Fine-tuning quantitative agronomic traits by manipulating gene copy number in rice

Chihiro Nomura[1], Hiroyuki Kanzaki[1], Eiko Kanzaki[1], Motoki Shimizu[1], Kaori Oikawa[1], Hiroe Utsushi[1], Kazue Ito[1], Yusaku Sugimura[1], Ryohei Terauchi[1,2], Akira Abe[1*]

[1] Iwate Biotechnology Research Center, Kitakami, Iwate, Japan

[2] Crop Evolution Laboratory, Kyoto University, Muko, Kyoto, Japan

* Corresponding author

E-mail: a-abe@ibrc.or.jp

Abstract

Although recent plant pan-genomes have revealed extensive copy number variations, studies addressing their phenotypic consequences are limited. Here, we manipulated gene copy number for *OsMADS18* through gene editing and evaluated the effects. The presence of one to three *OsMADS18* tandem copies was reflected in stepwise higher transcript levels and concomitant trait values for leaf blade and culm length, demonstrating that manipulating gene copy number can fine-tune important agronomic traits.

Main

Since the advent of high-throughput sequencing, numerous single-nucleotide polymorphisms (SNPs) and short insertion/deletion polymorphisms (indels) have been identified that affect phenotypes in many organisms, from humans[1] to crop plants[2]. Recent

24 developments in long-read sequencing technologies such as the PacBio and Oxford
25 Nanopore Technologies platforms have helped reveal copy number variations (CNVs)
26 and large structural variations in genomes, which were previously difficult to capture
27 using short-read sequencing methods, such as those involving the Illumina platform[3–5].
28 However, current *in silico* analyses using genome sequences obtained by long-read
29 sequences across plants to understand the relationships between CNV and phenotype are
30 based on materials with diverse genetic backgrounds, making it difficult to demonstrate
31 a direct dosage effect of CNV[6]. There is therefore a pressing need to study how variation
32 in gene dosage caused by CNVs results in different transcript levels and ultimately
33 influences plant phenotypes[7].
34
35 Genome editing is a powerful tool to introduce mutations at target genomic
36 locations[8]. Among gene editing methods, the clustered regularly interspaced short
37 palindromic repeats (CRISPR)/CRISPR-associated nuclease 9 (Cas9) system is the most
38 widely used for genetic research[9–11]. CRISPR/Cas9 typically cuts the genome near a target
39 site bound by a single guide RNA (sgRNA), after which the repair machinery introduces
40 mutations such as indels. CRISPR/Cas9-mediated gene editing can knock out genes to
41 explore gene function[12,13]. However, a complete loss-of-function mutant often displays
42 drastic changes in phenotypes compared to the wild type[14–17] and thus is not suitable for
43 studying the quantitative phenotypic effects associated with subtle changes in gene
44 expression. In addition to knocking out gene function, CRISPR/Cas9-mediated genome
45 editing can also delete larger genomic fragments by using two sgRNAs targeting two sites
46 flanking the genomic sequence to be deleted[11]. We hypothesized that this approach could
47 be used to increase or decrease the copy number of target genes, thereby revealing the

48 quantitative effects of gene CNV on expression levels and phenotypes.

49

50 MADS-box family proteins are transcription factors widely distributed across
51 eukaryotes. Among them, OsMADS18 from rice (*Oryza sativa*) is involved in the
52 transition from the vegetative to the reproductive stages[18,19] and exerts pleiotropic effects
53 on germination and tillering[20]. A previous study of a pan-genome from 33 genetically
54 diverse rice accessions reported CNVs for *OsMADS18* via tandem duplication[5]. Here, we
55 performed *de novo* genome assemblies for the elite *japonica* rice cultivar 'Hitomebore'
56 and the *japonica* cultivar 'Sasanishiki' using long reads from Oxford Nanopore
57 Technologies and compared the results to the rice reference genome IRGSP-1.0 from the
58 cultivar 'Nipponbare'[21] (Extended Data Fig. 1a,b). We also assessed sequencing coverage
59 with Illumina-based short reads between Hitomebore and Sasanishiki (Extended Data Fig.
60 1c and Supplementary Figs. 1 and 2). From these analyses, we confirmed the presence of
61 a 20-kb tandem duplication at 24.776–24.796 Mb on chromosome 7 in the Hitomebore
62 genome and its absence from the Sasanishiki and IRGSP-1.0 reference genomes.
63 According to the rice annotation project database (RAP-DB)[21,22], the only full-length
64 protein-coding gene contained in the 20-kb fragment was *OsMADS18*, which is
65 duplicated specifically in Hitomebore (Extended Data Fig. 1a–c and Supplementary Figs.
66 1–3). We detected only one SNP and three indels across the two tandem repeats of the 20-
67 kb fragment, with no variation in the two copies of the *OsMADS18* coding region
68 (Extended Data Fig. 1d). Given that *OsMADS18* is the only full-length protein-coding
69 gene within the duplicated region and that sequence variation between the two tandem
70 copies is minimal, we speculated that CNV of *OsMADS18* might offer an attractive model
71 for the genetic evaluation of gene dosage effects. We thus attempted to locally manipulate

72 gene copy number for *OsMADS18* using CRISPR/Cas9 and study how changing gene
73 copy number would affect plant phenotypes.
74
75 To this end, we designed an sgRNA targeting a site 12 kb upstream of the ATG of
76 *OsMADS18*, within the region tandemly duplicated in Hitomebore, for CRISPR/Cas9-
77 mediated editing in this cultivar. Since the two copies of *OsMADS18* have the same
78 sequence, this sgRNA has two target sites in this cultivar (Fig. 1a). From T_0 individuals
79 regenerated from Hitomebore calli, we checked for individuals with altered *OsMADS18*
80 copy number by quantitative PCR (qPCR), using two primer sets (Extended Data Fig.
81 2a): one set amplifying an amplicon outside of the duplicated region and thus having a
82 constant copy number and another with primers designed to anneal on either side of the
83 junction of the duplicated region to specifically amplify the junction if a tandem
84 duplication exists. We calculated the relative amplification ratio of these two primer sets
85 using an efficiency-corrected Pfaffl method[23]. When the relative amplification ratio of
86 Hitomebore (with two homozygous copies of the duplication) was set to 1, deviations of
87 the relative ratio from 1 in 0.5-unit increments or decrements reflected a change in
88 *OsMADS18* copy number (Extended Data Fig. 2a). The qPCR analysis identified nine T_0
89 plants with a relative ratio of ~0.5, out of 107 regenerated T_0 plants obtained from
90 different calli or from the same callus with different culture periods, as well as one T_0
91 plant with a relative ratio of ~1.5, suggesting a loss or gain of one *OsMADS18* copy,
92 respectively (Extended Data Fig. 2b,c). We allowed these plants to self-pollinate and
93 repeated the qPCR analysis on their progeny. From the parents with a relative ratio of
94 ~0.5, we obtained progeny with relative amplification ratios of 0 (reflecting absence of
95 the junction sequence of the duplicated region), ~0.5 or ~1 (Extended Data Fig. 2d); from

96 the parent with a relative ratio of ~1.5, we obtained progeny with relative amplification
97 ratios of ~1, ~1.5 or ~2 (Extended Data Fig. 2e). Furthermore, plant lines with relative
98 amplification ratios of ~0, ~1 or ~2 produced progeny with the same relative amplification
99 ratios, suggesting that these plants are homozygous for one, two or three copies of
100 *OsMADS18*, respectively (Fig. 1b). These results demonstrate that our qPCR-based
101 approach is a practical tool for efficiently detecting of genomic CNVs.
102
103 To confirm these copy numbers, we checked the sequencing coverage depth over the
104 *OsMADS18* genomic region using Illumina short reads from homozygous plants with an
105 estimated one, two or three tandem copies. The relative sequencing depth within the
106 *OsMADS18* region decreased in plants with one copy and increased in plants with three
107 copies, relative to those with two copies (Fig. 1c). A *de novo* genome assembly using
108 Nanopore long reads for the genomes from homozygous individuals with an estimated
109 one or three *OsMADS18* copies in the Hitomebore background showed that the
110 *OsMADS18* copy number was indeed one and three, respectively, as evidenced by a dot
111 matrix analysis (Extended Data Fig. 3).
112
113 To study the effect of CNV at *OsMADS18* on its transcript levels, we performed
114 reverse transcription qPCR (RT-qPCR) on individuals with one to three copies of the gene,
115 normalizing to *UBQ5*[24,25], with the mean relative *OsMADS18* transcript level for one-
116 copy plants set to 1. The relative *OsMADS18* transcript levels gradually rose from 1, on
117 average, to 2 or 3 in plants homozygous for one, two or three copies of *OsMADS18*,
118 respectively, thus neatly correlating with gene copy number (Fig. 1d). These findings
119 suggest that simple genome editing introduced CNV with stepwise changes in

120 *OsMADS18* expression levels.

121

122 We assessed the agronomic traits of lines with CNV (from one to three copies) at
123 *OsMADS18*. We observed no major difference in overall plant architecture among the rice
124 lines (Extended Data Fig. 4a). However, the flag leaf blade was significantly shorter when
125 the *OsMADS18* copy number was higher (Fig. 1e,f), as was the length of the leaf blade
126 just below the flag leaf (Extended Data Fig. 4b). By contrast, culm length was
127 significantly greater with more *OsMADS18* copies (Fig. 1g). To our knowledge, an effect
128 of *OsMADS18* on leaf blade size and culm length has not been previously reported.
129 Previous studies have reported that *OsMADS18* expression is upregulated during the
130 transition from vegetative to reproductive growth and is involved in the number of days
131 to heading[18,19]. However, in this study, we measured no significant difference in the
132 number of days to heading among the three sets of lines with CNV at *OsMADS18*
133 (Extended Data Fig. 4c).

134

135 Most agronomic traits are quantitative in nature, and their variations are controlled
136 by subtle differences in the expression levels of multiple genes. Thus, manipulating gene
137 expression is important to adjust traits. Recently, promoter editing has been developed to
138 generate a large allelic series of plants with different expression levels of a target gene by
139 inserting random mutations in the *cis*-regulatory elements of promoters[26]. However,
140 because this approach targets *cis*-regulatory elements, it is difficult to predict the direction
141 and degree of the changes in expression levels in the edited plants; moreover, the
142 spatiotemporal gene expression pattern may differ from that driven by the intact promoter.
143 Here, we propose the manipulation of gene copy number as a powerful alternative method

144 to create lines with a range of copy numbers leading to stepwise changes in the expression
145 levels of a target gene without randomly perturbing its spatiotemporal expression patterns.
146 This method should allow the fine-tuning of important traits by optimizing gene copy
147 number, thus accelerating plant breeding.
148

149 Methods

150 Plant materials and growth conditions

151 The genomic structure over the target *OsMADS18* region was compared between the rice
152 (*Oryza sativa*) ssp. *japonica* cultivars Hitomebore and Sasanishiki. Genome editing was
153 performed using Hitomebore, which carries two tandem copies of *OsMADS18*. Plants
154 with *OsMADS18* CNVs were selected from 107 regenerated gene-edited independent
155 lines. The selected individuals were allowed to self-pollinate to establish homozygous
156 lines with different *OsMADS18* copy numbers. These lines were grown in a plant
157 incubator when seedlings. After transplanting, they were grown in a temperature-
158 controlled greenhouse under short-day conditions, with day/night temperatures of
159 ~27°C/~23°C, and *OsMADS18* transcript levels and various phenotypes were evaluated.
160

161 Plasmid construction and plant transformation

162 To generate lines with CNV at *OsMADS18*, one sgRNA targeting the duplicated genomic
163 region was designed using CRISPRdirect (https://crispr.dbcls.jp/)[27]. To increase the
164 probability of mutations, a binary vector containing two sgRNA expression cassettes was
165 constructed based on a previous study[28]. The *OsU6a* and *OsU6b* promoters, amplified
166 using pYLsgRNA-OsU6a and pYLsgRNA-OsU6b as templates, were each fused to the
167 target sequence and sgRNA scaffold. The ccdB cassette of the pYLCRISPR/Cas9P_{ubi}-H

168 vector was replaced with the two sgRNA cassettes via the BsaI site using Golden Gate
169 assembly. This vector was introduced into *Agrobacterium tumefaciens* strain EHA105,
170 and rice transformation into Hitomebore was performed according to a previous study[29].
171 The pYLCRISPR/Cas9P_{ubi}-H (RRID:Addgene_66187), pYLsgRNA-OsU6a
172 RRID:Addgene_66194) and pYLsgRNA-OsU6b (RRID:Addgene_66196) vectors were
173 kindly provided by Yao-Guang Liu. The primers used in this study are listed in
174 Supplementary Tables 1.
175

176 ## ***De novo* assembly**

177 To reconstruct the Hitomebore and Sasanishiki genome, *de novo* assembly of each
178 genome was performed using Nanopore long reads and Illumina short reads following a
179 previously published method[30]. To extract high-molecular-weight genomic DNA from
180 rice leaf tissue for Nanopore sequencing, leaf blades were carefully ground using a mortar
181 and pestle while being cooled with liquid nitrogen, and genomic DNA was extracted using
182 a NucleoBond high-molecular-weight DNA kit (Macherey-Nagel, Düren, Nordrhein-
183 Westfalen, Germany). Following DNA extraction, low-molecular-weight DNA was
184 eliminated using a Short Read Eliminator Kit XL (Circulomics, Baltimore, Maryland,
185 USA). For Hitomebore, genomic library preparation was performed using a ligation
186 sequencing kit (SQK-LSK-114; Oxford Nanopore Technologies [ONT], Oxford,
187 Oxfordshire, UK) according to the manufacturer's instructions, and sequencing was
188 performed using PromethION 2 Solo (ONT). Additionally, long reads for Hitomebore
189 produced on a PromethION (ONT) platform, as reported by Sugihara et al.[30], were also
190 used. For Sasanishiki, genomic library preparation was performed using a ligation
191 sequencing kit (SQK-LSK-109; ONT), and sequencing was performed on a MinION

192 instrument.

193 Base-calling of the Nanopore long reads was conducted using Dorado v0.9.0
194 (https://github.com/nanoporetech/dorado). The first 50-bp of each read and reads with an
195 average read quality score below 10 were removed using Chopper v0.9.0[31]. The resulting
196 clean Nanopore long reads were assembled using NECAT v0.0.1[32], with the genome size
197 set to 380 Mb and using reads longer than 3,000 bp. To improve the accuracy of the
198 assembly, Racon v1.4.20[33] was employed for error correction using Nanopore reads,
199 followed by Medaka v1.4.1 (https://github.com/nanoporetech/medaka) for misassembly
200 correction. Two rounds of consensus correction were then performed using BWA
201 v0.7.17[34,35] and HyPo v1.0.3[36] with the Illumina short reads. Redundant contigs were
202 removed using purge-haplotigs v1.1.1[37], resulting in a *de novo* assembly of 376.2 Mb
203 comprising 59 contigs for Hitomebore and 378.7 Mb consisting of 49 contigs for
204 Sasanishiki.

205 *De novo* assembly was also performed for Hitomebore gene-edited lines with one or
206 three copies of the *OsMADS18* region. In both cases, high-molecular-weight genomic
207 DNA was extracted as described above. The library for the one-copy line was prepared
208 using an SQK-LSK110 kit and sequenced on MinION, while the library for the three-
209 copy line was prepared using an SQK-LSK114 kit and sequenced on a PromethION 2
210 Solo. Base-calling was performed using the same method as described above, and *de novo*
211 assembly was conducted using Flye v2.9.5[38].

212

213 Illumina-based DNA sequencing

214 Total genomic DNA was extracted from rice leaf blades using the CTAB-based method
215 with a NucleoSpin Plant II kit (Macherey-Nagel). Sequencing libraries were prepared

216 using a Collibri ES DNA Library Prep Kit for Illumina Systems (Thermo Fisher Scientific,
217 Waltham, Massachusetts, USA). Libraries were sequenced on a MiSeq (Illumina, San
218 Diego, California, USA) instrument in 251-bp paired-end mode for Hitomebore and in
219 301-bp paired-end mode for Sasanishiki and on a NovaSeq X Plus (Illumina) instrument
220 in 151-bp paired-end mode for lines with CNV (from one to three copies) at *OsMADS18*.
221 The resulting raw reads from the MiSeq instrument were filtered using Trimmomatic
222 v0.36[39] with the paired-end option. Adapter sequences were removed with the
223 ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10 option. Bases with quality scores of <20 at
224 the beginning and end of reads were removed (LEADING:20, TRAILING:20). A sliding
225 window of four bases was used to remove reads with average quality scores of <20
226 (SLIDINGWINDOW:4:20). Reads were removed if the read length after trimming was <
227 36 (MINLEN:36) and cut to 200 bases if the read length was > 200 bases (CROP:200).
228 The resulting raw reads from the NovaSeq X Plus instrument were filtered using fastp
229 v0.22.0[40] with default settings supplemented with the --detect_adapter_for_pe option to
230 automatically detect and trim adapter sequences and --trim_poly_g option to trim polyG
231 tails. The processed reads were aligned to the rice reference genome IRGSP-1.0 using
232 BWA v0.7.18. The obtained SAM files were converted to BAM format, sorted, and
233 indexed using SAMtools v1.9[41].
234

235 **Sequencing depth analysis**

236 The sequencing depth was calculated using mosdepth v0.3.3[42] with the BAM files
237 generated from high-throughput sequencing and subsequent processing. The sliding
238 window used in the analysis was created with BEDTools v2.31.1[43] using the command
239 'makewindows'. Relative sequencing depth between target and control was calculated

using the following equation:

$$D_{\mathrm{t}} = \left(\frac{d_{\mathrm{t}}}{\overline{D_{\mathrm{t}}}}\right) \times \left(\frac{\overline{D_{\mathrm{c}}}}{d_{c}}\right) \quad (1)$$

where D_{t} is the relative sequencing depth in the region of the target genome, d_{t} is the sequencing depth in the region of the target genome, $\overline{D_{\mathrm{t}}}$ is the average sequencing depth of the whole target genome, d_{c} is the sequencing depth in the target region of the control genome and $\overline{D_{\mathrm{c}}}$ is the average sequencing depth of the whole control genome.

Comparison of genome structures

The genomic structure over the target *OsMADS18* region was compared using MUMmer4 v4.0.1[44]. First, the command 'blastn' from BLAST+ v2.13.0[45] was used to identify the contig containing *OsMADS18* from each of the four *de novo* assemblies. The identified contigs were then extracted individually using SeqKit v2.4.0[46]. Each contig was compared to chromosome 7 of the IRGSP-1.0 reference genome using the command 'nucmer' of MUMmer4. The results were filtered with the command 'delta-filter', setting the minimum alignment identity to 95% and the minimum alignment length to 5,000 bp. Finally, the filtered results were visualized using the command 'mummerplot'.

To identify SNPs and indels within the duplicated regions of the Hitomebore genome, the duplicated sequences were separated into two segments corresponding to the upstream and downstream copies. These sequences and the corresponding region from the IRGSP-1.0 reference genome were aligned using MAFFT v7[47]. Sequence differences were then identified based on the alignment results. Visualization of the alignment of the tandem duplication was performed using the Integrative Genomics Viewer (IGV) v2.16.2[48], with the IRGSP-1.0 reference genome.

qPCR-based analysis of target genomic copy number

qPCR was conducted using a QuantStudio 3 Real-Time PCR System (Thermo Fisher Scientific) and Luna Universal qPCR Master Mix (New England Biolabs, Ipswich, Massachusetts, USA). To estimate CNV of *OsMADS18*, the relative amplification ratio was calculated using qPCR data from primer sets designed to span the duplicated junction and a non-duplicated control region. The analysis was based on the Pfaffl approach[23], which incorporates primer-specific amplification efficiencies. The mean C_{t} value of wild-type Hitomebore plants (homozygous for two copies of *OsMADS18*, $n = 3$) was used as the calibrator. The C_{t} difference (ΔC_{t}) for each primer set was calculated using Equations 2 and 3:

$$\Delta C_{\mathrm{t,junction}} = C_{\mathrm{t,junction,calibrator}} - C_{\mathrm{t,junction,sample}} \tag{2}$$

$$\Delta C_{\mathrm{t,control}} = C_{\mathrm{t,control,calibrator}} - C_{\mathrm{t,control,sample}} \tag{3}$$

The relative amplification ratio (R) was then calculated using Equation 4:

$$R = \frac{\left(E_{\mathrm{junction}}\right)^{\Delta C_{\mathrm{t,junction}}}}{\left(E_{\mathrm{control}}\right)^{\Delta C_{\mathrm{t,control}}}} \tag{4}$$

where E_{junction} and E_{control} represent the amplification efficiencies of the junction and control primer sets, respectively.

Further details are provided in Extended Data Fig. 2a, and the primer sequences used in the experiments are listed in Supplementary Table 1.

RT-qPCR analysis

Leaf blade samples were bulk collected from the uppermost fully expanded leaf of the main stem, with three individuals per sample, and immediately frozen in liquid nitrogen. After each sample was ground to powder, total RNA was extracted using a NucleoSpin

287 RNA Plant kit (Macherey-Nagel). Total RNA was then reverse transcribed into first-
288 strand cDNA using PrimeScript RT Master Mix (Takara Bio, Kusatsu, Shiga, Japan).
289 qPCR was performed using a QuantStudio 3 Real-Time PCR System with Luna Universal
290 qPCR Master Mix. The ΔΔCt method was used to quantify the transcript levels, and the
291 relative transcript levels of OsMADS18 were calculated using *UBQ5*[24,25] as the internal
292 control. The mean C_t value of biological replicates from plants with one homozygous
293 copy of the *OsMADS18* was used as the calibrator. The primers used in the experiment
294 are listed in Supplementary Table 1.
295

296 **Phenotyping**

297 After heading, fully expanded leaf blades were collected, scanned and analysed using a
298 custom-developed automated measurement program in Python to calculate their
299 morphology, including length. Leaf blade length was compared among individuals with
300 the same total number of leaves on the main stem. After heading was fully completed, the
301 length from the ground level to the panicle neck node was measured and defined as culm
302 length. The number of days from sowing to heading was recorded as the heading date.
303 All traits were measured on the main stem.
304

305 **Statistical analysis**

306 Trait comparisons were made between lines with different *OsMADS18* copy numbers,
307 and statistical significance was determined using the Tukey-Kramer test in R v.4.4.2
308 (https://www.r-project.org/). Details of the test are provided in the figure legends.
309

310 **Data availability**

The raw sequencing reads generated in this study have been deposited in the DNA Data Bank of Japan (DDBJ) under BioProject ID PRJDB 13864 and PRJDB20676. A detailed list of accession numbers for all samples is provided in Supplementary Table 2. Nanopore-based *de novo* genome assemblies for four samples are available at Zenodo (DOI: https://doi.org/10.5281/zenodo.15335260). The IRGSP-1.0 reference genome used in this study was obtained from RAP-DB (https://rapdb.dna.affrc.go.jp/download/irgsp1.html). Source data of this study are provided with this paper.

References

1. Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K. & Mardis, E. R. The next-generation sequencing revolution and its impact on genomics. *Cell* **155**, 27–38 (2013).
2. Singh, R., Kumar, K., Bharadwaj, C. & Verma, P. K. Broadening the horizon of crop research: a decade of advancements in plant molecular genetics to divulge phenotype governing genes. *Planta* **255**, 46 (2022).
3. Alonge, M. *et al.* Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell* **182**, 145–161.e23 (2020).
4. Liu, Y. *et al.* Pan-genome of wild and cultivated soybeans. *Cell* **182**, 162–176.e13 (2020).
5. Qin, P. *et al.* Pan-genome analysis of 33 genetically diverse rice accessions reveals

331 hidden genomic variations. *Cell* **184**, 3542–3558.e16 (2021).

332 6. Yuan, Y., Bayer, P. E., Batley, J. & Edwards, D. Current status of structural variation

333 studies in plants. *Plant Biotechnol. J.* **19**, 2153–2163 (2021).

334 7. Maino, E. *et al.* Targeted genome editing *in vivo* corrects a Dmd duplication restoring

335 wild-type dystrophin expression. *EMBO Mol. Med.* **13**, e13228 (2021).

336 8. Gaj, T., Gersbach, C. A. & Barbas, C. F., 3rd. ZFN, TALEN, and CRISPR/Cas-based

337 methods for genome engineering. *Trends Biotechnol.* **31**, 397–405 (2013).

338 9. Jinek, M. *et al.* A programmable dual-RNA–guided DNA endonuclease in adaptive

339 bacterial immunity. *Science* **337**, 816–821 (2012).

340 10. Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with

341 CRISPR-Cas9. *Science* **346**, 1258096 (2014).

342 11. Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Patron, N. J. & Nekrasov, V. Editing

343 plant genomes with CRISPR/Cas9. *Curr. Opin. Biotechnol.* **32**, 76–84 (2015).

344 12. Shalem, O., Sanjana, N. E. & Zhang, F. High-throughput functional genomics using

345 CRISPR-Cas9. *Nat. Rev. Genet.* **16**, 299–311 (2015).

346 13. Bao, A. *et al.* The CRISPR/Cas9 system and its applications in crop genome editing.

347 *Crit. Rev. Biotechnol.* **39**, 321–336 (2019).

348 14. Pnueli, L. *et al.* The *SELF-PRUNING* gene of tomato regulates vegetative to

349 reproductive switching of sympodial meristems and is the ortholog of *CEN* and
350 *TFL1*. *Development* **125**, 1979–1989 (1998).
351 15. Jeon, J. S. *et al. leafy hull sterile1* is a homeotic mutation in a rice MADS box gene
352 affecting rice flower development. *Plant Cell* **12**, 871–884 (2000).
353 16. Komatsu, M., Chujo, A., Nagato, Y., Shimamoto, K. & Kyozuka, J. *FRIZZY*
354 *PANICLE* is required to prevent the formation of axillary meristems and to establish
355 floral meristem identity in rice spikelets. *Development* **130**, 3841–3850 (2003).
356 17. Qi, J. *et al.* Mutation of the rice *Narrow leaf1* gene, which encodes a novel protein,
357 affects vein patterning and polar auxin transport. *Plant Physiol.* **147**, 1947–1959
358 (2008).
359 18. Fornara, F. *et al.* Functional characterization of *OsMADS18*, a member of the
360 *AP1*/*SQUA* subfamily of MADS box genes. *Plant Physiol.* **135**, 2207–2219 (2004).
361 19. Kobayashi, K. *et al.* Inflorescence meristem identity in rice is specified by
362 overlapping functions of three *AP1*/*FUL*-like MADS box genes and *PAP2*, a
363 *SEPALLATA* MADS box gene. *Plant Cell* **24**, 1848–1859 (2012).
364 20. Yin, X. *et al.* OsMADS18, a membrane-bound MADS-box transcription factor,
365 modulates plant architecture and the abscisic acid response in rice. *J. Exp. Bot.* **70**,
366 3895–3909 (2019).

367 21. Kawahara, Y. *et al.* Improvement of the *Oryza sativa* Nipponbare reference genome
368 using next generation sequence and optical map data. *Rice* **6**, 4 (2013).
369 22. Sakai, H. *et al.* Rice Annotation Project Database (RAP-DB): an integrative and
370 interactive database for rice genomics. *Plant Cell Physiol.* **54**, e6 (2013).
371 23. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-
372 PCR. *Nucleic Acids Res.* **29**, e45 (2001).
373 24. Jain, M., Nijhawan, A., Tyagi, A. K. & Khurana, J. P. Validation of housekeeping
374 genes as internal control for studying gene expression in rice by quantitative real-
375 time PCR. *Biochem. Biophys. Res. Commun.* **345**, 646–651 (2006).
376 25. Jain, N., Vergish, S. & Khurana, J. Validation of house-keeping genes for
377 normalization of gene expression data during diurnal/circadian studies in rice by RT-
378 qPCR. *Sci. Rep.* **8**, 3203 (2018).
379 26. Rodríguez-Leal, D., Lemmon, Z. H., Man, J., Bartlett, M. E. & Lippman, Z. B.
380 Engineering quantitative trait variation for crop improvement by genome editing.
381 *Cell* **171**, 470–480.e8 (2017).
382 27. Naito, Y., Hino, K., Bono, H. & Ui-Tei, K. CRISPRdirect: software for designing
383 CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics* **31**, 1120–
384 1123 (2015).

385 28. Ma, X. *et al.* A robust CRISPR/Cas9 system for convenient, high-efficiency
386 multiplex genome editing in monocot and dicot plants. *Mol. Plant* **8**, 1274–1284
387 (2015).
388 29. Toki, S. *et al.* Early infection of scutellum tissue with *Agrobacterium* allows high-
389 speed transformation of rice. *Plant J.* **47**, 969–976 (2006).
390 30. Sugihara, Y. *et al.* Disentangling the complex gene interaction networks between rice
391 and the blast fungus identifies a new pathogen effector. *PLoS Biol.* **21**, e3001945
392 (2023).
393 31. De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M. & Van Broeckhoven, C.
394 NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**,
395 2666–2669 (2018).
396 32. Chen, Y. *et al.* Efficient assembly of nanopore reads via highly accurate and intact
397 error correction. *Nat. Commun.* **12**, 60 (2021).
398 33. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome
399 assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).
400 34. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler
401 transform. *Bioinformatics* **26**, 589–595 (2010).
402 35. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-

403 MEM. Preprint at https://doi.org/10.48550/arXiv.1303.3997 (2013).

404 36. Kundu, R., Casey, J. & Sung, W.-K. HyPo: Super fast & accurate polisher for long
405 read genome assemblies. Preprint at https://doi.org/10.1101/2019.12.19.882506
406 (2019).

407 37. Roach, M. J., Schmidt, S. A. & Borneman, A. R. Purge Haplotigs: allelic contig
408 reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**, 460
409 (2018).

410 38. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone
411 reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).

412 39. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
413 sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

414 40. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ
415 preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

416 41. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008
417 (2021).

418 42. Pedersen, B. S. & Quinlan, A. Mosdepth: quick coverage calculation for genomes
419 and exomes. *Bioinformatics* **34**, 867–868 (2017).

420 43. Quinlan, A. R. BEDTools: the Swiss-army tool for genome feature analysis. *Curr.*

421 *Protoc. Bioinformatics* **47**, 11.12.1–11.12.34 (2014).

422 44. Marçais, G. *et al.* MUMmer4: A fast and versatile genome alignment system. *PLoS*

423 *Comput. Biol.* **14**, e1005944 (2018).

424 45. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**,

425 421 (2009).

426 46. Shen, W., Sipos, B. & Zhao, L. SeqKit2: A Swiss army knife for sequence and

427 alignment processing. *Imeta* **3**, e191 (2024).

428 47. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version

429 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

430 48. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

431

432 Acknowledgements

433 We appreciate Yao-Guang Liu for providing the pYLCRISPR/Cas9P_{ubi}-H, pYLsgRNA-

434 OsU6a and pYLsgRNA-OsU6b plasmids, and Yumiko Ogasawara and Miyuki Saito for

435 technical support. Part of the analysis was performed using the National Institute of

436 Genetics (NIG) supercomputer at the ROIS National Institute of Genetics. This research

437 was partially supported by the following grants: the research program on development of

438 innovative technology (JPJ007097) from the Project of the Bio-oriented Technology

439 Research Advancement Institution (BRAIN); and JSPS KAKENHI (JP22K20584 to C.N.

440 and JP23K26882 to A.A.).

Author information

Authors and Affiliations

Iwate Biotechnology Research Center, Kitakami, Iwate, Japan

Chihiro Nomura, Hiroyuki Kanzaki, Eiko Kanzaki, Motoki Shimizu, Kaori Oikawa, Hiroe Utsushi, Kazue Ito, Yusaku Sugimura, Ryohei Terauchi & Akira Abe

Crop Evolution Laboratory, Kyoto University, Muko, Kyoto, Japan

Ryohei Terauchi

Contributions

C.N. and A.A. conceived and designed the research. C.N., H.K., E.K., M.S., K.O., H.U., K.I., Y.S., and A.A. performed the experiments. C.N. and A.A. conducted bioinformatics analyses. C.N., R.T., and A.A. wrote and revised the manuscript. All authors read and approved the final version of the manuscript.

Corresponding author

Correspondence to Akira Abe.

Ethics declarations

Competing interests

The authors declare no competing interests.

Figure legends

464 **Fig. 1 | Phenotypic variation associated with manipulation of *OsMADS18* copy**
465 **number.**
466 **a**, Diagram of *OsMADS18* CNV. The sgRNA targets two identical sequences within the
467 tandem-duplicated *OsMADS18* genomic region. CRISPR/Cas9-mediated editing can
468 result in plants with either one or three copies of *OsMADS18*. Target sites of the sgRNA
469 are indicated with scissor icons. **b**, Relative amplification ratio of the tandem-duplicated
470 *OsMADS18* region, as determined by qPCR (n = 18 for each CNV). **c**, Relative
471 sequencing depth of Illumina short reads around *OsMADS18*. A sliding window with a
472 window size of 1,000 bp and a step size of 500 bp was used. Thin lines represent
473 individual plants, and the thick line indicates the mean depth for plants with each
474 *OsMADS18* CNV (n = 5). The mean value of plants with the one homozygous copy of
475 the *OsMADS18* CNV is used as the control. The orange shaded area marks the position
476 of *OsMADS18*. **d**, Relative *OsMADS18* transcript levels in leaf blades, as determined by
477 RT-qPCR, normalized to *UBQ5*, with the mean transcript level of one-copy plants set to
478 1. Leaf blades were bulk collected from three individuals at 58 days after sowing (n = 10;
479 total of 30 plants per *OsMAD18* CNV). **e**, Representative photograph of flag leaves. **f**,**g**,
480 Lengths of the flag leaf (n = 29, 34, and 29 for 1, 2, and 3 CNV, respectively) (**f**) and culm
481 (n = 30, 35, and 36 for 1, 2, and 3 CNV, respectively) (**g**). "1 CNV," "2 CNV" and "3
482 CNV" refer to plants with one, two and three homozygous copies of the *OsMADS18* CNV,
483 respectively, for (**b**)–(**g**). Different lowercase letters indicate significant differences (P <
484 0.05) by Tukey-Kramer test for (**d**), (**f**) and (**g**). All violin plots were generated using
485 kernel density estimation with a smoothing adjustment factor of 1.5.
486

487 # Extended data

488 **Extended Data Fig. 1 | Comparison of the genomic regions flanking the**
489 ***OsMADS18* locus on chromosome 7 in the Hitomebore and Sasanishiki cultivars.**
490 **a**,**b**, Dot matrix plots of the *OsMADS18* genomic region between Hitomebore and IRGSP-
491 1.0 (**a**) and between Sasanishiki and IRGSP-1.0 (**b**), using one contig from each cultivar.
492 Grid lines are spaced every 0.03 Mb. **c**, Relative sequencing depth of Hitomebore as target
493 to Sasanishiki as control using Illumina short reads around *OsMADS18*. A sliding window
494 with a window size of 5,000 bp and a step size of 500 bp was used. The orange shaded
495 area marks the position of *OsMADS18* for (**a**)–(**c**). **d**, One SNP and three indels were
496 detected between the *OsMADS18* tandem duplications in Hitomebore, representing
497 sequence differences observed in duplicated region 2 when aligned with duplicated region
498 1.
499
500 **Extended Data Fig. 2 | Determination of *OsMADS18* copy number by qPCR.**
501 **a,** Diagram of the method used to determine *OsMADS18* copy number by qPCR. When
502 the relative amplification ratio of Hitomebore (homozygous for two *OsMADS18* copies)
503 is set to 1, the values for different genotypes are as follows: homozygous for one copy, 0;
504 heterozygous for one and two copies, 0.5; heterozygous for two and three copies, 1.5;
505 homozygous for three copies, 2. **b**, Frequency distribution of the relative amplification
506 ratio for copy number determination in transgenic plants (T_0), as determined by qPCR. **c**,
507 Summary table showing the numbers and percentages of T_0 plants with estimated copy
508 number alterations. T_0 plants derived from different calli or from the same callus with
509 different culture periods were counted as independent lines. The relative amplification
510 ratio was used to classify heterozygous plants as follows: heterozygous for one copy/two
511 copies, $0.3 < x < 0.7$; heterozygous for two copies/three copies, $1.3 < x < 1.7$; x represents

512 the relative amplification ratio. **d**, Frequency distribution of progeny (n = 400) derived
513 from estimated heterozygous for one copy/two copies (relative amplification ratio is ~0.5).
514 **e**, Frequency distribution of progeny (n = 12) derived from estimated heterozygous for
515 two copies/three copies (relative amplification ratio is ~1.5).
516
517 **Extended Data Fig. 3 | Genomic structure of the *OsMADS18* locus in IRGSP-1.0**
518 **and CNV-edited lines.**
519 **a**,**b**, Dot matrix plots comparing the *OsMADS18* genomic region in the Nipponbare
520 reference genome (IRGSP-1.0) to that in plants with one homozygous copy (1 CNV) (**a**)
521 or three homozygous copies (3 CNV) (**b**) of *OsMADS18*, using a contig containing the
522 *OsMADS18* from each edited line. Grid lines are spaced at 0.03-Mb intervals. The orange
523 shaded region indicates the position of *OsMADS18*.
524
525 **Extended Data Fig. 4 | Effects of *OsMADS18* CNV on other agronomic traits.**
526 **a**, Representative photograph of overall plant architecture with different *OsMADS18* copy
527 number in the Hitomebore background. **b**, Leaf blade length of the leaf just below the
528 flag leaf (n = 29, 35, and 29 for 1, 2, and 3 CNV, respectively). **c**, Days to heading (n =
529 30, 35, and 36 for 1, 2, and 3 CNV, respectively). Different lowercase letters indicate
530 significant differences ($P < 0.05$) based on Tukey-Kramer test. "1 CNV," "2 CNV" and
531 "3 CNV" refer to plants with one, two and three homozygous copies of the *OsMADS18*
532 CNV, respectively. All violin plots were generated using kernel density estimation with a
533 smoothing adjustment factor of 1.5.
534
535 # Supplementary information

Supplementary Information

Supplementary Figs. 1–3, and Figs. 1 and 2.

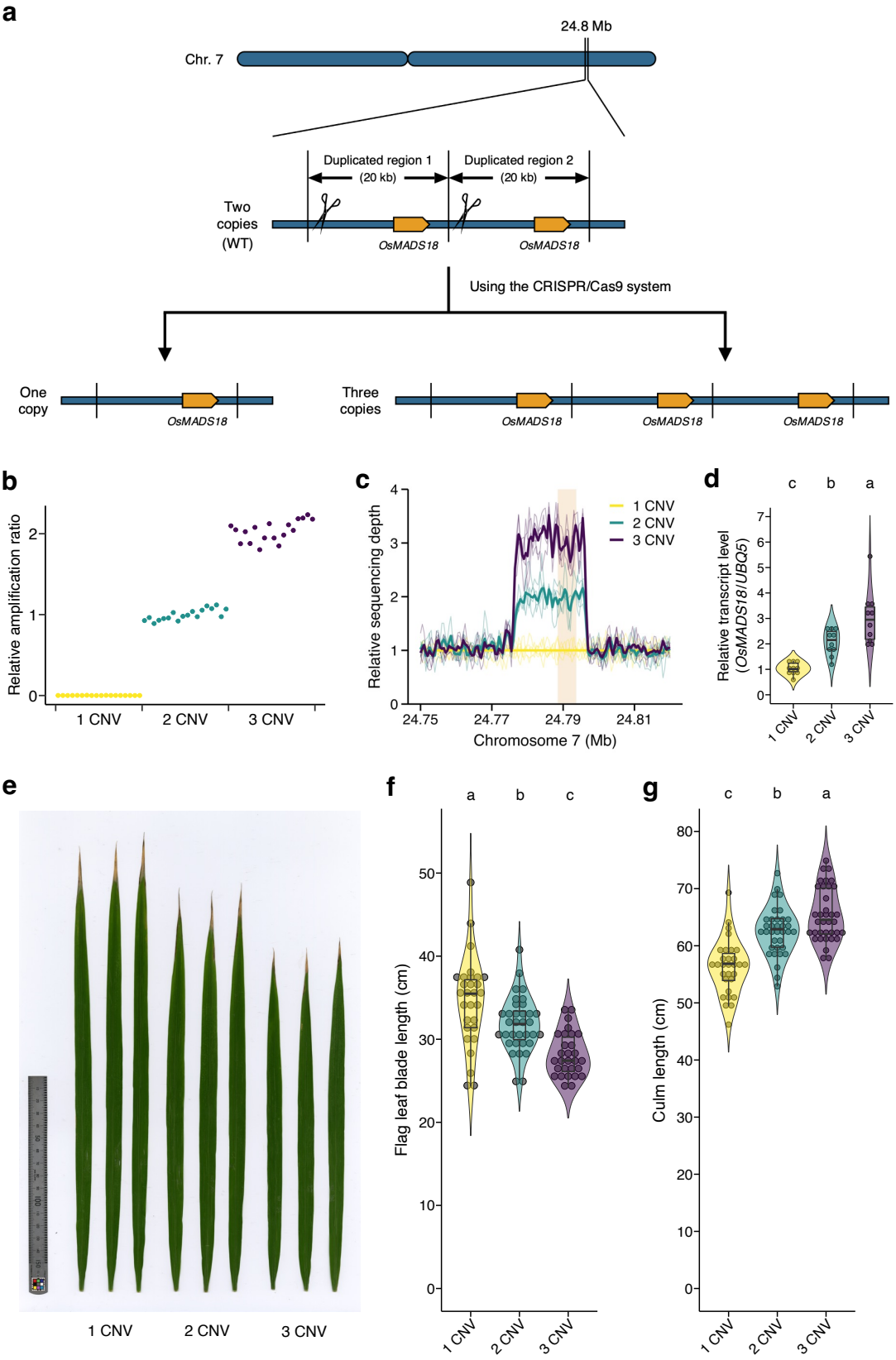
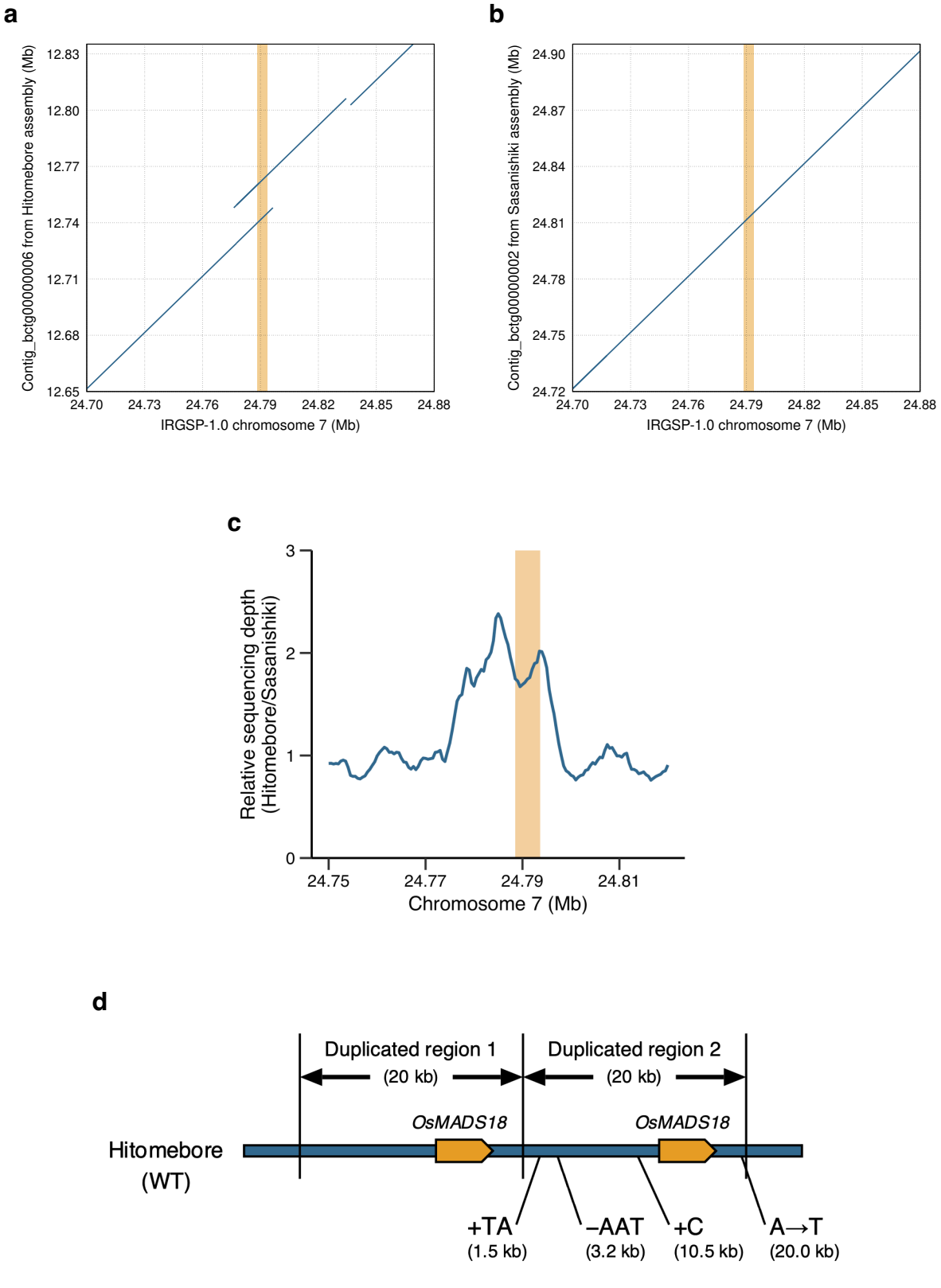
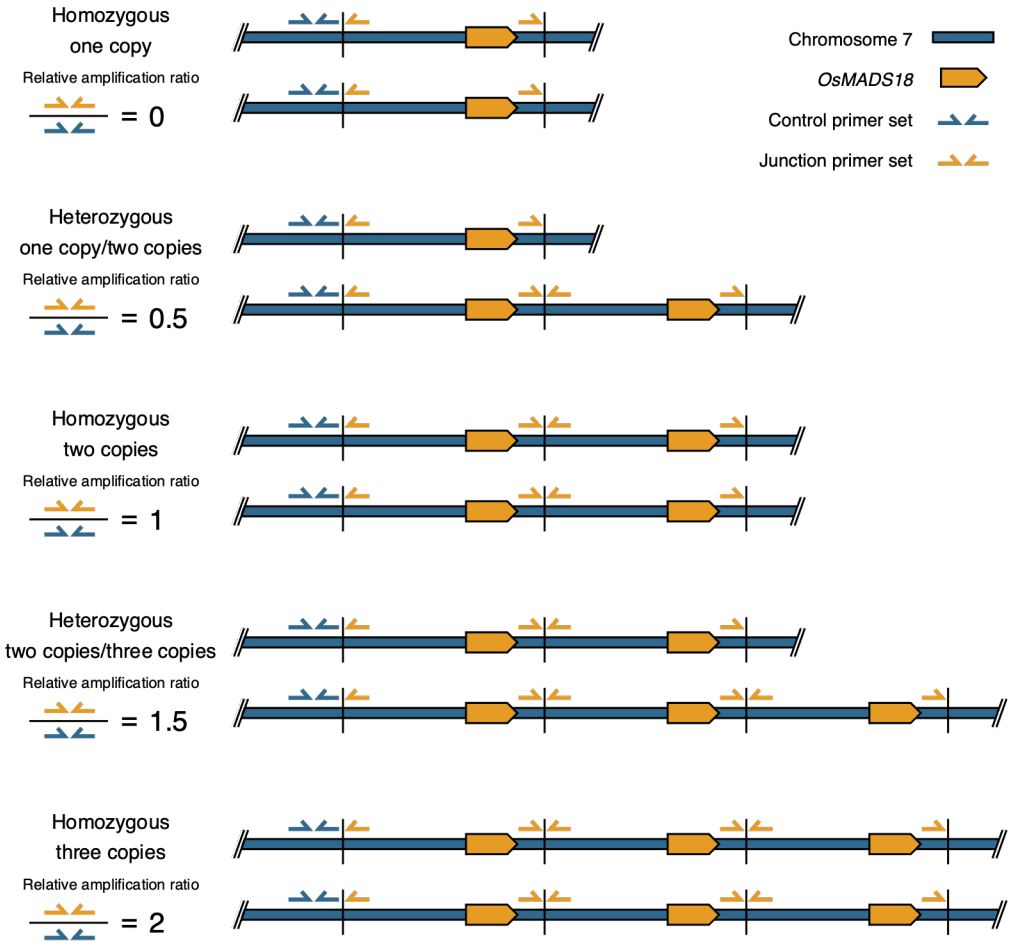

a
24.8 Mb
Chr. 7
Duplicated region 1
(20 kb)
Duplicated region 2
(20 kb)
Two copies (WT)
OsMADS18
OsMADS18
Using the CRISPR/Cas9 system
One copy
OsMADS18
Three copies
OsMADS18
OsMADS18
OsMADS18
b
Relative amplification ratio
1 CNV
2 CNV
3 CNV
c
Relative sequencing depth
1 CNV
2 CNV
3 CNV
Chromosome 7 (Mb)
d
Relative transcript level (OsMADS18/UBQ5)
e
1 CNV
2 CNV
3 CNV
f
Flag leaf blade length (cm)
g
Culm length (cm)

Fig.1

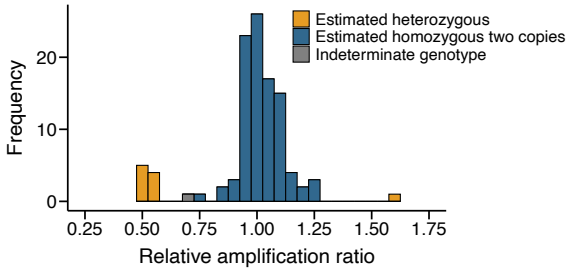


Extended Data Fig.1

a



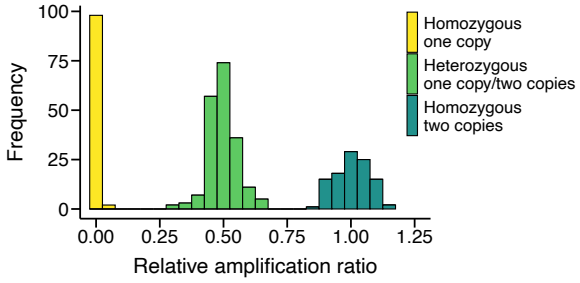
b



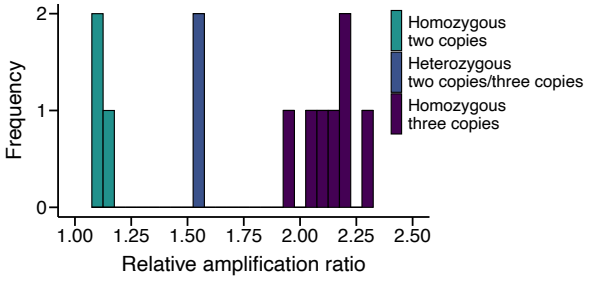
c

	Culture period on selection medium		Overall
	~6 weeks	~10 weeks	
Independently cultured T_0 plants	76	31	107
Estimated heterozygous one copy/two copy ($0.3 < x < 0.7$)	5 (6.6%)	4 (12.9%)	9 (8.4%)
Estimated heterozygous two copies/three copies ($1.3 < x < 1.7$)	0 (0.0%)	1 (3.2%)	1 (0.9%)

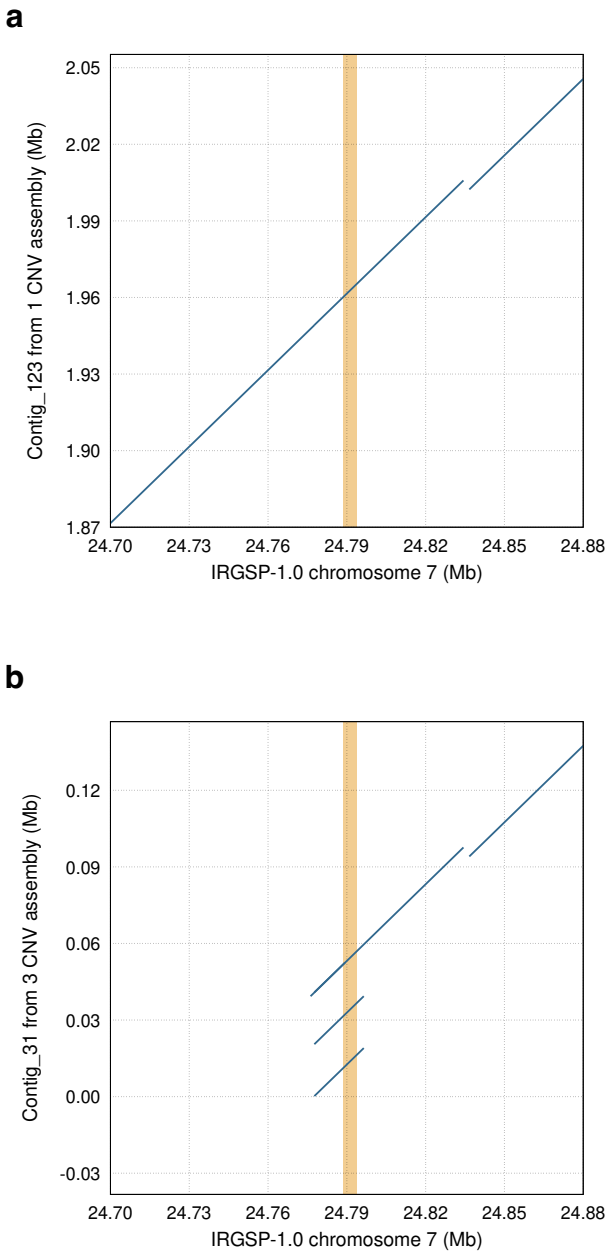
d



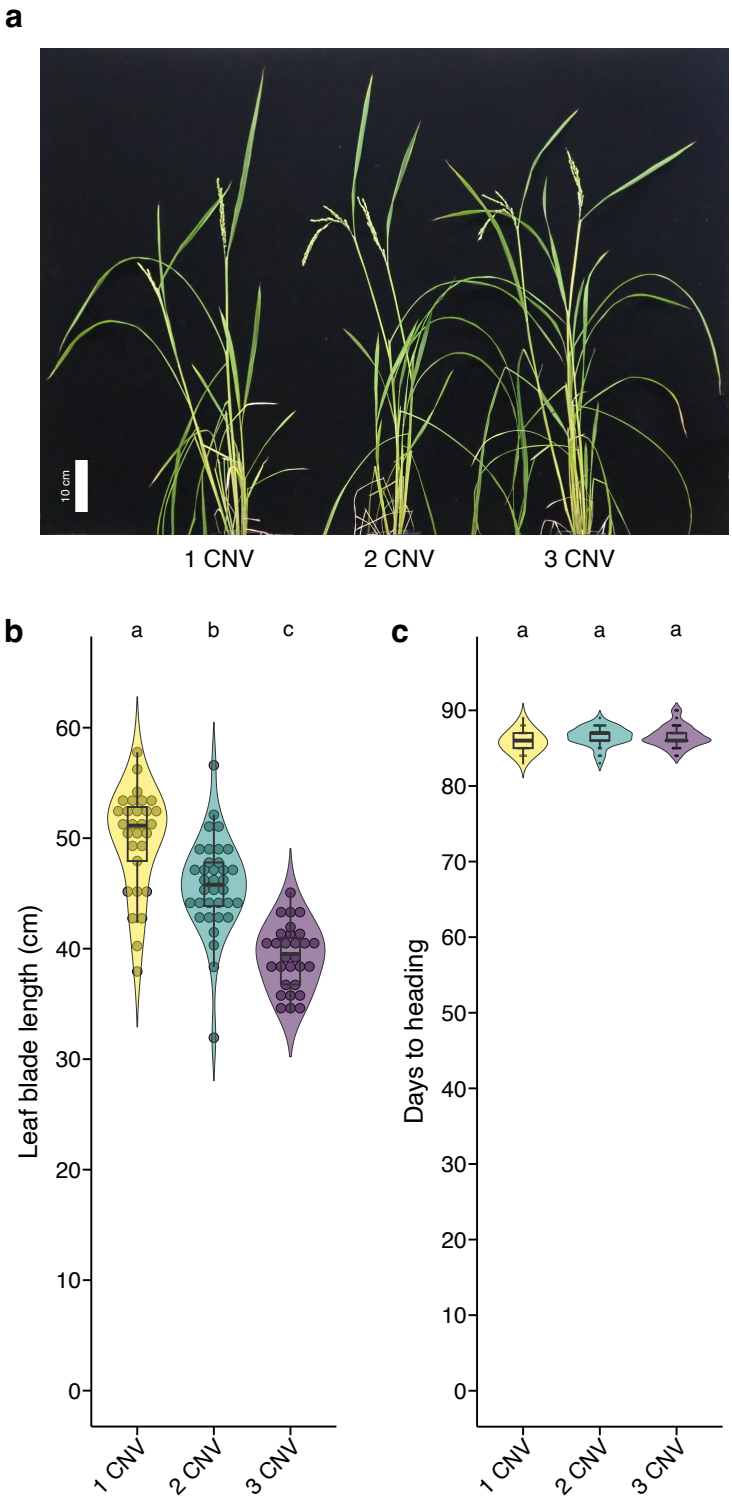
e



Extended Data Fig. 2



Extended Data Fig. 3



Extended Data Fig. 4