	WL1312/LunHui422	MH63
FH838	226	γ552 [a]
WH88	WanZhong3	704(Yuai/MH63)/IR1544

[a] γ552: Gamma irradiation from MH63.

Table S4. Expected proportions of MH63 and MH63's blocks from IR30 in 8 MH63 descendants. The proportion of MH63 blocks inherited from IR30 is 48.99% (=1-51.01%; the proportion in Gui630 is shown in *SI Appendix*, Table S6). Only those blocks of MH63 derived from IR30 could contain ancestral blocks of Peta or DGWG.

MH63 descendant	Proportion inherited from MH63	Proportions of MH63 blocks that are inherited from IR30
CDR22	50.0%	24.5%
CH047	50.0%	24.5%

FH838	50.0%	24.5%
HH570	50.0%	24.5%
MH70	50.0%	24.5%
MH725	25.0%	12.2%
MH86	37.5%	18.4%
WH88	12.5%	6.1%
P [a]	3.66E-04	1.21E-06

[a] P denotes the probability of a block shared by all 8 MH63 descendants.

Table S5. Number of blocks inherited in a cultivar.

Block size	Cultivar	Chromosome												Total
		1	2	3	4	5	6	7	8	9	10	11	12	
10kb ≦ size<400kb	IR8	0	10	17	9	4	11	2	1	1	10	4	13	82
	IR24	1	45	28	16	14	12	18	5	13	6	3	17	178
	IR30	15	21	10	8	10	14	1	20	5	25	24	10	163
	MH63	2	15	15	5	11	2	4	2	6	5	10	10	87
size≧ 400kb	IR8	8	20	17	15	9	11	7	4	4	15	7	12	129
	IR24	4	21	16	15	16	11	19	6	14	7	10	16	155
	IR30	17	13	16	9	10	21	2	11	14	15	14	15	157
	MH63	6	16	16	6	16	6	11	6	12	8	14	14	131

Table S6. Length statistics of blocks in MH63 inherited from different ancestors

Ancestor	No. of blocks	Total length	Minimum length	Maximum length	Average length	Genome coverage
DGWG	57	27,569,346	14,721	7,575,570	483,673	7.39%
Peta	59	23,515,592	13,554	2,183,924	398,569	6.30%
IR24-OP	79	25,269,808	10,918	2,210,496	319,871	6.77%
IR30-OP	102	41,127,359	15,524	5,996,696	403,209	11.02%
Gui630	109	190,397,494	13,104	14,327,468	1,746,766	51.01%
Total	408	307,045,232	10,918	14,327,468	752,562	82.26%

Table S7. Number of Peta and DGWG blocks inherited in a cultivar. The number of inherited blocks in a cultivar was counted as a whole, no matter how many pieces for a block arose in the next generation. As one large block may be broken into several segments from the parent to an offspring, the number of blocks observed will increase with generations in general.

Inherited source	Cultivar	Chromosome												Total
		1	2	3	4	5	6	7	8	9	10	11	12	
DGWG	IR8	4	15	17	12	7	11	4	3	3	13	6	13	108
	IR24	4	29	16	15	12	16	10	6	12	14	3	15	152
	IR30	7	26	14	13	12	16	10	10	10	13	12	10	153
	MH63	3	8	3	2	7	5	2	1	4	7	10	5	57
	Shared [a]	2	2	0	0	4	0	0	1	0	3	2	1	15
Peta	IR8	4	15	17	12	6	11	5	2	2	12	5	12	103
	IR24	4	17	13	7	9	6	12	1	2	12	4	11	98
	IR30	8	15	11	7	9	2	12	0	2	19	5	11	101
	MH63	2	6	9	1	8	0	7	0	2	14	5	5	59
	Shared [a]	1	3	0	0	5	0	6	1	0	13	1	0	30

[a] Shared: the number of inherited blocks in the regions shared by all 8 descendants of MH63.

Table S8. Details of the blocks in MH63 inherited from DGWG or Peta and the proportion of the segments shared by all 8 MH63 descendants.

Inherited source	Chromo-some	Blocks [a]	Average length	Total length	Shared length [b]
DGWG	chr01	3	1,236,103	3,708,308	2,948,707
	chr02	8	141,973	1,135,787	512,080
	chr03	3	427,396	1,282,189	-
	chr04	2	342,322	684,644	-
	chr05	7	377,948	2,645,637	1,100,465
	chr06	5	722,095	3,610,477	-
	chr07	2	936,114	1,872,228	-
	chr08	1	7,575,570	7,575,570	1,077,696
	chr09	4	184,825	739,300	-
	chr10	7	208,271	1,457,899	276,829
	chr11	10	207,410	2,074,096	85,153
	chr12	5	156,642	783,211	264,366
Peta	chr01	2	1,271,075	2,542,150	737,516
	chr02	6	185,928	1,115,570	321,982
	chr03	9	349,836	3,148,521	-
	chr04	1	859,240	859,240	-
	chr05	8	239,757	1,918,054	1,344,352
	chr07	7	688,932	4,822,522	2,093,665
	chr09	2	791,380	1,582,760	205,138
	chr10	14	218,299	3,056,183	2,773,870
	chr11	5	601,649	3,008,245	1,283,919
	chr12	5	292,469	1,462,347	-
Total		116	18,015,235	51,084,938	15,025,738

[a] Blocks: Numbers observed in MH63.

[b] Shared length: the total length of regions shared by all 8 MH63 descendants.

Table S9. Observed plant height of eight tall varieties in field experiments. All analyzed varieties were planted in an outdoor field. Plant height was measured at the end of the heading stage. The height of MH63 is provided here as a reference.

Variety	Plant Height (cm)
NT	147.2±2.2
WLX	167.4±2.3
CN422	177.6±1.8
GC13	136.2±3.9
LC3	164.4±4.0
SLX	153.0±2.1
Tetep	154.4±4.4
ZN4	153.8±3.8
MH63	113.0±2.0

Table S10. Detailed information of functionally-unknown gene loci identified under strong artificial selection. The 123 gene loci listed here correspond to the ones that passed the filtering steps in Fig. 1B. Gene annotations were retrieved from RAP-DB (http://rapdb.dna.affrc.go.jp/)
[a] Gene annotations were predicted with the transcript variant in the block. For example, Os01t0266600-01 is a transcript variant of Os01g0266600.
[b] Loci consist of two or more overlapping genes.

Ancestor	Region	Length (bp)	Locus ID	Description
DGWG	chr01:8813485-9679511	866027	Os01g0266600	Thioredoxin fold domain containing protein. (Os01t0266600-01); [a] Thioredoxin fold domain containing protein. (Os01t0266600-02)
			Os01g0266800	Cystinosin/ERS1p repeat containing protein. (Os01t0266800-01)
			Os01g0267050	Non-protein coding transcript. (Os01t0267050-01)
			Os01g0267100 [b]	MT-A70 family protein. (Os01t0267100-01)
DGWG	chr01:37602014-39226171	1624158	Os01g0868550	Hypothetical protein. (Os01t0868550-00)
			Os01g0868600 [b]	Hypothetical conserved gene. (Os01t0868600-01)
			Os01g0877300	Mitotic checkpoint family protein. (Os01t0877300-01)
			Os01g0883900	Protein of unknown function DUF248, methyltransferase putative family protein. (Os01t0883900-01);Protein of unknown function DUF248, methyltransferase putative family protein. (Os01t0883900-02)
			Os01g0884050 [b]	Non-protein coding transcript. (Os01t0884050-00)

			Os01g0884200	Hypothetical gene. (Os01t0884200-01)
			Os01g0884400	Armadillo domain containing protein. (Os01t0884400-01)
			Os01g0884500	Plus-3 domain containing protein. (Os01t0884500-01)
			Os01g0884600 [b]	Hypothetical protein. (Os01t0884600-00)
			Os01g0885000	Similar to Cytochrome c. (Os01t0885000-01);Similar to Cytochrome c. (Os01t0885000-02)
			Os01g0885200	Uncharacterised conserved protein UCP031277 domain containing protein. (Os01t0885200-01);Hypothetical conserved gene. (Os01t0885200-02)
			Os01g0885600	Alpha/beta hydrolase fold-1 domain containing protein. (Os01t0885600-01)
			Os01g0885700	Protein kinase-like domain containing protein. (Os01t0885700-01)
			Os01g0885900	Similar to transcriptional factor TINY. (Os01t0885900-00);CGSNL Gene Symbol=ERF17;CGSNL Gene Name=ETHYLENE RESPONSE FACTOR 17;CGSNL Gene Symbol Synonym(s)=OsERF#017, OsERF017, OsERF17, AP2/EREBP#091, AP2/EREBP91;CGSNL Gene Name Synonym(s)=ethylene response factor 17, APETALA2/ethylene-responsive element binding protein 91

			Os01g0886000	Protein of unknown function DUF179 family protein. (Os01t0886000-01)
			Os01g0887700	Zinc finger, RING/FYVE/PHD-type domain containing protein. (Os01t0887700-01);Similar to Nucleic acid binding protein. (Os01t0887700-02);Similar to Nucleic acid binding protein. (Os01t0887700-03)
			Os01g0892500	Similar to carboxylic ester hydrolase. (Os01t0892500-01)
			Os01g0900700	Similar to predicted protein. (Os01t0900700-00)
DGWG	chr01:39790237-40248758	458522	Os01g0914000	Conserved hypothetical protein. (Os01t0914000-01)
			Os01g0914100	Plant lipid transfer protein/seed storage/trypsin-alpha amylase inhibitor domain containing protein. (Os01t0914100-00);CGSNL Gene Symbol= ;CGSNL Gene Name= ;CGSNL Gene Symbol Synonym(s)=OsLTPd1;CGSNL Gene Name Synonym(s)=non-specific lipid transfer protein d1, lipid transfer protein d1, type D non-specific lipid transfer protein 1
			Os01g0914200	Hypothetical protein. (Os01t0914200-01)

			Os01g0914300 [b]	Plant lipid transfer protein/seed storage/trypsin-alpha amylase inhibitor domain containing protein. (Os01t0914300-01);CGSNL Gene Symbol= ;CGSNL Gene Name= ;CGSNL Gene Symbol Synonym(s)=OsLTPd2;CGSNL Gene Name Synonym(s)=non-specific lipid transfer protein d2, lipid transfer protein d2, type D non-specific lipid transfer protein 2
Peta	chr01:40248759-40971796	723038	Os01g0925400	No apical meristem (NAM) protein domain containing protein. (Os01t0925400-01);CGSNL Gene Symbol=NAC41;CGSNL Gene Name=NAC DOMAIN-CONTAINING PROTEIN 41;CGSNL Gene Symbol Synonym(s)=ONAC041, ONAC41, ONAC050, ONAC50;CGSNL Gene Name Synonym(s)=NAC domain-containing protein 041, NAC domain-containing protein 41, NAC domain-containing protein 50
			Os01g0925600	Conserved hypothetical protein. (Os01t0925600-01)
			Os01g0925700 [b]	Similar to predicted protein. (Os01t0925700-01)
			Os01g0926501	Non-protein coding transcript. (Os01t0926501-00)
			Os01g0926600	Similar to Pectin-glucuronyltransferase. (Os01t0926600-01)
			Os01g0926800	Cellular retinaldehyde-binding/triple function, N-terminal domain containing protein. (Os01t0926800-01);CGSNL Gene Symbol= ;CGSNL Gene Name= ;CGSNL Gene Symbol

				Synonym(s)= ;CGSNL Gene Name Synonym(s)=sec14 like protein
			Os01g0927600	Similar to Auxin response factor 2 (ARF1-binding protein) (ARF1-BP). (Os01t0927600-01);CGSNL Gene Symbol=ARF2;CGSNL Gene Name=AUXIN RESPONSE FACTOR 2;CGSNL Gene Symbol Synonym(s)=OsARF2, OsARF4;CGSNL Gene Name Synonym(s)=auxin response factor-2, auxin response factor 2, auxin response factor-4, auxin response factor 4
			Os01g0927900	Similar to Aspartate kinase precursor (EC 2.7.2.4). (Os01t0927900-01)
			Os01g0928100	Similar to expressed protein. (Os01t0928100-01)
			Os01g0928200	Conserved hypothetical protein. (Os01t0928200-00)
			Os01g0928300	Prefoldin domain containing protein. (Os01t0928300-01)
			Os01g0928600	Serine palmitoyltransferase. (Os01t0928600-01)
			Os01g0928700	Similar to Serine palmitoyltransferase. (Os01t0928700-00)
			Os01g0930300	RNA-directed DNA polymerase (reverse transcriptase), related domain containing protein. (Os01t0930300-01)

			Os01g0930800	Glycoside hydrolase, family 1 protein. (Os01t0930800-01);CGSNL Gene Symbol=BGLU5;CGSNL Gene Name=BETA-GLUCOSIDASE 5;CGSNL Gene Symbol Synonym(s)=Os1bglu5, Os1Bglu5;CGSNL Gene Name Synonym(s)=beta-glucosidase 5
			Os01g0930900	Similar to Short chain alcohol dehydrogenase-like. (Os01t0930900-00)
			Os01g0931100	Conserved hypothetical protein. (Os01t0931100-01);Protein of unknown function DUF1624 domain containing protein. (Os01t0931100-02)
			Os01g0931400	Similar to thiamin pyrophosphokinase 1. (Os01t0931400-01);Thiamin pyrophosphokinase, eukaryotic domain containing protein. (Os01t0931400-02);CGSNL Gene Symbol= ;CGSNL Gene Name= ;CGSNL Gene Symbol Synonym(s)=ROX1;CGSNL Gene Name Synonym(s)=Regulator of XA21-1
			Os01g0931600	Tubby, C-terminal domain containing protein. (Os01t0931600-00)
Peta	chr02:8971712-9219957	248246	Os02g0258800	Conserved hypothetical protein. (Os02t0258800-01)

			Os02g0258900 [b]	Similar to Molybdopterin biosynthesis CNX2 protein (Molybdenum cofactor biosynthesis enzyme CNX2). (Os02t0258900-01);Similar to CNX2 (COFACTOR OF NITRATE REDUCTASE AND XANTHINE DEHYDROGENASE 2); catalytic. (Os02t0258900-02)
			Os02g0259100	Conserved hypothetical protein. (Os02t0259100-01);Similar to predicted protein. (Os02t0259100-02);CGSNL Gene Symbol=IIP2;CGSNL Gene Name=ILA1 INTERACTING PROTEIN 2;CGSNL Gene Symbol Synonym(s)= ;CGSNL Gene Name Synonym(s)=ILA1 interacting protein 2
			Os02g0259850	Hypothetical gene. (Os02t0259850-00)
			Os02g0259900	Conserved hypothetical protein. (Os02t0259900-01)
			Os02g0260000	Non-protein coding transcript. (Os02t0260000-01)
			Os02g0261100	Ubiquitin-conjugating enzyme OsUBC5b. (Os02t0261100-01);CGSNL Gene Symbol=UBC5B;CGSNL Gene Name=UBIQUITIN CONJUGATING ENZYME 5B;CGSNL Gene Symbol Synonym(s)=OsUBC5b;CGSNL Gene Name Synonym(s)=Ubiquitin conjugating enzyme 5b
			Os02g0261150 [b]	Non-protein coding transcript. (Os02t0261150-00)
DGWG	chr05:4172316-4217313	44998	Os05g0169600	Serine/threonine protein kinase domain containing protein. (Os05t0169600-01)

			Os05g0169700	Similar to Sugar transporter family protein, expressed. (Os05t0169700-00)
			Os05g0169701 [b]	Hypothetical protein. (Os05t0169701-00)
			Os05g0169800	Similar to Carboxyl-terminal proteinase. (Os05t0169800-00)
			Os05g0169900 [b]	Hypothetical gene. (Os05t0169900-01)
			Os05g0170200	Similar to Universal stress protein. (Os05t0170200-01);Similar to universal stress protein family protein. (Os05t0170200-02)
Peta	chr05:4772591-5865858	1093268	Os05g0191500	Similar to stem 28 kDa glycoprotein. (Os05t0191500-01);Similar to Stem 28 kDa glycoprotein. (Os05t0191500-02);Similar to stem 28 kDa glycoprotein. (Os05t0191500-03)
			Os05g0191700	Similar to Stem 28 kDa glycoprotein. (Os05t0191700-00)
Peta	chr07:24281836-24713675	431840	Os07g0598100	Similar to Hydroxyproline-rich glycoprotein DZ-HRGP precursor. (Os07t0598100-01)
Peta	chr07:25344163-26086625	742463	Os07g0616500	Similar to (S)-2-hydroxy-acid oxidase, peroxisomal (EC 1.1.3.15) (Glycolate oxidase) (GOX) (Short chain alpha-hydroxy acid oxidase). (Os07t0616500-01);CGSNL Gene Symbol=GLO4;CGSNL Gene Name=GLYCOLATE OXIDASE 4;CGSNL Gene Symbol Synonym(s)=GLO5, OsGLO5, OsGLO4,;CGSNL Gene Name Synonym(s)=Glycolate oxidase 5,

				Glycolate oxidase 4
			Os07g0616750	Hypothetical gene. (Os07t0616750-01)
			Os07g0616800[b]	Sucrose synthase 3 (EC 2.4.1.13) (Sucrose-UDP glucosyltransferase 3). (Os07t0616800-01);CGSNL Gene Symbol=RSUS3;CGSNL Gene Name=SUCROSE SYNTHASE 3;CGSNL Gene Symbol Synonym(s)=RSus3(t), SUS3, SS3, OsScS3, RSs3, RSus3, OsSUS3;CGSNL Gene Name Synonym(s)=Rice sucrose synthase-3(t), Sucrose synthase 3, Sucrose-UDP glucosyltransferase 3, Sucrose synthase-3, rice SuSy gene 3
			Os07g0617000	Similar to Ethylene response factor 2. (Os07t0617000-01);Similar to Ethylene response factor 2. (Os07t0617000-02);CGSNL Gene Symbol=ERF65;CGSNL Gene Name=ETHYLENE RESPONSE FACTOR 65;CGSNL Gene Symbol Synonym(s)=OsERF#065, OsERF065, OsERF65, AP2/EREBP#107, AP2/EREBP107;CGSNL Gene Name Synonym(s)=ethylene response factor 65, APETALA2/ethylene-responsive element binding protein 107
			Os07g0617600	Hypothetical conserved gene. (Os07t0617600-01)

			Os07g0617700	Cyclin-like F-box domain containing protein. (Os07t0617700-01)
			Os07g0618266	Hypothetical conserved gene. (Os07t0618266-00)
			Os07g0618450	Non-protein coding transcript. (Os07t0618450-00)
			Os07g0619400	EF-Hand type domain containing protein. (Os07t0619400-01)
Peta	chr10:12931668-13082317	150650	Os10g0391100	Conserved hypothetical protein. (Os10t0391100-01)
			Os10g0391200	Cyclin-like F-box domain containing protein. (Os10t0391200-00)
			Os10g0391300	DNA-binding SAP domain containing protein. (Os10t0391300-00);CGSNL Gene Symbol=C3H62;CGSNL Gene Name=ZINC FINGER CCCH DOMAIN-CONTAINING PROTEIN 62;CGSNL Gene Symbol Synonym(s)=OsC3H62;CGSNL Gene Name Synonym(s)=Zinc finger CCCH domain-containing protein 62
			Os10g0391400	Tify domain containing protein. (Os10t0391400-01);CGSNL Gene Symbol=TIFY11E;CGSNL Gene Name=TIFY GENE 11E;CGSNL Gene Symbol Synonym(s)=OsJAZ 13, OsJAZ13,OsTIFY11e;CGSNL Gene Name Synonym(s)=Jasmonate ZIM-domain protein 13
			Os10g0391500	Conserved hypothetical protein. (Os10t0391500-01);Conserved hypothetical protein. (Os10t0391500-02);Conserved hypothetical protein. (Os10t0391500-03);Conserved hypothetical protein.

				(Os10t0391500-04)
			Os10g0392400	Tify domain containing protein. (Os10t0392400-01);CGSNL Gene Symbol=TIFY11D;CGSNL Gene Name=TIFY GENE 11D;CGSNL Gene Symbol Synonym(s)=OsTIFY11d, OsJAZ1, OsJAZ 12, OsJAZ12;CGSNL Gene Name Synonym(s)=Jasmonate ZIM-domain protein 1, Jasmonate ZIM-domain protein 12
DGWG	chr10:14904592-14947517	42926	Os10g0422000	Conserved hypothetical protein. (Os10t0422000-01)
Peta	chr10:15005837-15814847	809011	Os10g0423400	Similar to BTB/POZ domain containing protein. (Os10t0423400-00)
			Os10g0423501	Conserved hypothetical protein. (Os10t0423501-00)
			Os10g0423800	Similar to BTB/POZ domain containing protein. (Os10t0423800-00)
			Os10g0423900	BTB/POZ-like domain containing protein. (Os10t0423900-00)
			Os10g0425400	BTB/POZ-like domain containing protein. (Os10t0425400-01)
			Os10g0425700	MATH domain containing protein. (Os10t0425700-00)
			Os10g0425900	Kelch related domain containing protein. (Os10t0425900-01)
			Os10g0426100	Similar to HAT family dimerisation domain containing protein. (Os10t0426100-00)

			Os10g0426300	Similar to Exonuclease family protein, expressed. (Os10t0426300-00)
			Os10g0426450[b]	Hypothetical protein. (Os10t0426450-00)
			Os10g0426600	Conserved hypothetical protein. (Os10t0426600-01)
			Os10g0426700[b]	Hypothetical protein. (Os10t0426700-01)
			Os10g0427100	Conserved hypothetical protein. (Os10t0427100-01)
			Os10g0427300	Similar to BTB/POZ domain containing protein, expressed. (Os10t0427300-01);Similar to BTB/POZ domain containing protein, expressed. (Os10t0427300-02)
			Os10g0430200	Similar to Sinapyl alcohol dehydrogenase. (Os10t0430200-01);Similar to Sinapyl alcohol dehydrogenase. (Os10t0430200-02);Similar to Sinapyl alcohol dehydrogenase. (Os10t0430200-03);CGSNL Gene Symbol=CAD3;CGSNL Gene Name=CINNAMYL ALCOHOL DEHYDROGENASE 3;CGSNL Gene Symbol Synonym(s)=OsCAD3;CGSNL Gene Name Synonym(s)=Cinnamyl alcohol dehydrogenase 3
			Os10g0433200	Hypothetical conserved gene. (Os10t0433200-00)
			Os10g0435900	Similar to BTB/POZ domain containing protein. (Os10t0435900-00)
			Os10g0436100	BTB/POZ-like domain containing protein. (Os10t0436100-00)

			Os10g0438300	Conserved hypothetical protein. (Os10t0438300-00)
			Os10g0438301	Hypothetical conserved gene. (Os10t0438301-00)
Peta	chr10:21769689-21922126	152438	Os10g0554800	Plant lipid transfer protein and hydrophobic protein, helical domain containing protein. (Os10t0554800-01)
			Os10g0554900	Protein of unknown function DUF566 family protein. (Os10t0554900-01);Protein of unknown function DUF566 domain containing protein. (Os10t0554900-02)
			Os10g0555100	Similar to DNA chromosome 4, ESSA I CONTIG fragment NO. 6 (Glucosyltransferase like protein). (Os10t0555100-01);CGSNL Gene Symbol= ;CGSNL Gene Name= ;CGSNL Gene Symbol Synonym(s)=OsGGT;CGSNL Gene Name Synonym(s)=glycogenin glucosyltransferase
			Os10g0555200	Resolvase, holliday junction-type, YqgF-like domain containing protein. (Os10t0555200-01)
			Os10g0555300	Zinc finger, C2H2-like domain containing protein. (Os10t0555300-01)
			Os10g0555600	Beta-expansin precursor. (Os10t0555600-01);CGSNL Gene Symbol=EXPB6;CGSNL Gene Name=BETA-EXPANSIN 6;CGSNL Gene Symbol Synonym(s)=OsEXPB6, osaEXPb1.8;CGSNL Gene Name Synonym(s)=Expansin-B6,

				Beta-expansin-6
			Os10g0555651[b]	Conserved hypothetical protein. (Os10t0555651-00)
			Os10g0555700	Beta-expansin. (Os10t0555700-02);CGSNL Gene Symbol=EXPB2;CGSNL Gene Name=BETA-EXPANSIN 2;CGSNL Gene Symbol Synonym(s)=OsEXPB2, OsaEXPb1.9;CGSNL Gene Name Synonym(s)=Expansin-B2, Beta-expansin-2
			Os10g0555900	Similar to Beta-expansin. (Os10t0555900-01);Beta-expansin precursor. (Os10t0555900-02);CGSNL Gene Symbol=EXPB3;CGSNL Gene Name=BETA-EXPANSIN 3;CGSNL Gene Symbol Synonym(s)=OsEXPB3, osaEXPb1.10;CGSNL Gene Name Synonym(s)=Expansin-B3, Beta-expansin-3
			Os10g0556000[b]	Hypothetical protein. (Os10t0556000-00)
			Os10g0556100	Similar to beta-expansin EXPB4. (Os10t0556100-01);Similar to beta-expansin EXPB4. (Os10t0556100-02);Similar to Expansin-B4. (Os10t0556100-03);CGSNL Gene Symbol=EXPB4;CGSNL Gene Name=BETA-EXPANSIN 4;CGSNL Gene Symbol

				Synonym(s)=OsEXPB4, OsaEXPb1.7;CGSNL Gene Name Synonym(s)=Expansin-B4, Beta-expansin-4
			Os10g0556200	Hypothetical conserved gene. (Os10t0556200-01);CGSNL Gene Symbol= ;CGSNL Gene Name= ;CGSNL Gene Symbol Synonym(s)=OsbHLH113;CGSNL Gene Name Synonym(s)=basic helix-loop-helix protein 113
			Os10g0556500	Hypothetical conserved gene. (Os10t0556500-01)
			Os10g0556700	Hypothetical conserved gene. (Os10t0556700-00)
			Os10g0556900	Conserved hypothetical protein. (Os10t0556900-01)
Peta	chr10:21992900-22072751	79852	Os10g0558750	Similar to Oxidoreductase, 2OG-Fe oxygenase family protein, expressed. (Os10t0558750-01)
			Os10g0558801[b]	Non-protein coding transcript. (Os10t0558801-00)
			Os10g0558850	Hypothetical protein. (Os10t0558850-00)
			Os10g0559200	Similar to Oxidoreductase, 2OG-Fe oxygenase family protein, expressed. (Os10t0559200-01)
			Os10g0559300	Hypothetical protein. (Os10t0559300-01)

			Os10g0559400[b]	Similar to cDNA clone:J023138C05, full insert sequence. (Os10t0559400-01);Similar to cDNA clone:J023138C05, full insert sequence. (Os10t0559400-02)
			Os10g0559450	Conserved hypothetical protein. (Os10t0559450-01)
			Os10g0559500[b]	2OG-Fe(II) oxygenase domain containing protein. (Os10t0559500-01);2OG-Fe(II) oxygenase domain containing protein. (Os10t0559500-02)
			Os10g0559650	Hypothetical protein. (Os10t0559650-00)
			Os10g0559800	Protein of unknown function DUF547 domain containing protein. (Os10t0559800-01)
			Os10g0559833[b]	Hypothetical protein. (Os10t0559833-00)
			Os10g0559866	Hypothetical protein. (Os10t0559866-00)
			Os10g0559900[b]	Ribosomal protein L18/L5 domain containing protein. (Os10t0559900-01)
Peta	chr11:6540176-7824094	1283919	Os11g0242400	Similar to Rieske. (Os11t0242400-01)
			Os11g0242600	Protein of unknown function DUF579, plant family protein. (Os11t0242600-00)
			Os11g0242700	Similar to Ribosomal protein L37. (Os11t0242700-01)

			Os11g0242800	Similar to ASCAB9-A (ASCAB9-B) (Fragment). (Os11t0242800-01);CGSNL Gene Symbol= ;CGSNL Gene Name= ;CGSNL Gene Symbol Synonym(s)=ASCAB9-A, CP26, Lhcb5;CGSNL Gene Name Synonym(s)=light-harvesting protein ASCAB9-A, PSII CP26, PSII Lhcb5
DGWG	chr12:1-264366	264366	Os12g0100100	Similar to ALY protein. (Os12t0100100-01);Similar to ALY protein. (Os12t0100100-02);Similar to ALY protein. (Os12t0100100-03)
			Os12g0100200[b]	Conserved hypothetical protein. (Os12t0100200-00)
			Os12g0102100	Alcohol dehydrogenase superfamily, zinc-containing protein. (Os12t0102100-01);Similar to predicted protein. (Os12t0102100-02)
			Os12g0102200	LisH dimerisation motif domain containing protein. (Os12t0102200-01)
			Os12g0102300	Similar to WRKY transcription factor 57. (Os12t0102300-01);CGSNL Gene Symbol=WRKY57;CGSNL Gene Name=WRKY GENE 57;CGSNL Gene Symbol Synonym(s)=OsWRKY57;CGSNL Gene Name Synonym(s)=Rice WRKY gene57
			Os12g0102350	Similar to WRKY transcription factor 57. (Os12t0102350-01)

			Os12g0102400	Zinc finger, RING/FYVE/PHD-type domain containing protein. (Os12t0102400-01);CGSNL Gene Symbol= ;CGSNL Gene Name= ;CGSNL Gene Symbol Synonym(s)=OsRINGC2-2;CGSNL Gene Name Synonym(s)=RING-C2 type protein 2
			Os12g0103000	EF-Hand type domain containing protein. (Os12t0103000-01)
			Os12g0103150	Conserved hypothetical protein. (Os12t0103150-00)
			Os12g0103300	Preprotein translocase Sec, Sec61-beta subunit domain containing protein. (Os12t0103300-01)
			Os12g0103500	Ribosome-inactivating protein domain containing protein. (Os12t0103500-00)
			[b]Os12g0103540[b]	Non-protein coding transcript. (Os12t0103540-01)
			Os12g0103580	Peptidase C48, SUMO/Sentrin/Ubl1 domain containing protein. (Os12t0103580-00)
			Os12g0104250	Hypothetical gene. (Os12t0104250-01)
			Os12g0104400	Fatty acid desaturase, type 1 domain containing protein. (Os12t0104400-01)
			Os12g0104700	Protein of unknown function DUF231, plant domain containing protein. (Os12t0104700-01)
			Os12g0104733	Hypothetical protein. (Os12t0104733-00)

			Os12g0104766[b]	Similar to Clathrin heavy chain. (Os12t0104766-00)
			Os12g0104850	Hypothetical gene. (Os12t0104850-01)
			Os12g0104900[b]	EF-HAND 2 domain containing protein. (Os12t0104900-00);CGSNL Gene Symbol=CML26;CGSNL Gene Name=CALMODULIN-LIKE PROTEIN 26;CGSNL Gene Symbol Synonym(s)=OsCML26;CGSNL Gene Name Synonym(s)=calmodulin-like protein 26

Table S11. Expression levels in callus of the 57 genes chosen for the knockout experiment.

Expression level[a]	Gene loci with successful knockout		Gene loci with failed knockout	Total
	Phenotypic change	No phenotypic change		
Low	1	0	4	5
Medium	5	3	18	26
High	9	1	16	26
Total	15	4	38	57

[a] The expression data were retrieved from Genevestigator, a database of microarray expression data. For those loci that were absent in Genevestigator, we used the FPKM values downloaded from 9311 callus RNA-seq data (NCBI, BioProject PRJNA117345, SRR037711~SRR037724). In Genevestigator, RMA (Robust Multi-array Average), is used to summarize probe-level measurements of all arrays to give an estimate expression level, the so-called “average expression value”. The expression levels of genes in all samples are divided into three groups: "LOW", "MEDIUM" or "HIGH". "LOW"(<8.1) corresponds to the first quartile, "MEDIUM"(8.1~11.8) to the interquartile range and "HIGH" (>11.8) to the fourth quartile. The FPKM values were divided to three groups: low (<1), medium (1~10), and high (>10)

Table S12.Positive control genes and their phenotypic changes when knocked out. This table lists gRNA, gRNA site, Sanger sequencing data and phenotypic change for 6 positive control genes. The gene *rl14* was not knocked out successfully. All of the five successfully knocked out lines showed phenotypic changes.

Gene	Name	Reported phenotype	gRNA [a]	gRNA sites	gRNA	Double Knockout lines /total	Target genotypes in the haplotypes of a mutant detected by Sanger sequencing	Phenotype studied		
								Quantified phenotype	WT measurement	Mutant measurement
Os01g0883800	*sd1*	Semi-dwarf. Gibberellin biosynthesis.	TCGGCCACACGAATGGCTCC	chr01:38382613-38382632	TCGGCCACACGAATGGCTCC	4/7	H1-1:DEL(38382617-38382629); H1-2:SNP(38382615-AGCCA-G) [b]	Plant height	132.4	62.8
							H4-1:DEL(38382617-38382629); H4-2:SNP(38382615-AGCCA-G)			
							H5-1:SNP(38382615-AGCCATTCGTGTGGC-GG); H5-2:IN(38382617-G)			
							H6-1:DEL(38382617-38382629); H6-2:SNP(38382615-AGCCA-G)			
							H2-1:WT;	No	-	-

							H2-2:WT	phenotypic change		
							H3-1:WT; H3-2:WT			
							H7-1:WT; H7-2:WT			
Os01g0884300	*OsNAC6*	Seedling root growth	GACCGCCGCTCATCTCTCTC	chr01:38401360-38401379	GACCGCCGCTCATCTCTCTC	2/2	H1-1:IN(38401361-CCG); H1-2:IN(38401362-T)	Plant height	129.3	65.5
							H3-1:IN(38401361-CCG); H3-2:IN(38401362-T)			
Os05g0170000	*bc10*	Brittle culm. Culm mechanical strength. Cellulose and arabinogalactan protein levels.	GTCCCGCCAGAACTCTGGTA	chr05:4197196-4197215	GTCCCGCCAGAACTCTGGTA	4/4	H5-1:DEL(4197200-4197221); H5-2:DEL(4197189-4197208)	Plant height	130.3	47.8
							H6-1:DEL(4197195-4197199); H6-2:DEL(4197198)			
							H7-1:DEL(4197200-4197221); H7-2:DEL(4197189-4197208)			
							H8-1:DEL(4197197-4197205); H8-2:DEL(4197199-4197204)			

Os05g0187500	*GW5*	Grain size.	TTCCGCAACATGTGGGGAGG	chr05:5365143-5365162	TTCCGCAACATGTGGGGAGG	1/4	H8-1:IN(5365159-A); H8-2:IN(5365159-G)	Grain size Fig. S8	-	-
							H9-1:WT; H9-2:WT	No phenotypic change	-	-
							H10-1:WT; H10-2:WT			
							H11-1:WT; H11-2:WT			
Os10g0558900	*rl14*	Leaf rolling. Cellulose and ligin content. Stomata size. Transpiration rate.	GGCACGGATAGTAGAACAGT	chr10:21996026-21996045	GGCACGGATAGTAGAACAGT	0/3	H1-1:WT; H1-2:WT	No phenotypic change	-	-
							H2-1:WT; H2-2:WT			
							H3-1:WT; H3-2:WT			
Os02g0260200	*lp*	Panicle branching. Tillering. Culm mechanical	GAGCCTGCGTAAGTGGTATG	chr02:9043621-9043640	GAGCCTGCGTAAGTGGTATG	1/3	H1-1:DEL(9043637); H1-2:IN(9043637-A)	Panicle shape Refer to Fig. S8		Refer to Fig. S8

		strength.					H2-1:WT; H2-2:WT	No phenotypic change	-	-
							H3-1:WT; H3-2:WT			

[a] gRNAs were designed to bind the 5' region of the coding sequence of the target gene.

[b]H1-1 and H1-2 refer to the two haplotypes of the H1 mutant. The genotype of each haplotype was annotated by comparison with the wild type. WT: wild type. DEL means that this haplotype has a deletion in this region compared to the wild type. IN refers to a insertion at the downstream of the following site and SNP means that there was a single nucleotide variation at the site.

Table S13. The positions of the 10 randomly sampled genes as the negative controls for the CRISPR knockout experiments.

Gene ID	Adjacent block	Distance (bp)
Os01g0936100	Peta 40248759 40971796	124115
Os05g0375600	Peta 18169283 18185538	39647
Os05g0571300	Peta 28264741 28350688	122880
Os05g0571700		139048
Os05g0573600		227217
Os10g0341700	Peta 10158206 10350586	63171
Os10g0341750		54016
Os10g0342300		35306
Os10g0342650		35262
Os10g0558400	Peta 21992900 22072751	11874

The 10 control genes were randomly selected from within the 300 kb regions from the original Peta ancestor blocks that were shared in all 8 MH63 descendants, ranging from 11874 to 227217 bp. In total, there were 3124 unknown genes near Peta blocks. No observable changes occurred in phenotypes when these genes were knocked out.

Table S14. Negative control genes and the phenotypic changes in their knockout lines. This table lists gRNA, gRNA site, Sanger sequencing results and the phenotypic change in the 10 negative control genes.

Gene ID	Predicted function	gRNA [a]	gRNA sites	Double knockout lines /total	Target genotypes in the haplotypes of a mutant detected by Sanger sequencing	Plant height (cm) [c]	WT plant height (cm)
Os01g0936100	Similar to Protein kinase.	GATAAGGAACCTCTTGACTG	chr01:41098121-41098140	3/3	H6-1:SNP(41098124-41098127-AAGG-T); H6-2:DEL(41098124) [b]	130.3	131.5
					H7-1:DEL(41098124-41098126); H7-2:DEL(41098125)		
					H8-1:DEL(41098124-41098126); H8-2:DEL(41098124-41098126)		
Os05g0375600	Similar to Peptide chain release factor 2.	TCTTAAGGGTCCGTACGACA	chr05:18127613-18127632	7/7	H1-1:DEL(18127629-18127631); H1-2:DEL(18127629)	134.0	132.4
					H2-1:DEL(18127629); H2-2:DEL(18127630)		
					H3-1:DEL(18127629); H3-2:DEL(18127631)		
					H4-1:DEL(18127629-18127631); H4-2:DEL(18127629)		

					H5-1:DEL(18127629); H5-2:DEL(18127630)		
					H6-1:DEL(18127629); H6-2:DEL(18127630)		
					H7-1:DEL(18127629); H7-2:DEL(18127630)		
Os05g0571700	Cyclin-like F-box domain containing protein.	TTGGGTCTGAACACTTCGCT	chr05:28492551-28492570	2/4	H1-1:DEL(28492551-28492569); H1-2:IN(28492568-G)	126.5	129.2
					H3-1:DEL(28492551-28492569); H3-2:IN(28492568-G)		
					H2-1:WT; H2-2:WT	133.5	
					H4-1:WT; H4-2:WT		
Os05g0573600	Similar to predicted protein.	CATGTAGCCCGTGGCCCGCG	chr05:28578326-28578345	2/2	H1-1:DEL(28578327-28578335); H1-2:DEL(28578330-28578332)	132.0	130.1
					H3-1:DEL(28578329-28578334); H3-2:DEL(28578330-28578332)		

Os10g0341750	Hypothetical protein.	TTTGGTTGCGCATTGCAAAT	chr10:10102678-10102697	2/3	H2-1:DEL(10102691-10102693); H2-2:DEL(10102689-10102693)	134.0	130.1
					H1-1:DEL(10102677-10102697); H1-2:DEL(10102690-10102695)		
					H1-1:DEL(10102677-10102697); H1-2:WT	128.0	
Os10g0342300	Serine/threonine protein kinase-related domain containing protein.	TCCGTAGTGGCATAGCTCGT	chr10:10122833-10122852	18/18	H13-1:DEL(10122845-10122849); H13-2:SNP(10122848-CT-A)	132.0	129.2
					H14-1:DEL(10122845-10122849); H14-2:SNP(10122848-CT-A)		
					H15-1:DEL(10122847-10122849); H153-2:DEL(10122847-10122849)		
					H16-1:DEL(10122847-10122849); H16-2:DEL(10122847-10122849)		

					H17-1:DEL(10122847-10122849); H17-2:DEL(10122847-10122849)		
					H18-1:DEL(10122847-10122850); H18-2:IN(10122847-A)		
					H2-1:DEL(10122847-10122850); H2-2:IN(10122847-A)		
					H3-1:DEL(10122848-10122849); H3-2:IN(10122847-A)		
					H4-1:DEL(10122847-10122850); H4-2:IN(10122847-A)		
					H7-1:DEL(10122848-10122849); H7-2:IN(10122847-A)		
					H8-1:DEL(10122848-10122849); H8-2:IN(10122847-A)		
					H9-1:DEL(10122848-10122849); H9-2:IN(10122847-A)		
					H10-1:DEL(10122848-10122849); H10-2:IN(10122847-A)		

					H11-1:DEL(10122848-10122849); H11-2:IN(10122847-A)		
					H12-1:DEL(10122848-10122849); H12-2:IN(10122847-A)		
					H13-1:DEL(10122848-10122849); H13-2:IN(10122847-A)		
					H14-1:DEL(10122848-10122849); H14-2:IN(10122847-A)		
					H15-1:DEL(10122848-10122849); H15-2:IN(10122847-A)		
Os10g0341700	Cellulose synthase-like H1.	TGCAGGAGAGGGTGCCCATC	chr10:10089587-10089606	3/3	H1-1:DEL(10089601-10089603); H1-2:DEL(10089601-10089605)	133.0	130.2
					H2-1:DEL(10089601-10089602); H2-2:DEL(10089601)		
					H3-1:DEL(10089601-10089603); H3-2:DEL(10089601-10089605)		

Os05g0571300	Conserved hypothetical protein.	ATCACCGTGAAATGCAAACG	chr05:28476736-28476755	3/3	H1-1:DEL(28476752-28476753); H1-2:IN(28476753-AG)	134.3	132.4
					H2-1:DEL(28476752-28476753); H2-2:DEL(28476749-28476751)		
					H3-1:DEL(28476752-28476753); H3-2:IN(28476753-AG)		
Os10g0558400	Similar to Oxidoreductase, 2OG-Fe oxygenase family protein.	GTCTCCTCGTCTGACGACAA	chr10:21980728-21980747	2/2	H1-1:DEL(21980706-21980787); H1-2:DEL(21980737-21980750)	128.5	132.4
					H2-1:DEL(21980706-21980787); H2-2:DEL(21980737-21980750)		
Os10g0342650	Hypothetical protein.	ATTGAATCCACTTCGGCTCT	chr10:10121252-10121271	6/6	H1-1:DEL(10121255-10121258); H1-2:DEL(10121255-10121258)	131.3	131.5
					H2-1:DEL(10121255-10121257); H2-2:DEL(10121255-10121257)		
					H3-1:DEL(10121255-10121258); H3-2:IN(10121254-T)		

					H4-1:DEL(10121255-10121256); H4-2:DEL(10121255-10121258)		
					H5-1:IN(10121254-TT); H5-2:IN(10121254-TT)		
					H6-1:DEL(10121255-10121256); H6-2:IN(10121254-G)		

[a] gRNAs were designed to bind the 5' region of the coding sequence of the target gene;

[b] H1-1 and H1-2 refer to the two haplotypes of the H1 mutant. The genotype of each haplotype was annotated by comparison with the wild type. WT: wild type. DEL means that this haplotype has a deletion on this region relative to the wild type. IN refers to a insertion at the downstream of the following site and SNP means that there was a single nucleotide variation at the site;

[c] No obvious phenotypic changes were observed. Only plant height was measured to compare with the mutants of target genes.

Table S15. Target gene loci and phenotypic changes in the knockout lines. This table lists the gRNA, gRNA site, Sanger sequencing results and phenotypic change for each target gene locus. The 20 loci were sampled from the 123 loci with unknown function and all except Os01g884600 were successfully knocked out. For each gene locus, the gRNA was designed to target specific site at the beginning of the CDS region to cause a frame-shift mutation. Then a pair of DNA oligonucleotides with appropriate cloning linkers were synthesized and ligated into the BsaI-digested pRGEB31 vector. Each construct was transformed into callus of Kasalath (indica) or Wuyugeng24 (japonica). At least 10 transformed individuals were produced in two recipients for each vector. The rice plants were measured under natural field conditions in the Experimental station of Nanjing University. Plant phenotypes were examined every three days against the corresponding wild type plants. For each plant, genomic DNA was extracted from fresh leaves by the CTAB method. Each target region was amplified by PCR and genotyped by Sanger sequencing. Within all our knockout experiments, 82.1% of transgenic plants have a knockout allele and 79.5% have double knockout alleles. Quantification of plant height and heading date are recorded in details.

Gene locus	Phenotypes	gRNA [a]	gRNA site	Double knockout lines /total	Genotypes of target regions in the two haplotypes of a mutant detected by Sanger sequencing	Phenotype studied [c]		
						Quantified phenotype	WT measurement	Mutant measurement
Os01g0884200	Dwarf, sterile	GTCGGTTCAATGCGCACGTG	chr01:38398664-38398683	2/2	H1-1:DEL(38398658-38398683) H1-2:IN(38398680-A) [b]	Plant height	129.3cm	110.7cm
					H2-1:DEL(38398658-38398683) H2-2:IN(38398680-A)			
Os01g0884400Os01g0884450	Late heading, sterile	GATTGCTCCGCAGTTTGCGA	chr01:38412019-38412038	3/4	H1-1:IN(38412035-A) H1-2:DEL(38412033-38412035)	Heading date	110d	143d
					H2-1:DEL(38412035) H2-2:IN(38412035-T)			
					H3-1:IN(38412035-A) H3-2:IN(38412035-T)			
					H4-1:WT H4-2:WT	No phenotypic change	-	-

Os01g884600	No phenotypic change	GTAAGTATAAATTCATCAAG	chr01:38422244-38422263	0/3	H2-1:WT H2-2:WT	No phenotypic change		
					H3-1:WT H3-2:WT			
					H4-1:WT H4-2:WT			
Os01g0885000	Small, growth retard, less tillers	CATGCCGTCGACAAGGCCGC	chr01:38443777-38443796	2/2	H1-1:DEL(38443794) H1-2:DEL(38443792-38443793)	Refer to Fig. S8 due to inability to quantitatively describe	-	-
					H2-1:DEL(38443793-38443794) H2-2:DEL(38443791)			
Os01g0886000	Late heading, less tillers, sterile	TGACGCGCTGACCGCCAGCT	chr01:38493921-38493940	4/4	H1-1:DEL(38493937) H1-2:DEL(38493932-38493941)	Heading date	110d	122d
					H2-1:DEL(38493937) H2-2:DEL(38493933-38493937)			
					H3-1:DEL(38493937) H3-2:DEL(38493933-38493937)			
					H4-1:DEL(38493937) H4-2:DEL(38493932-38493941)			
Os01g0925600 Os01g0925700	Leaf rolling, shorter panicle	TATCGGAAGCTGGGGGATCC	chr01:40584644-40584663	2/4	H3-1:DEL(40584659-40584662) H3-2:DEL(40584659-40584662)	Refer to Fig. S8	-	-
					H4-1:IN(40584659-T) H4-1:IN(40584659-T)			

					H8-1:SNP(40584656-C-T;40584658-C-T) H8-2:SNP(40584656-C-T;40584658-C-T)			
					H6-1:WT H6-2:WT	No phenotypic change	-	-
Os01g0930800	Late heading, sterile	ATCCGAAGACGAAATCCTCC	chr01:40842388-40842407	4/4	H1-1:IN(40842402-G) H1-2:DEL(40842403-40842404)	Heading date	110d	125d
					H2-1:DEL(40842404-40842405) H2-2:DEL(40842405)			
					H3-1:IN(40842402-G) H3-2:DEL(40842403-40842404)			
					H4-1:IN(40842402-G) H4-2:DEL(40842403-40842404)			
Os01g0930900	No phenotypic change	CGGCCGCGCCGAACAGCGTC	chr01:40849696-40849715	5/7	H1-1:DEL(40849711-40849712) H1-2:DEL(40849710-40849712)	No phenotypic change	-	-
					H3-1::IN(40849711-G) H3-2::IN(40849716-CC)			
					H5-1:DEL(40849711-40849712) H5-2:DEL(40849710-40849712)			
					H6-1:DEL(40849711-40849712) H6-2:DEL(40849710-40849712)			

					H7-1:DEL(40849711-40849712) H7-2:DEL(40849710-40849712)			
					H2-1 : WT H2-2::IN(40849711-G)	No phenotypic change	-	-
					H4-1 : WT H4-2::IN(40849711-G)			
Os10g0555600 Os10g0555651	Dwarf	GATGGAGCGGTAGTTGGAGT	chr10:21830493-21830512	2/2	H1-1:DEL(21830508-21830509) H1-2:DEL(21830508-21830509)			
					H2-1:DEL(21830508-21830509) H2-2:DEL(21830508-21830509)			
Os10g0555900Os10g0556000	Dwarf, late heading	AGAACGGTAGACGGCCATGG	chr10:21848246-21848265	8/12	H1-1:DEL(21848261-21848262) H1-2:IN(21848262-G)	Plant height [d]	132.4cm	64.6cm
					H2-1:DEL(21848262-21848263) H2-2:DEL(21848262-21848263)			
					H5-1:DEL(21848261-21848262) H5-2:IN(21848262-G)			
					H6-1:DEL(21848263) H6-2:DEL(21848263)			
					H10-1:DEL(21848261-21848262) H10-2:IN(21848262-G)	Heading date [d]	81d	90d
					H11-1:DEL(21848261-218			

					48262) H11-2:IN(21848262-G)			
					H12-1:DEL(21848261-21848262) H12-2:IN(21848262-G)			
					H13-1:DEL(21848262) H13-2:IN(21848262-A)			
					H4-1:WT H4-2:WT	No phenotypic change	-	-
					H7-1:WT H7-2:WT			
					H8-1:WT H8-2:WT			
					H9-1:WT H9-2:WT			
Os10g0556200	Dwarf	GGGTCGTCCAGATCCCAGCT	chr10:21867486-21867505	1/1	H7-1:DEL(21867498-21867505) H7-2:DEL(21867498-21867505)	Plant height	122.1cm	73.7cm
Os10g0556900	No phenotypic change	GGAAAAGGTTTCAGGGCCAA	chr10:21909319-21909338	2/4	H2-1:IN(21909335-A) H2-2:IN(21909335-C)	No phenotypic change	-	-
					H4-1:IN(21909335-A) H4-2:IN(21909335-C)			
					H1-1:WT H1-2:WT	No phenotypic change	-	-
					H3-1:WT H3-2:WT			
Os10g0555100	Dwarf, spike shape change，	CGACGCGCTCATAGTCAACC	chr10:21800366-21800385	1/1	H2-1:DEL(21800381) H2-2:DEL(21800381)	Plant height	130.1cm	99.6cm

Os10g0555200	Dwarf, sterile	GCGTTGGCTCTGACACTGCG	chr10:21804189-21804208	3/4	H3-1:DEL(21804201-21804204) H3-2:DEL(21804202-21804205)	Plant height	130.3cm	99.7cm
					H4-1:DEL(21804201-21804204) H4-2:DEL(21804202-21804205)			
					H6-1:DEL(21804201-21804204) H6-2:DEL(21804202-21804205)			
					H8-1:WT H8-2:WT	No phenotypic change	-	-
Os10g0555300	Dwarf, sterile	CGCCGAGCGCCTGGTACGAC	chr10:21809888-21809907	3/4	H5-1:DEL(21809904) H5-2:DEL(21809904)	Plant height	129.2cm	105.2cm
					H6-1:IN(21809904-G) H6-2:IN(21809904-G)			
					H7-1:IN(21809904-G) H7-2:IN(21809904-G)			
					H4-1:WT H4-2:WT	No phenotypic change	-	-
Os10g0555700	Sterile	TCCGTGATGATCACCGTCCT	chr10:21836806-21836825	1/1	H2-1:DEL(21836817-21836821) H2-2:DEL(21836817-21836821)	Sterility	Normal	Sterile
Os10g0556100	Small, growth retard, leaf rolling	GATTTTGTGAGCCACATTGC	chr10:21859441-21859460	1/1	H1-1:IN(21859458-T) H1-2:DEL(21859457)	Plant height	89.4cm	60.5cm
Os10g055	Leaf	GAGCTGTA	chr10:219950	12/12	H1-1:DEL(21995082-2199	Plant height	130.2cm	117.2cm

8850	rolling, dwarf, weak	CACCCACGCCAA	68-21995087		5086) H1-2:IN(21995084-C)			
					H2-1:DEL(21995082-21995086) H2-2:IN(21995084-C) H3-1:DEL(21995082-21995086) H3-2:IN(21995084-T) H4-1:DEL(21995082-21995086) H4-2:IN(21995084-T)			
					H5-1:DEL(21995082-21995086) H5-2:IN(21995084-C)			
					H6-1:DEL(21995082-21995086) H6-2:IN(21995084-C)			
					H7-1:DEL(21995082-21995086) H7-2:IN(21995084-T)			
					H8-1:DEL(21995082-21995086) H8-2:IN(21995084-T)			
					H9-1:DEL(21995082-21995086) H9-2:IN(21995084-T)			
					H10-1:DEL(21995082-21995086) H10-2:IN(21995084-T)			
					H11-1:DEL(21995082-21995086) H11-2:IN(21995084-T)			

					H12-1:DEL(21995082-21995086) H12-2:IN(21995084-T)			
Os10g0559800 Os10g0559833	No phenotypic change	AAAGAACTACTTCCTGAATA	chr10:22041457-22041476	3/3	H1-1:DEL(22041472-22041481) H1-2:DEL(22041471-22041472)	No phenotypic change	-	-
					H2-1:DEL(22041472-22041481) H2-2:DEL(22041471-22041472)			
					H3-1:DEL(22041472-22041481) H3-2:DEL(22041471-22041472)			
Os11g0242400	No phenotypic change	CGGCACCACCAGAGGGTCGT	chr11:7642767-7642786	2/2	H1-1:IN(7642784-T) H1-2:IN(7642784-T)	No phenotypic change	-	-
					H2-1:IN(7642784-T) H2-2:IN(7642784-T)			

[a] gRNAs were designed to bind the 5' region of the coding sequence of the target gene locus.

[b] H1-1 and H1-2 refer to the two haplotypes of the H1 mutant. Genotype of each haplotype was annotated by comparison with the wild type. WT: wild type. DEL means that this haplotype has a deletion on this region relative to the wild type. IN refers to an insertion at the downstream of the following site and SNP means that there was a single nucleotide variation at the site.

[c] Plant height was in centimeters and heading date was in days.

[d] All of the 8 double knockout mutants showed both changes in plant height and heading date.

Table S16. Expression levels in callus of the 26 genes chosen for the knockdown experiment.

Expression level [a]	gene loci with successful knockdown		Gene loci with failed knockdown	Total
	Phenotypic change	No phenotypic change		
Low	1	1	1	3
Medium	3	2	7	12
High	6	3	2	11
Total	10	6	10	26

Among the 38 gene loci for which we could not get knockout mutants even after at least two times of transformation, 26 were chosen randomly to construct dCas9 knockdown plasmids to reduce their expression level. Similar to the knockout results, (i) there was still a high percentage (38.5%) of genes for which we could not get knockdown mutants; (ii) there was a high percentage (62.5%) of knockdown genes that exhibited distinctly phenotypic changes and many of them died during seedling. These results are consistent with the possibility that most of these genes are essential.

[a] The expression data were retrieved from Genevestigator, a database of microarray expression data. For those loci that were absent in Genevestigator, we used the FPKM values downloaded from 9311 callus RNA-seq data (NCBI, BioProject PRJNA117345, SRR037711~SRR037724). In Genevestigator, RMA (Robust Multi-array Average), is used to summarize probe-level measurements of all arrays to give an estimate expression level, the so-called “average expression value”. The expression levels of genes in all samples are divided to three groups: "LOW", "MEDIUM" or "HIGH". "LOW"(<8.1) corresponds to the first quartile, "MEDIUM"(8.1~11.8) to the interquartile range and "HIGH" (>11.8) to the fourth quartile. The FPKM values were divided to three groups: low (<1), medium (1~10), and high (>10)

Table S17. Estimation of crossovers in each cultivar through exchange of blocks larger than 400kb.

Cultivar	Chromosome												All
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	
IR8	7	13	11	9	6	7	4	2	2	10	4	7	82
IR24	2	15	9	11	9	6	12	4	8	4	6	9	95
IR30	10	9	9	6	7	16	0	9	10	11	6	8	101
MH63	3	10	8	4	10	3	6	3	5	6	9	8	75
Total	22	47	37	30	32	32	22	18	25	31	25	32	353

Dataset S1. (separate file)
Full list of identified SNP markers for the IR30-Gui630- MH63 trio.

Dataset S2. (separate file)
Identified markers in DGWG-Peta-IR8 inferred or corrected using pedigree information.