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] Department of Genomics and Breeding, Iwate Biotechnology Research Center, Kitakami, Iwate, 024-0003, Japan

[2] Crop Evolution Laboratory, Kyoto University, Muko, Kyoto, 617-0001, Japan

* Corresponding author

Phone: +81-197-68-2911

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

This manuscript was posted as a preprint at bioRxiv on May 16, 2025 (doi: https://doi.org/10.1101/2025.05.16.654416) and is available under International license CC-BY 4.0.

Keywords: copy number variation, CRISPR, OsMADS18, gene dosage, rice

Abstract

Although recent plant pan-genome studies have revealed extensive copy number variations, there are limited studies addressing the phenotypic consequences of these variations. Here, we manipulated the copy number of OsMADS18 in rice (Oryza sativa) through gene editing and evaluated the effects. OsMADS18, which encodes a MADS-box transcription factor known to regulate plant development, is tandemly duplicated in the elite rice cultivar 'Hitomebore' as a natural variant. We therefore edited the OsMADS18 copy number using a basic CRISPR/Cas9 system and established rice lines harboring one to three tandem copies of OsMADS18, as identified through high throughput qPCR screening. The presence of one to three OsMADS18 tandem copies was reflected in stepwise increases in transcript levels and concomitant trait values for leaf blade length and culm length. These results demonstrate that manipulating gene copy number can fine-tune important agronomic traits, providing a novel breeding strategy for crop improvement.

Introduction

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 (Koboldt *et al.*, 2013) to crop plants (Singh *et al.*, 2022). Recent developments in long-read sequencing technologies, such as the PacBio and Oxford Nanopore Technologies platforms, have helped reveal copy number variations (CNVs) and large structural variations in the genome, which were previously difficult to capture using short-read sequencing methods,

47 such as those involving the Illumina platform (Alonge *et al.*, 2020; Liu *et al.*, 2020; Qin
48 *et al.*, 2021). However, current *in silico* analyses exploring the relationships between
49 CNV and phenotype using genome sequences obtained through long-read sequencing
50 among accessions of a single species are based on materials with diverse genetic
51 backgrounds, making it difficult to demonstrate a direct dosage effect of CNV (Yuan *et*
52 *al.*, 2021). Therefore, there is a pressing need to study how variation in gene dosage
53 caused by CNV results in different transcript levels and ultimately influences plant
54 phenotypes (Maino *et al.*, 2021).

55 Genome editing is a powerful tool for introducing mutations into target genomic
56 locations (Gaj *et al.*, 2013). Among gene-editing methods, the clustered regularly
57 interspaced short palindromic repeats (CRISPR)/CRISPR-associated nuclease 9 (Cas9)
58 system is the most widely used for genetic research (Jinek *et al.*, 2012; Doudna &
59 Charpentier, 2014; Belhaj *et al.*, 2015). CRISPR/Cas9 typically cuts the genome near a
60 target site bound by a single guide RNA (sgRNA), after which the repair machinery
61 introduces mutations such as indels. CRISPR/Cas9-mediated gene editing can be used to
62 knock out genes to explore gene function (Shalem *et al.*, 2015; Bao *et al.*, 2019). However,
63 a complete loss-of-function mutant often displays drastic changes in phenotype compared
64 to the wild type (Pnueli *et al.*, 1998; Jeon *et al.*, 2000; Komatsu *et al.*, 2003; Qi *et al.*,
65 2008) and thus is not suitable for studying quantitative phenotypic effects associated with
66 subtle changes in gene expression. In addition to knocking out gene function,
67 CRISPR/Cas9-mediated genome editing can also delete larger genomic fragments by
68 using two sgRNAs that target two sites flanking the genomic sequence to be deleted
69 (Belhaj *et al.*, 2015).

70 We hypothesized that this approach could be used to increase or decrease the copy

number of a target gene, thereby revealing the quantitative effects of gene CNV on expression levels and phenotypes. In this study, we manipulated the copy number of the MADS-box transcription factor gene *OsMADS18* in rice (*Oryza sativa*) through gene editing and evaluated the effects, providing a powerful approach for fine-tuning complex traits in crops.

Materials and Methods

Plant Materials and Growth Conditions

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

Plasmid Construction and Plant Transformation

To generate lines with CNV at *OsMADS18*, one sgRNA targeting the duplicated genomic region was designed using CRISPRdirect (https://crispr.dbcls.jp/) (Naito *et al.*, 2015). To increase the probability of mutations, a binary vector containing two sgRNA expression cassettes was constructed based on a previous study (Ma *et al.*, 2015). The *OsU6a* and

95 *OsU6b* promoters, amplified using pYLsgRNA-OsU6a and pYLsgRNA-OsU6b as
96 templates, were each fused to the target sequence and sgRNA scaffold. The ccdB cassette
97 of the pYLCRISPR/Cas9P_{ubi}-H vector was replaced with the two sgRNA cassettes via the
98 BsaI site using Golden Gate assembly. This vector was introduced into *Agrobacterium*
99 *tumefaciens* strain EHA105, and rice transformation in the Hitomebore background was
100 performed as described previously (Toki *et al.*, 2006). The pYLCRISPR/Cas9P_{ubi}-H
101 (RRID:Addgene_66187), pYLsgRNA-OsU6a (RRID:Addgene_66194), and
102 pYLsgRNA-OsU6b (RRID:Addgene_66196) vectors were kindly provided by Yao-
103 Guang Liu. The primers used in this study are listed in Table **S1**.
104

105 ## ***De Novo* Assembly**

106 To reconstruct the Hitomebore and Sasanishiki genomes, *de novo* assembly of each
107 genome was performed using Nanopore long reads and Illumina short reads following a
108 previously published method (Sugihara *et al.*, 2023). To extract high-molecular-weight
109 genomic DNA from rice leaf tissue for Nanopore sequencing, leaf blades were carefully
110 ground using a mortar and pestle while being cooled with liquid nitrogen. Genomic DNA
111 was then extracted using a NucleoBond High-molecular-weight DNA Kit (Macherey-
112 Nagel, Düren, Nordrhein-Westfalen, Germany). Following DNA extraction, low-
113 molecular-weight DNA was eliminated using a Short Read Eliminator Kit XL
114 (Circulomics, Baltimore, MD, USA). For Hitomebore, genomic library preparation was
115 performed using a Ligation Sequencing Kit (SQK-LSK-114; Oxford Nanopore
116 Technologies [ONT], Oxford, Oxfordshire, UK) according to the manufacturer's
117 instructions, and sequencing was performed using PromethION 2 Solo (ONT). Long
118 reads for Hitomebore produced on the PromethION (ONT) platform, as described by

119 Sugihara *et al.* (2023), were also used. For Sasanishiki, genomic library preparation was
120 performed using a Ligation Sequencing Kit (SQK-LSK-109; ONT), and sequencing was
121 performed on a MinION instrument.

122 Base-calling of the Nanopore long reads was conducted using Dorado v0.9.0
123 (https://github.com/nanoporetech/dorado). The first 50 bp of each read and reads with an
124 average read quality score below 10 were removed using Chopper v0.9.0 (De Coster &
125 Rademakers, 2022). The resulting clean Nanopore long reads were assembled using
126 NECAT v0.0.1 (Chen *et al.*, 2021), with the genome size set to 380 Mb and with reads
127 longer than 3,000 bp. To improve the accuracy of the assembly, Racon v1.4.20 (Vaser *et*
128 *al.*, 2017) was employed for error correction using Nanopore reads, followed by Medaka
129 v1.4.1 (https://github.com/nanoporetech/medaka) for misassembly correction. Two
130 rounds of consensus correction were then performed using BWA v0.7.17 (Li, 2013) and
131 HyPo v1.0.3 (Kundu *et al.*, 2019) with the Illumina short reads. Redundant contigs were
132 removed using Purge Haplotigs v1.1.1 (Roach *et al.*, 2018), resulting in a *de novo*
133 assembly of 376.2 Mb comprising 59 contigs for Hitomebore and 378.7 Mb consisting of
134 49 contigs for Sasanishiki.

135 *De novo* assembly was also performed for Hitomebore gene-edited lines with one or
136 three copies of the *OsMADS18* region. In both cases, high-molecular-weight genomic
137 DNA was extracted as described above. The library for the single-copy line was prepared
138 using an SQK-LSK110 Kit and sequenced on a MinION instrument, while the library for
139 the three-copy line was prepared using an SQK-LSK114 Kit and sequenced on a
140 PromethION 2 Solo instrument. Base-calling was performed as described above, and *de*
141 *novo* assembly was conducted using Flye v2.9.5 (Kolmogorov *et al.*, 2019).

142

Illumina-Based DNA Sequencing

Total genomic DNA was extracted from rice leaf blades using the CTAB-based method with a NucleoSpin Plant II Kit (Macherey-Nagel). Sequencing libraries were prepared using a Collibri ES DNA Library Prep Kit for Illumina Systems (Thermo Fisher Scientific, Waltham, MA, USA). Libraries were sequenced on a MiSeq (Illumina, San Diego, CA, USA) instrument in 251-bp paired-end mode for Hitomebore and in 301-bp paired-end mode for Sasanishiki and on a NovaSeq X Plus (Illumina) instrument in 151-bp paired-end mode for lines with CNV (ranging from one to three copies) at *OsMADS18*. The resulting raw reads from the MiSeq instrument were filtered using Trimmomatic v0.36 (Bolger *et al.*, 2014) with the paired-end option. Adapter sequences were removed using the ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10 option. Bases with quality scores of <20 at the beginning and end of reads were removed (LEADING:20, TRAILING:20). A sliding window of four bases was used to remove reads with average quality scores of <20 (SLIDINGWINDOW:4:20). Reads were removed if the read length after trimming was <36 (MINLEN:36) and were cut to 200 bases if the read length was >200 bases (CROP:200). The resulting raw reads from the NovaSeq X Plus instrument were filtered using fastp v0.22.0 (Chen *et al.*, 2018) with default settings supplemented with the --detect_adapter_for_pe option to automatically detect and trim adapter sequences and the --trim_poly_g option to trim polyG tails. The processed reads were aligned to the rice reference genome IRGSP-1.0 using BWA v0.7.18. The resulting SAM files were converted to BAM format, sorted, and indexed using SAMtools v1.9 (Danecek *et al.*, 2021).

Analysis of Sequencing Depth

Sequencing depth was calculated using mosdepth v0.3.3 (Pedersen & Quinlan, 2017) with the BAM files generated from high-throughput sequencing and subsequent processing. The sliding window used in the analysis was created with BEDTools v2.31.1 (Quinlan, 2014) using the command 'makewindows'. Relative sequencing depth between the target and control was calculated using the following equation:

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

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 of the target *OsMADS18* region was compared using MUMmer4 v4.0.1 (Marçais *et al.*, 2018). First, the 'blastn' command from BLAST+ v2.13.0 (Camacho *et al.*, 2009) 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 (Shen *et al.*, 2024). Each contig was compared to chromosome 7 of the IRGSP-1.0 reference genome using the 'nucmer' command of MUMmer4. The results were filtered using the 'delta-filter' command, setting the minimum alignment identity to 95% and the minimum alignment length to 5,000 bp. Finally, the filtered results were visualized using the 'mummerplot' command.

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 (Katoh & Standley, 2013). Sequence differences were then identified based on the alignment results.

The duplicated regions in Hitomebore and the corresponding region in Sasanishiki were visualized using the Integrative Genomics Viewer (IGV) v2.16.2 (Robinson *et al.*, 2011), with the IRGSP-1.0 reference genome.

qPCR-Based Analysis of Target Gene 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, MA, USA). To estimate the 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 (Pfaffl, 2001), 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:

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

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

Further details are provided in Fig. **S1**, and the primer sequences used in the

experiments are listed in Table **S1**.

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. Each sample was ground to a powder, and total RNA was extracted from the samples using a NucleoSpin RNA Plant Kit (Macherey-Nagel). The total RNA was reverse transcribed into first-strand cDNA using PrimeScript RT Master Mix (Takara Bio, Kusatsu, Shiga, Japan). qPCR was performed using a QuantStudio 3 Real-Time PCR System with Luna Universal qPCR Master Mix. The $\Delta\Delta C_t$ method was used to quantify transcript levels, and the relative transcript levels of *OsMADS18* were calculated using *UBQ5* (Jain *et al.*, 2006, 2018) as the internal control. The mean C_t value of biological replicates from plants with one homozygous copy of *OsMADS18* was used as the calibrator. The primers used in the experiment are listed in Table **S1**.

Phenotyping

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

238 ## Statistical Analysis

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

242

243 # Results

244 ## *OsMADS18* is Tandemly Duplicated in Hitomebore

245 MADS-box transcription factors are widely distributed across eukaryotes. For example,
246 OsMADS18 from rice is involved in the transition from the vegetative to the reproductive
247 stages (Fornara *et al.*, 2004; Kobayashi *et al.*, 2012) and exerts pleiotropic effects on
248 germination and tillering (Yin *et al.*, 2019). A previous study of a pan-genome from 33
249 genetically diverse rice accessions revealed CNVs for *OsMADS18* that arose via tandem
250 duplication (Qin *et al.*, 2021). Here, we generated *de novo* genome assemblies for the
251 elite *japonica* rice cultivar 'Hitomebore' and the *japonica* cultivar 'Sasanishiki' using
252 long reads generated by Oxford Nanopore sequencing and compared the results to the
253 reference genome IRGSP-1.0 from the rice cultivar 'Nipponbare' (Kawahara *et al.*, 2013)
254 (Fig. **S2a,b**). We also assessed sequencing coverage using Illumina-based short reads
255 between Hitomebore and Sasanishiki (Figs. **S2c**, **S3a,b**).

256 We confirmed the presence of a 20-kb tandem duplication at 24.776–24.796 Mb
257 on chromosome 7 in the Hitomebore genome and its absence from the Sasanishiki and
258 IRGSP-1.0 reference genomes. According to the Rice Annotation Project Database (RAP-
259 DB) (Kawahara *et al.*, 2013; Sakai *et al.*, 2013), the only full-length protein-coding gene
260 contained in this 20-kb fragment is *OsMADS18*, which is duplicated specifically in
261 Hitomebore (Figs. **S2a–c**, **S3**). We detected only one SNP and three indels across the two

262 tandem repeats of the 20-kb fragment, with no variation in the two copies of the
263 *OsMADS18* coding region (Fig. **S2d**). Given that *OsMADS18* is the only full-length
264 protein-coding gene within the duplicated region and that sequence variation between the
265 two tandem copies is minimal, we reasoned that CNV of *OsMADS18* might provide an
266 attractive model for the genetic evaluation of gene dosage effects. We thus attempted to
267 locally manipulate the copy number of *OsMADS18* using CRISPR/Cas9 and study how
268 changes in gene copy number would affect plant phenotypes.
269

270 CRISPR-Mediated Manipulation of *OsMADS18* Copy Number

271 To manipulate the *OsMADS18* copy number, we designed an sgRNA targeting a site 12
272 kb upstream of the ATG of *OsMADS18*, within the region tandemly duplicated in
273 Hitomebore, for CRISPR/Cas9-mediated editing in this cultivar. Since the two copies of
274 *OsMADS18* have the same sequence, this sgRNA has two target sites in this cultivar (Fig.
275 **1a**). Among the T_0 individuals regenerated from Hitomebore calli, we looked for
276 individuals with altered *OsMADS18* copy numbers using quantitative PCR (qPCR) with
277 two primer sets (Fig. **S1**): one set amplifying an amplicon outside of the duplicated region,
278 thus maintaining a constant copy number, and another set designed to anneal to either
279 side of the junction of the duplicated region to specifically amplify the junction in the
280 presence of a tandem duplication. We calculated the relative amplification ratio of these
281 two primer sets using the efficiency-corrected Pfaffl method (Pfaffl, 2001). When the
282 relative amplification ratio of Hitomebore (with two homozygous copies of the
283 duplication) was set to 1, deviations of the relative ratio from 1 in increments or
284 decrements of 0.5 units reflected a change in *OsMADS18* copy number (Fig. **S1**).
285 Of the 107 regenerated T_0 plants obtained from different calli or from the same

286 callus with different culture periods, qPCR analysis identified nine plants with a relative
287 amplification ratio of ~0.5 and one with a relative amplification ratio of ~1.5, suggesting
288 the loss or gain of one *OsMADS18* copy, respectively (Fig. **S4a,b**; Dataset **S1**). We
289 allowed these plants to self-pollinate and repeated the qPCR analysis on their progeny.
290 From the parents with a relative amplification ratio of ~0.5, we obtained progeny with
291 relative amplification ratios of 0 (reflecting absence of the junction sequence of the
292 duplicated region), ~0.5, or ~1 (Fig. **S4c**; Dataset **S1**). From the parents with a relative
293 amplification ratio of ~1.5, we obtained progeny with relative amplification ratios of ~1,
294 ~1.5, or ~2 (Fig. **S4d**; Dataset **S1**). Furthermore, plant lines with relative amplification
295 ratios of ~0, ~1, or ~2 produced progeny with the same relative amplification ratios,
296 suggesting that these plants were homozygous for one, two, or three copies of *OsMADS18*,
297 respectively (Fig. **1b**; Dataset **S2**). These results demonstrate that our qPCR-based
298 approach is practical for efficiently detecting genomic CNVs.

299 To confirm these copy numbers, we examined the sequencing coverage depth over
300 the *OsMADS18* genomic region using Illumina short reads from homozygous plants with
301 an estimated one, two, or three tandem copies. The relative sequencing depth within the
302 *OsMADS18* region was lower in plants with one copy of *OsMADS18* and higher in plants
303 with three copies of *OsMADS18* relative to those with two copies of this gene (Fig. **1c**).
304 *De novo* genome assembly using Nanopore long reads for the genomes of homozygous
305 individuals with an estimated one or three *OsMADS18* copies in the Hitomebore
306 background showed that the *OsMADS18* copy number was indeed one and three,
307 respectively, as evidenced by dot matrix analysis (Fig. **S5**).

308

309 **Copy Number Variation at *OsMADS18* Modulates Agronomic Traits in**

310 **a Stepwise Manner**

311 To study the effect of CNV at *OsMADS18* on its transcript levels, we performed reverse-
312 transcription qPCR (RT-qPCR) on individuals with one to three copies of the gene,
313 normalizing to *UBQ5* (Jain *et al.*, 2006, 2018), with the mean relative *OsMADS18*
314 transcript level for plants with one copy set to 1. The relative *OsMADS18* transcript levels
315 gradually increased from an average of 1 to 2 or 3 in plants homozygous for one, two, or
316 three copies of *OsMADS18*, respectively, thus neatly correlating with gene copy number
317 (Fig. **1d**; Dataset **S2**). These findings suggest that simple genome editing introduced CNV
318 that resulted in stepwise changes in *OsMADS18* expression levels.

319 We assessed the agronomic traits of lines with CNV (from one to three copies) at
320 *OsMADS18*. We observed no major difference in overall plant architecture among the rice
321 lines (Fig. **S6a**). However, the flag leaf blade was significantly shorter when the
322 *OsMADS18* copy number was higher (Fig. **1e,f**; Dataset **S2**), as was the length of the leaf
323 blade just below the flag leaf (Fig. **S6b**; Dataset **S3**). By contrast, culm length was
324 significantly greater in plants with more *OsMADS18* copies (Fig. **1g**; Dataset **S2**). To our
325 knowledge, effects of *OsMADS18* on leaf blade size and culm length have not been
326 previously reported. Previous studies have shown that *OsMADS18* is upregulated during
327 the transition from vegetative to reproductive growth and is involved in the number of
328 days to heading (Fornara *et al.*, 2004; Kobayashi *et al.*, 2012). However, in the current
329 study, we measured no significant difference in the number of days to heading among the
330 three sets of lines with CNV at *OsMADS18* (Fig. **S6c**; Dataset **S3**).

331

332 # Discussion

333 Most agronomic traits are quantitative in nature, and their variations are controlled by

subtle differences in the expression levels of multiple genes. Thus, manipulating gene expression is an important method for adjusting traits of interest. Promoter editing was recently developed as a means to generate a large allelic series of plants with different expression levels of a target gene by inserting random mutations in the *cis*-regulatory elements of promoters (Rodríguez-Leal *et al.*, 2017). However, because this approach targets *cis*-regulatory elements, it is difficult to predict the direction and degree of changes in expression levels in the edited plants. Moreover, the spatiotemporal gene expression pattern may differ from that driven by the intact promoter. We propose that manipulating gene copy number is a powerful alternative method to create lines with a range of copy numbers, leading to stepwise changes in the expression levels of a target gene without randomly perturbing its spatiotemporal expression patterns. This method should allow for the fine-tuning of important traits by optimizing gene copy number, thus accelerating plant breeding.

Acknowledgements

We thank Yao-Guang Liu for providing the pYLCRISPR/Cas9P_{ubi}-H, pYLsgRNA-OsU6a, and pYLsgRNA-OsU6b plasmids, as well as Yumiko Ogasawara and Miyuki Saito for technical support. Part of the analysis was performed using the National Institute of Genetics (NIG) supercomputer at the ROIS National Institute of Genetics. This research was partially supported by the following grants: the Research Program on the Development of Innovative Technology (JPJ007097) from the Project of the Bio-oriented Technology Research Advancement Institution (BRAIN); and JSPS KAKENHI (JP22K20584 to C.N. and JP23K26882 to A.A.).

Competing interest

None declared.

Author contributions

CN and AA conceived and designed the research. CN, HK, EK, MS, KO, HU, KI, YS, and AA performed the experiments. CN and AA conducted bioinformatics analyses. CN, RT, and AA wrote and revised the manuscript. All authors read and approved the final version of the manuscript.

ORCID

Chihiro Nomura https://orcid.org/0009-0001-4574-4279

Motoki Shimizu https://orcid.org/0000-0002-5622-5554

Yusaku Sugimura https://orcid.org/0000-0001-7271-7703

Ryohei Terauchi https://orcid.org/0000-0002-0095-4651

Akira Abe https://orcid.org/0000-0002-0344-2643

Data availability

The raw sequencing reads have been deposited in the DNA Data Bank of Japan (DDBJ) under BioProject IDs PRJDB13864 and PRJDB20676, with the accession numbers provided in Table S2. 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 used in this study are

provided in Datasets S1, S2, and S3.

References

Alonge M, Wang X, Benoit M, Soyk S, Pereira L, Zhang L, Suresh H, Ramakrishnan S, Maumus F, Ciren D, *et al.* 2020. Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell* **182**: 145–161.e23.

Bao A, Burritt DJ, Chen H, Zhou X, Cao D, Tran L-SP. 2019. The CRISPR/Cas9 system and its applications in crop genome editing. *Critical Reviews in Biotechnology* **39**: 321–336.

Belhaj K, Chaparro-Garcia A, Kamoun S, Patron NJ, Nekrasov V. 2015. Editing plant genomes with CRISPR/Cas9. *Current Opinion in Biotechnology* **32**: 76–84.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Chen Y, Nie F, Xie S-Q, Zheng Y-F, Dai Q, Bray T, Wang Y-X, Xing J-F, Huang Z-J, Wang D-P, *et al.* 2021. Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nature Communications* **12**: 60.

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

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, *et al.* 2021. Twelve years of SAMtools and BCFtools. *GigaScience* **10**: giab008.

De Coster W, Rademakers R. 2022. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* **39**: btad311.

Doudna JA, Charpentier E. 2014. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**: 1258096.

Fornara F, Parenicová L, Falasca G, Pelucchi N, Masiero S, Ciannamea S, Lopez-

Dee Z, Altamura MM, Colombo L, Kater MM. 2004. Functional characterization of *OsMADS18*, a member of the *AP1*/*SQUA* subfamily of MADS box genes. *Plant Physiology* **135**: 2207–2219.

Gaj T, Gersbach CA, Barbas CF 3rd. 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology* **31**: 397–405.

Jain M, Nijhawan A, Tyagi AK, Khurana JP. 2006. Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. *Biochemical and Biophysical Research Communications* **345**: 646–651.

Jain N, Vergish S, Khurana J. 2018. Validation of house-keeping genes for normalization of gene expression data during diurnal/circadian studies in rice by RT-qPCR. *Scientific Reports* **8**: 3203.

Jeon JS, Jang S, Lee S, Nam J, Kim C, Lee SH, Chung YY, Kim SR, Lee YH, Cho YG, *et al.* 2000. *leafy hull sterile1* is a homeotic mutation in a rice MADS box gene affecting rice flower development. *Plant Cell* **12**: 871–884.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**: 816–821.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780.

Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S, *et al.* 2013. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* **6**: 4.

Kobayashi K, Yasuno N, Sato Y, Yoda M, Yamazaki R, Kimizu M, Yoshida H, Nagamura Y, Kyozuka J. 2012. Inflorescence meristem identity in rice is specified by overlapping functions of three *AP1*/*FUL*-like MADS box genes and *PAP2*, a *SEPALLATA* MADS box gene. *Plant Cell* **24**: 1848–1859.

Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. 2013. The next-generation sequencing revolution and its impact on genomics. *Cell* **155**: 27–38.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology* **37**: 540–546.

Komatsu M, Chujo A, Nagato Y, Shimamoto K, Kyozuka J. 2003. *FRIZZY PANICLE* is required to prevent the formation of axillary meristems and to establish floral meristem identity in rice spikelets. *Development* **130**: 3841–3850.

Kundu R, Casey J, Sung W-K. 2019. HyPo: super fast & accurate polisher for long read genome assemblies. *bioRxiv*. doi: https://doi.org/10.1101/2019.12.19.882506

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. doi: https://doi.org/10.48550/arXiv.1303.3997

Liu Y, Du H, Li P, Shen Y, Peng H, Liu S, Zhou G-A, Zhang H, Liu Z, Shi M, *et al.* 2020. Pan-genome of wild and cultivated soybeans. *Cell* **182**: 162–176.e13.

Ma X, Zhang Q, Zhu Q, Liu W, Chen Y, Qiu R, Wang B, Yang Z, Li H, Lin Y, *et al.* 2015. A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Molecular Plant* **8**: 1274–1284.

Maino E, Wojtal D, Evagelou SL, Farheen A, Wong TWY, Lindsay K, Scott O, Rizvi SZ, Hyatt E, Rok M, *et al.* 2021. Targeted genome editing *in vivo* corrects a *Dmd* duplication restoring wild-type dystrophin expression. *EMBO Molecular Medicine* **13**: e13228.

Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4: a fast and versatile genome alignment system. *PLoS Computational Biology* **14**: e1005944.

Naito Y, Hino K, Bono H, Ui-Tei K. 2015. CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics* **31**: 1120–1123.

Pedersen BS, Quinlan A. 2017. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**: 867–868.

Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* **29**: e45.

Pnueli L, Carmel-Goren L, Hareven D, Gutfinger T, Alvarez J, Ganal M, Zamir D, Lifschitz E. 1998. The *SELF-PRUNING* gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of *CEN* and *TFL1*.

Development **125**: 1979–1989.

Qi J, Qian Q, Bu Q, Li S, Chen Q, Sun J, Liang W, Zhou Y, Chu C, Li X, *et al.* 2008. Mutation of the rice *Narrow leaf1* gene, which encodes a novel protein, affects vein patterning and polar auxin transport. *Plant Physiology* **147**: 1947–1959.

Qin P, Lu H, Du H, Wang H, Chen W, Chen Z, He Q, Ou S, Zhang H, Li X, *et al.* 2021. Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell* **184**: 3542–3558.e16.

Quinlan AR. 2014. BEDTools: the Swiss-army tool for genome feature analysis. *Current Protocols in Bioinformatics* **47**: 11.12.1–11.12.34.

Roach MJ, Schmidt SA, Borneman AR. 2018. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**: 460.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nature Biotechnology* **29**: 24–26.

Rodríguez-Leal D, Lemmon ZH, Man J, Bartlett ME, Lippman ZB. 2017. Engineering quantitative trait variation for crop improvement by genome editing. *Cell* **171**: 470–480.e8.

Sakai H, Lee SS, Tanaka T, Numa H, Kim J, Kawahara Y, Wakimoto H, Yang C-C, Iwamoto M, Abe T, *et al.* 2013. Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics. *Plant and Cell Physiology* **54**: e6.

Shalem O, Sanjana NE, Zhang F. 2015. High-throughput functional genomics using CRISPR-Cas9. *Nature Reviews Genetics* **16**: 299–311.

Shen W, Sipos B, Zhao L. 2024. SeqKit2: a Swiss army knife for sequence and alignment processing. *iMeta* **3**: e191.

Singh R, Kumar K, Bharadwaj C, Verma PK. 2022. Broadening the horizon of crop research: a decade of advancements in plant molecular genetics to divulge phenotype governing genes. *Planta* **255**: 46.

Sugihara Y, Abe Y, Takagi H, Abe A, Shimizu M, Ito K, Kanzaki E, Oikawa K, Kourelis J, Langner T, *et al.* 2023. Disentangling the complex gene interaction networks between rice and the blast fungus identifies a new pathogen effector. *PLoS Biology* **21**: e3001945.

Toki S, Hara N, Ono K, Onodera H, Tagiri A, Oka S, Tanaka H. 2006. Early infection of scutellum tissue with *Agrobacterium* allows high-speed transformation of rice. *Plant Journal* **47**: 969–976.

Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research* **27**: 737–746.

Yin X, Liu X, Xu B, Lu P, Dong T, Yang D, Ye T, Feng Y-Q, Wu Y. 2019. OsMADS18, a membrane-bound MADS-box transcription factor, modulates plant architecture and the abscisic acid response in rice. *Journal of Experimental Botany* **70**: 3895–3909.

Yuan Y, Bayer PE, Batley J, Edwards D. 2021. Current status of structural variation studies in plants. *Plant Biotechnology Journal* **19**: 2153–2163.

Supporting information

Dataset S1 Source data for all panels in Fig. S4.

Dataset S2 Source data for panels (b), (d), (f), and (g) in Fig. 1.

Dataset S3 Source data for panels (b) and (c) in Fig. S6.

Fig. S1 Diagram of the method used to determine *OsMADS18* copy number by qPCR.

Fig. S2 Tandem duplication of the *OsMADS18* genomic region on chromosome 7 in Hitomebore.

Fig. S3 Genome browser views of the *OsMADS18* genomic region on chromosome 7.

Fig. S4 *OsMADS18* CNV in T_0 plants and progeny as determined by qPCR.

Fig. S5 Genomic structure of the *OsMADS18* locus in IRGSP-1.0 and CNV-edited lines.

Fig. S6 Effects of *OsMADS18* CNV on other agronomic traits.

Table S1 List of primers used in this study.

Table S2 List of sequencing data and corresponding Sequence Read Archive (SRA) accession numbers.

525 **Figure legend**

526 **Fig. 1.** Phenotypic variation associated with the manipulation of *OsMADS18* copy
527 number. (a) Diagram of the method used to induce CNV of *OsMADS18*. The sgRNA
528 targets two identical sequences within the tandemly duplicated *OsMADS18* genomic
529 region. CRISPR/Cas9-mediated editing can result in plants with either one or three copies
530 of *OsMADS18*. Target sites of the sgRNA are indicated by scissor icons. (b) Relative
531 amplification ratio of the tandemly duplicated *OsMADS18* region, as determined by
532 qPCR (n = 18 for each CNV). (c) Relative sequencing depth of Illumina short reads
533 around *OsMADS18*. A sliding window with a window size of 1,000 bp and a step size of
534 500 bp was used. Thin lines represent individual plants, and the thick lines indicate the
535 mean depth for plants with each *OsMADS18* CNV (n = 5). The mean value for plants with
536 one homozygous copy of *OsMADS18* was set to 1. The orange shaded area marks the
537 position of *OsMADS18*. (d) Relative *OsMADS18* transcript levels in leaf blades, as
538 determined by RT-qPCR and normalized to *UBQ5*, with the mean transcript level of plants
539 with one copy set to 1. Leaf blades were bulk collected from three individuals at 58 days
540 after sowing (n = 10; total of 30 plants per *OsMAD18* CNV). (e) Representative
541 photograph of flag leaves. (f, g) Lengths of the flag leaf blade (n = 29, 34, and 29 for 1,
542 2, and 3 CNV, respectively) (f) and culm (n = 30, 35, and 36 for 1, 2, and 3 CNV,
543 respectively) (g). "1 CNV," "2 CNV," and "3 CNV" refer to plants with one, two, and
544 three homozygous copies of *OsMADS18*, respectively, for (b–g). Different lowercase
545 letters indicate significant differences (P < 0.05) based on the Tukey-Kramer test for
546 panels (d), (f), and (g). All violin plots were generated using kernel density estimation
547 with a smoothing adjustment factor of 1.5. All internal boxplots represent the first and
548 third quartiles (the lower and upper hinges, respectively), the median (the middle

549 horizontal line), and the range from the hinges to the smallest and largest observations
550 within 1.5 times the interquartile range (the bottom and top whiskers, respectively). All
551 data points, including outliers in the boxplot, are displayed as dots in the dot plot.

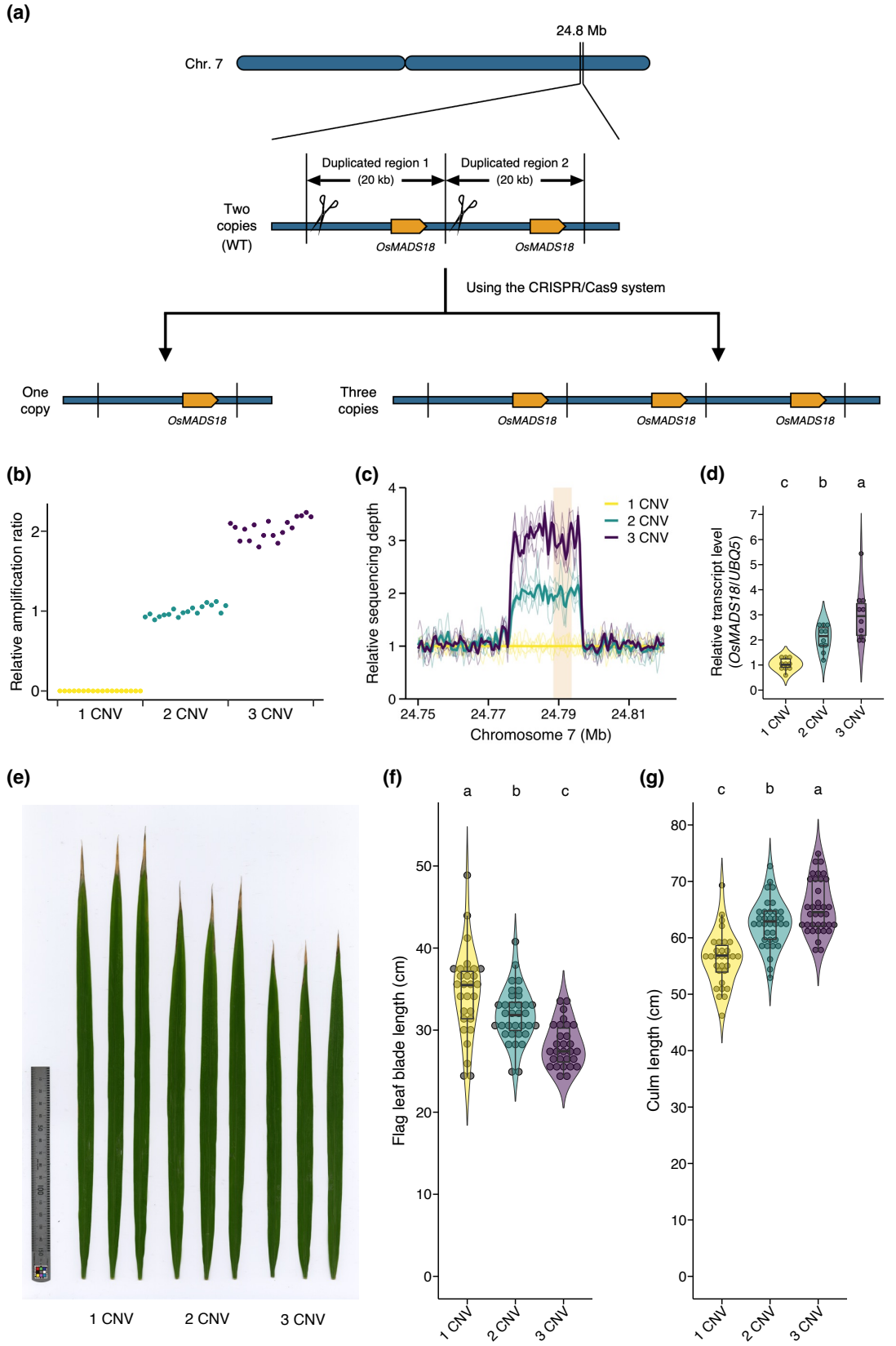
(a)
24.8 Mb
Chr. 7
Duplicated region 1
(20 kb)
Duplicated region 2
(20 kb)
Two copies (WT)
OsMADS18
OsMADS18
Using the CRISPR/Cas9 system
One copy
OsMADS18
Three copies
OsMADS18
OsMADS18
OsMADS18
(b)
Relative amplification ratio
1 CNV
2 CNV
3 CNV
(c)
Relative sequencing depth
1 CNV
2 CNV
3 CNV
Chromosome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

Fig. 1