Article

A route to *de novo* domestication of wild allotetraploid rice

Graphical Abstract

Highlights

- Establishing efficient transformation and genome editing system in allotetraploid rice

- The high-quality assembly of the genome of allotetraploid rice

- Identification of *O. alta* homologs of agronomically important genes from diploid rice

- Achieved targeted improvement of various traits in *O. alta* through genome editing
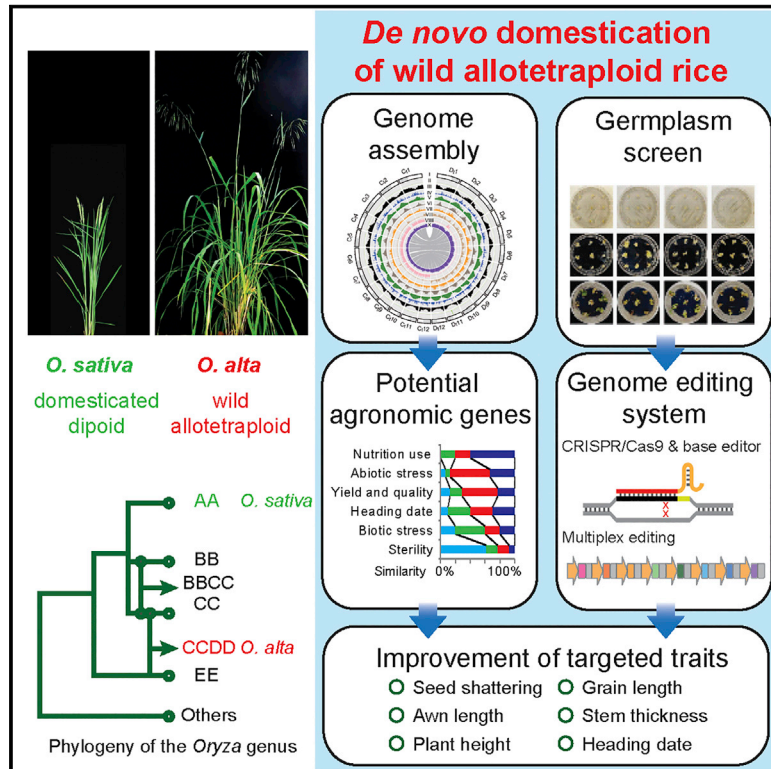
Authors

Hong Yu, Tao Lin, Xiangbing Meng, ..., Caixia Gao, Chengzhi Liang, Jiayang Li

Correspondence

hyu@genetics.ac.cn (H.Y.),
cxgao@genetics.ac.cn (C.G.),
jyli@genetics.ac.cn (J.L.),
cliang@genetics.ac.cn (C.L.)

In Brief

Li and colleagues developed a breeding route to *de novo* domestication of wild allotetraploid rice that provides a rational strategy for creating novel crops and generated a series of allotetraploid rice lines edited in domestication-related and agronomically important genes.

CellPress

Cell

CellPress

Article

A route to *de novo* domestication of wild allotetraploid rice

Hong Yu,[1,15,*] Tao Lin,[1,14,15] Xiangbing Meng,[1,15] Huilong Du,[1,2,15] Jingkun Zhang,[1,2,15] Guifu Liu,[1] Mingjiang Chen,[1] Yanhui Jing,[1] Liquan Kou,[1] Xiuxiu Li,[1,2] Qiang Gao,[1] Yan Liang,[1] Xiangdong Liu,[3,4] Zhilan Fan,[5] Yuntao Liang,[6] Zhukuan Cheng,[1,2] Mingsheng Chen,[1,2] Zhixi Tian,[2,7] Yonghong Wang,[1,2] Chengcai Chu,[1,2] Jianru Zuo,[1,2] Jianmin Wan,[8] Qian Qian,[9] Bin Han,[10] Andrea Zuccolo,[11,12] Rod A. Wing,[11,13] Caixia Gao,[2,7,*] Chengzhi Liang,[1,2,*] and Jiayang Li[1,2,3,16,*]

[1]State Key Laboratory of Plant Genomics, and National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Innovation Academy for Seed Design, Chinese Academy of Sciences, Beijing 100101, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China
[3]Guangdong Laboratory of Lingnan Modern Agriculture, Guangzhou 510642, China
[4]State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, South China Agricultural University, Guangzhou 510642, China
[5]National Field Genebank for Wild Rice (Guangzhou), Rice Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China
[6]Rice Research Institute, Guangxi Academy of Agricultural Science, Nanning 530007, China
[7]State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Innovation Academy for Seed Design, Chinese Academy of Sciences, Beijing 100101, China
[8]National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China
[9]State Key Laboratory of Rice Biology, China National Rice Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou 310006, China
[10]National Center of Plant Gene Research Shanghai Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences and CAS Center of Excellence for Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 200233, China
[11]Center for Desert Agriculture, Biological and Environmental Sciences & Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia
[12]Institute of Life Sciences, Scuola Superiore Sant'Anna, Pisa 56127, Italy
[13]Arizona Genomics Institute, School of Plant Sciences, University of Arizona, Tucson, AZ 85721, USA
[14]Present address: College of Horticulture, China Agricultural University, Beijing 100193, China
[15]These authors contributed equally
[16]Lead contact
*Correspondence: hyu@genetics.ac.cn (H.Y.), cxgao@genetics.ac.cn (C.G.), cliang@genetics.ac.cn (C.L.), jyli@genetics.ac.cn (J.L.)
https://doi.org/10.1016/j.cell.2021.01.013

SUMMARY

Cultivated rice varieties are all diploid, and polyploidization of rice has long been desired because of its advantages in genome buffering, vigorousness, and environmental robustness. However, a workable route remains elusive. Here, we describe a practical strategy, namely *de novo* domestication of wild allotetraploid rice. By screening allotetraploid wild rice inventory, we identified one genotype of *Oryza alta* (CCDD), polyploid rice 1 (PPR1), and established two important resources for its *de novo* domestication: (1) an efficient tissue culture, transformation, and genome editing system and (2) a high-quality genome assembly discriminated into two subgenomes of 12 chromosomes apiece. With these resources, we show that six agronomically important traits could be rapidly improved by editing *O. alta* homologs of the genes controlling these traits in diploid rice. Our results demonstrate the possibility that *de novo* domesticated allotetraploid rice can be developed into a new staple cereal to strengthen world food security.

INTRODUCTION

To meet the world summit on food security, 50% more food is required by 2050, putting enormous pressure on increasing the yield per unit area (Bailey-Serres et al., 2019; Godfray et al., 2010; Long et al., 2015). This challenge becomes even greater owing to rapid climate changes, which results in global warming, extreme high and low temperatures, more droughts and floods, new emergence of pests and diseases, and has already become an immense threat to agriculture and food security (Rosenzweig and Parry, 1994; Wheeler and von Braun, 2013). Environmental adaptation of major crops will involve a huge cost, and there is

a great need to develop new high yield crops that are more resilient to climate variability and disasters (Nelson et al., 2009). Polyploidy, mostly resulting from whole genome duplications or interspecific hybridizations, is a common mode of evolution in flowering plants, and polyploid plants often have significant advantages in terms of biomass, vigorousness, and robust adaptation to environmental changes (Comai, 2005; Fang and Morrell, 2016; Lichman et al., 2020; Van de Peer et al., 2017). Therefore, crop polyploidization may play an important role in next-generation crop improvement aimed at facing food security challenges.

The domestication of cultivated rice from its ancestral progenitor wild diploid rice *Oryza rufipogon* is considered one of the most important developments in human history, and rice is now the staple food feeding more than half of the world's population (Huang et al., 2012). During domestication, various important agronomical traits were selected, including seed shattering, erect plant architecture, panicle shape, awn length, grain size and quality, hull color, and so on (Chen et al., 2019). In the genomic era, diploid rice has become a model monocot species owing to its relatively simple genome and the development of efficient transformation systems (Izawa and Shimamoto, 1996). Agronomically important genes controlling plant height, tiller number, panicle morphology, seed weight, etc. are well-studied and have been utilized to breed new elite varieties by rational design (Qian et al., 2016; Sasaki et al., 2002; Zeng et al., 2017). Even so, new strategies for rice improvement are badly needed. Polyploidization of cultivated diploid rice species has also been proposed, but this presents difficulties mainly because of their restricted genetic backgrounds (Cai et al., 2007; Wu et al., 2014; Xu et al., 2014).

To date, 27 species of the genus *Oryza* have been identified and classified into 11 distinct genome types, comprising six diploids (AA, BB, CC, EE, FF, and GG) and five allotetraploids (BBCC, CCDD, HHJJ, HHKK, and KKLL) (Wing et al., 2018). Of these, the species with CCDD genome from South America have much larger biomass and stronger biotic and abiotic resistance than the cultivated diploid rice (Ammiraju et al., 2010a; Khush, 1997; Prusty et al., 2018). Allotetraploid rice with the CCDD genome originated from a single hybridization event, with the CC genome species (*O. officinalis* or *O. rhizomatis*) serving as the maternal parent and an extinct species with the DD genome type serving as the paternal donor (*O. australiensis* with the EE genome type is the extant closest relative of DD) (Ammiraju et al., 2010b; Bao and Ge, 2004; Lu et al., 2009; Wing et al., 2018). Research on allotetraploid rice has largely lagged behind the well-studied diploid rice species with AA genome because of the lack of a reference genome (Stein et al., 2018) and the difficulty of transformation and regeneration (Zhang et al., 2019).

Domestication of wild plants is believed to have enormous potential for agriculture. The traditional domestication of wild plants into commercially available crops has usually taken hundreds or even thousands of years, but recently developed genome editing technologies make it possible to achieve this in a few generations (DeHaan et al., 2020; Eshed and Lippman, 2019; Lemmon et al., 2018; Li et al., 2018; Van Tassel et al., 2020; Zsögön et al., 2018). Therefore, efficient genome editing systems are badly needed to accelerate the pace of domestication. *Agrobacte-*

rium-mediated transformation systems have been developed efficiently only in *O. sativa* subspecies *japonica* and *indica* (Hiei et al., 1994; Shimamoto et al., 1989); however, the transformation efficiency is still very low in *O. glaberrima* and in wild rice with AA genome (Lacchini et al., 2020; Lin and Zhang, 2005; Mohammed et al., 2019). As for allotetraploid wild rice, their transformation remains a formidable challenge, although some of them could be regenerated at an extremely low rate (Zhang et al., 2019; Shimizu-Sato et al., 2020).

Here, we describe a route for *de novo* domestication and improvement of a wild rice relative—the allotetraploid *O. alta* (CCDD)—by screening the wild allotetraploid rice inventory, establishing an efficient transformation and genome editing system, generating a high-quality annotated assembly of 24 chromosomes that illuminates the hereditary blueprint and evolutionary history of allotetraploid rice genomes, and genome-editing of domestication-related genes and agronomically important genes based on the knowledge of diploid rice. Our work offers a practical strategy for creating new crops to meet future food demand.
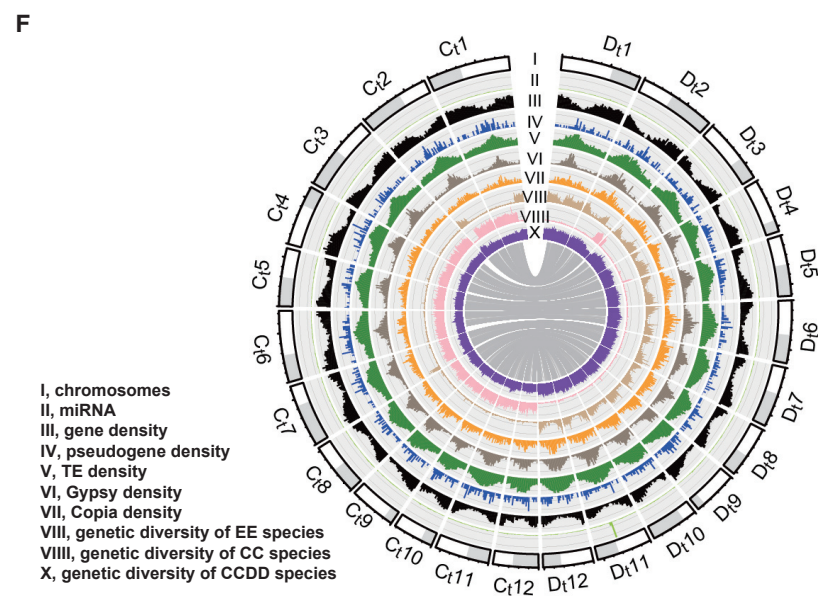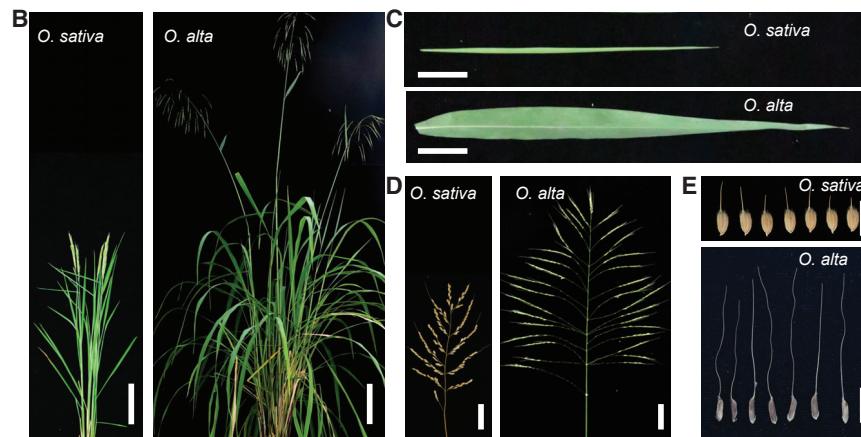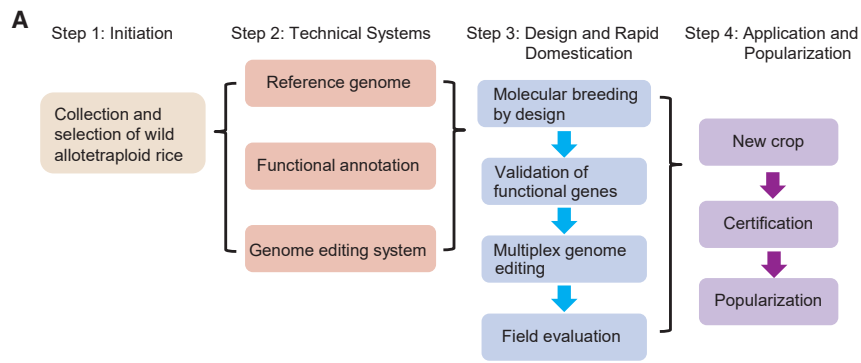
RESULTS

Roadmap to *de novo* domestication of wild allotetraploid rice

To generate polyploid rice crops, we initiated a roadmap strategy, namely a *de novo* domestication of wild allotetraploid rice (Figure 1A). The whole process can be divided into four steps: step 1, selection of an appropriate starting material for *de novo* domestication; step 2, establishment of desirable technical systems including a reference genome, the functional annotation of genes and, importantly, an efficient transformation and genome editing system; step 3, design and rapid introduction by multiplex genome editing of domestication-related genes or alleles, followed by field evaluation; and step 4, certification and popularization of the new crop. However, neither a suitable starting material nor the required technical systems were initially available for wild allotetraploid rice.

Establishment of a transformation system for allotetraploid *O. alta* genotype PPR1

To identify the most promising starting material, we considered that callus induction and regeneration abilities were essential attributes, followed by suitable biomass characteristics and resistance to abiotic and biotic stresses. Therefore, we collected seeds for 28 lines with CCDD genome from germplasm banks, comprising eight *O. alta*, two *O. grandiglumis*, and 18 *O. latifolia* lines, and tested their abilities of callus induction and regeneration using the typical methods for diploid rice (Zhang et al., 2016). We found that some of the lines had modest callus induction ability but most gave a very low rate of regeneration, except for one genotype of *O. alta* (accession ID 2007-24 from National Nursery of Wild Rice Germplasm, Nanning, China), which performed well for both callus induction and regeneration (Figure S1A; Table S1). Focusing on this line, we optimized the protocols and finally achieved a satisfactory transformation system with ~80% transformation efficiency and ~40% regeneration ability (Figures S1B–S1E). We therefore considered this

Figure 1. Roadmap of *de novo* domestication of *O. alta* and its structures and genomic features

(A) Roadmap of *de novo* domestication of wild allotetraploid rice.

(B–E) Whole plants (B), blades (C), panicles (D), and seeds (E) of *O. alta* and *O. sativa*. Scale bars, 20 cm (B); 5 cm (C and D); and 1 cm (E).

(F) Distribution of genomic features of *O. alta*. Black lines connect the collinear blocks of 24 homologous chromosomes of C_t1 to C_t12 and D_t1 to D_t12 subgenomes.

See also Figure S1 and Tables S1, S3, and S4.

length with sparse spikelets (Figure 1D), and small grain size (thousand-grain weight ~8.79 g) with awns >4 cm (Figure 1E). In summary, we selected PPR1 as the starting model material for *de novo* domestication and established an efficient tissue culture and transformation system of it, thus overcoming one of the biggest hurdles in the process of *de novo* domestication.

Assembling a high-quality genome of PPR1 and homeologous subgenomes

The second challenge was to assemble a high-quality genome of PPR1. We took an integrated genome sequencing and assembling approach by generating 121.1 Gb (~123.42X) of single-molecule real-time sequencing data, 287.1 Gb (~292.62X) of BioNano data, 103.2 Gb (~105.18X) of high-throughput chromatin conformation capture (Hi-C) data, and 85.2 Gb (~86.79X) Illumina short reads (Figures S2B and S2C), and achieved a high-quality assembly comprising 24 pseudo-chromosomes with contig N50 of 18.2 Mb, superscaffold N50 of 37.1 Mb (PPR1 V1.0) (Table 1), and a total genome size of 894.6 Mb (876.4 Mb anchored on chromosomes). We then constructed a gene expression atlas by generating 235 Gb of RNA sequencing (RNA-seq) data for ten representative tissues, which yielded an average 92.81% mapping rate to PPR1 V1.0 (Figures S3A and S3B). Of the 482,997 transcripts derived from transcriptome sequencing, 90.16% were detected in PPR1 V1.0 with more than 99% identity (Figure S3C). Finally, 99,312 protein-coding genes, including 81,421 high-confidence (HC) and 17,891 low-confidence genes with an average length of 3,021 bp, were predicted and functionally annotated (Figures S3D and S3E). In addition, we annotated 1,253 pseudogenes and 17,104 non-coding RNAs, including 12,332 microRNAs,

material a model for further research into *de novo* domestication and named it polyploid rice 1 (PPR1). Importantly, PPR1 also had a low rate of heterozygosity (<0.2%) (Figure S2A), which is a positive feature for genome sequencing and assembly. Although it has a substantially larger biomass than cultivated rice, it has typical non-domesticated features, such as height >2.7 m (Figure 1B), broad and long leaves (Figure 1C), panicle >48 cm in

Table 1. The *O. alta* genome assembly and gene annotation

	Total genome	C_t subgenome	D_t subgenome
Estimate of genome size (Mb)	789.9	–	–
Genome assembly (Mb)	894.6	441.3	435.1
Number of superscaffolds	274	12	12
N50 of superscaffolds (Mb)	37.1	–	–
N50 of supercontigs (Mb)	18.2	–	–
N50 of contigs (Mb)	18.2	–	–
N90 of superscaffolds (Mb)	27.8	–	–
N90 of supercontigs (Mb)	8.5	–	–
N90 of contigs (Mb)	7.2	–	–
Longest sequence length (Mb)	50.4	50.4	49.0
GC content (%)	43.9	43.7	44.0
Number of genes	99,312	52,861	46,388
Number of high confident genes	81,421	43,481	37,908
Percentage of gene length in genome/subgenome (%)	33.5	36.4	32.0
Mean gene length (bp)	3,020.5	3,040.5	2,999.6
Gene density (per Mb)	111.0	119.8	106.6
Mean coding sequence length (bp)	910.2	908.8	912.0
Mean exon length (bp)	358.5	359.2	357.7
Mean intron length (bp)	621.4	629.1	612.7
Subgenome specific genes	–	6,283	4,931
Masked repeat sequence length (Mb)	551.5	261.3	272.3
Repeats percentage of genome size (%)	61.7	59.2	62.6

See also Figures S2 and S3 and Table S2.

2,032 rRNAs, 1,147 tRNAs, 207 small nuclear RNAs (snRNAs), 1,057 small nucleolar RNAs (snoRNAs), 10 signal recognition particle RNAs (srpRNAs), and 319 other non-coding RNAs (ncRNAs) (Table S2). Of the predicted protein-coding genes, 99.94% were unevenly distributed over the 24 chromosomes with more genes on chromosome arms/ends (Figure 1F). We found 98.2% complete genes of the Benchmarking Universal Single-Copy Orthologs (BUSCO) gene set (Seppey et al., 2019) were recalled, demonstrating the high accuracy and completeness of the genome assembly.

By aligning 229,886 BAC end sequences and 278 genes from CC and EE genome species (Table S3) (Ammiraju et al., 2006; Zou et al., 2013) to the PPR1 genome, we divided the 24 pseudo-chromosomes into C_t and D_t subgenomes, which contained 52,861 and 46,388 protein-coding genes, respectively (Table 1). Among them, we identified 19,958 homeologous gene pairs, and the transcriptome analysis indicated that 5,732 homeologous gene pairs on the homologous chromosomes showed expression dominance in one gene. Although genome sizes differed noticeably between CC (~651 Mb) and EE (~965 Mb) species (Zuccolo et al., 2007), the C_t and D_t subgenomes of *O. alta* had similar sizes of ~441.3 Mb and ~435.1 Mb (Table 1), which were also highly collinear with 100 syntenic blocks containing 17,639 and 17,527 genes, respectively, on the homologous chromosomes of the C_t and D_t subgenomes (Figure 1F; Table S4). These findings further demonstrate the high quality and coverage of the allotetraploid PPR1 genome sequence.

Genomic features of allotetraploid *O. alta*

To explore gene similarity between *O. alta* and *O. sativa*, we compared the PPR1 HC genes (81,839 genes) and the Nipponbare MSU gene set (55,986 genes) and found that 43,599 (77.87%) genes of Nipponbare had homologs in *O. alta*, including 39,543 (70.41%) genes with homologous genes on both C_t and D_t subgenomes (Figure 2A). Meanwhile, we found that 10,862 and 9,982 genes on the C_t and D_t subgenomes of *O. alta* found no homologs in cultivated diploid rice, possibly because the highly diverged homologs could not be reliably identified. As *O. alta* showed strong biotic and abiotic resistance phenotypes (Khush, 1997; Prusty et al., 2018), we further examined the resistance gene analogs in *O. alta* (Table S5). Although *O. sativa* ssp. *japonica/Geng* Nipponbare, and *indica/Xian* R498 contain much more nucleotide-binding site (NBS) type genes compared to the C_t or D_t subgenome (Figure 2B) due to the expansion of NBS-leucine rich repeat (NBS-LRR) genes in the AA genome after the split of *O. meridionalis* and *O. glumaepatula* (Stein et al., 2018), *O. alta* contains more genes in total for all four types of resistance gene analogs, showing a potential for broader resistant spectrum. Moreover, the number of transcription factors and transcription regulators is similar among the C_t and D_t subgenomes and AA genome (Figures S4A–S4C; Table S5). These results indicate that the *O. alta* genome contains considerable genetic resources which have not been utilized in cultivated rice.

Furthermore, based on protein homologies among *O. alta* and nine other species, a total of 42,642 gene families were identified, with 19,925 and 18,519 in the C_t and D_t subgenomes, respectively (Figure 2C). Of these families, 14,125 were shared among C_t, D_t, and *O. sativa* and 525 C_t-specific and 363 D_t-specific. Phylogenetic analysis based on single-copy gene families showed that the *O. alta* C_t subgenome and *O. sativa* genome were the nearest neighbors, confirming the previous hypothesis that the AA genome shared a more recent common ancestor with the CC genome than with the DD/EE genome (Wing et al., 2018). The estimated divergence times between the AA and CC genomes, and between the CC and DD genomes, were 4.57 and 5.45 million years ago (MYA), respectively (Wing et al., 2018). Both the C_t and D_t subgenomes of allotetraploid
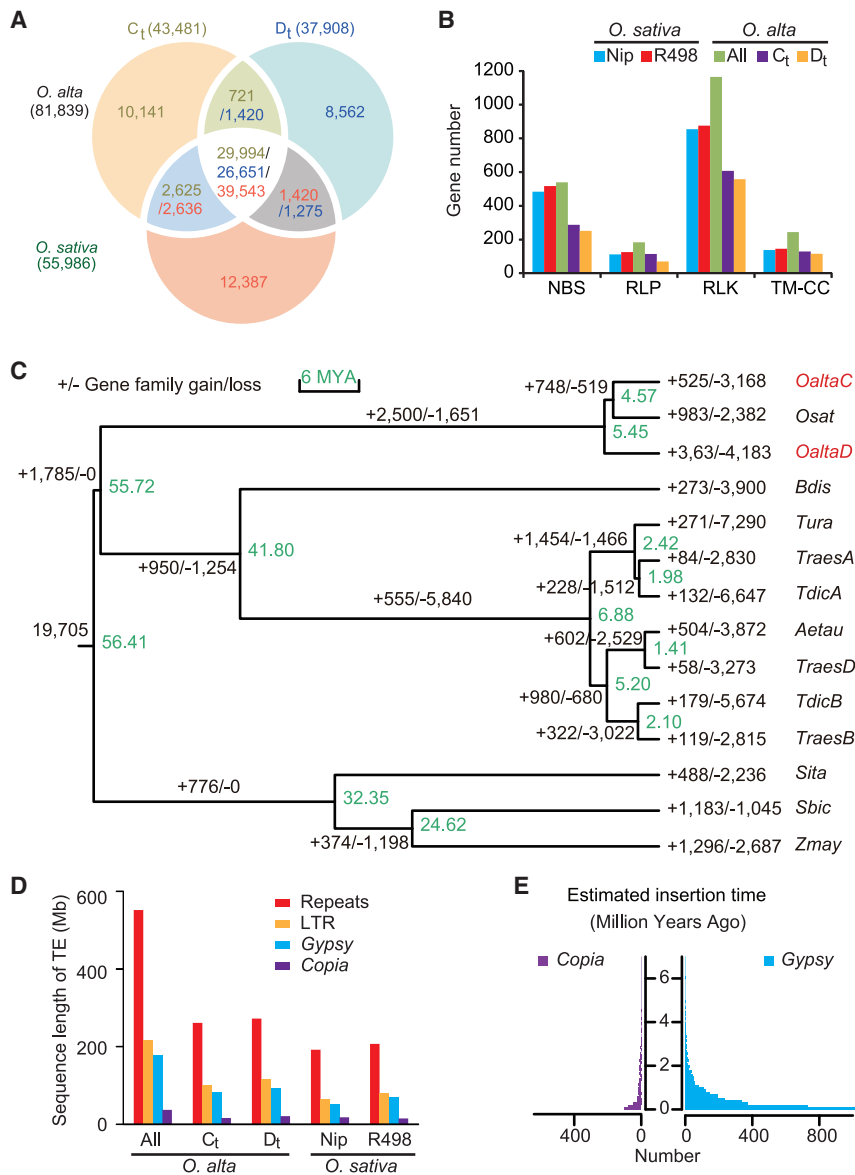
Figure 2. Genomic features of allotetraploid *O. alta*

(A) Count of homologs between *O. sativa* and *O. alta* high confident (HC) genes.

(B) Resistant gene analogs in *O. alta*, *O. sativa* ssp. *indica* (R498), and *japonica* (Nip) genomes. NBS, nucleotide-binding site; RLK, receptor-like kinase; RLP, receptor-like protein; TM-CC, transmembrane coiled-coil.

(C) Phylogenetic tree of ten species based on orthologs of single-gene families. The values above each branch represent the number of gene family gains/losses, and the green number to the right of each node represents divergent time (MYA) from the common ancestor. *Aetau*, *Aegilops tauschii*; *Bdis*, *Brachypodium distachyon*; *OaltaC* and *OaltaD*, *Oryza alta* C_t and D_t subgenomes; *Osat*, *Oryza sativa*; *Sbic*, *Sorghum bicolor*; *Sita*, *Setaria italic*; *TdicA* and *TdicB*, *Triticum dicoccoides* A and B subgenomes; *TraesA*, *TraesB*, and *TraesD*, *Triticum aestivum* A, B, and D subgenomes; *Tura*, *Triticum urartu*; and *Zmay*, *Zea mays*.

(D) Length of repeats, LTR, *Copia* and *Gypsy* among *O. alta*, and *O. sativa* ssp. *indica* (R498) and *japonica* (Nip) genomes.

(E) Estimated insertion times for *Copia* and *Gypsy* LTR retrotransposons.

See also Figure S4 and Tables S2 and S5.

higher than 59.2% and 18.7% of the C_t subgenome (Figure 2D). PIF-Harbinger transposons were enriched within the 2-kb upstream and downstream of coding regions (Figure S2D), indicating their possible roles in the evolution of cis-elements. These results suggest that the amplification of *Gypsy*-type LTRs is the major driving force for the expansion of the *O. alta* genome.

Genetic diversity of allotetraploid rice with CCDD genome

We resequenced a total of 44 wild rice accessions, including 28 CCDD, six EE and ten CC accessions, generating a total of 922.45 Gb of sequence data (Table S1) for population genomic analysis. Sequence reads for each accession were aligned to the PPR1 V1.0 with an average mapping rate of 98.47%, 88.20% and 57.17% for CCDD, CC and EE species, respectively (Table S1). We identified 30.77 million high-quality variants consisting of 21,281,771 SNPs and 9,490,896 indels with an average of 23.8 SNPs and 10.6 indels per kilobase. A total of 7,373,962 SNPs and 3,075,405 indels were located in genic regions, including 1,183,604 nonsynonymous and 288,439 frameshift variants. 70,779 variants had potentially large effects, including 54,340 SNPs and 16,439 indels that cause the introduction of start codons, premature stop codons, and longer-than-usual transcripts. These data indicate that substantial variations exist within the population of CCDD accessions, as well as between C_t and C, and between D_t and E genomes.

O. alta were considerably larger than those of Nipponbare and R498. The analysis showed that 61.7% (551.5 Mb) of the *O. alta* genome consisted of repetitive sequences (Table 1), a much higher proportion than the 40.43% in Nipponbare and 42.05% in R498 (Du et al., 2017; International Rice Genome Sequencing Project, 2005). The major difference was in long terminal repeats (LTRs), which are the predominant repeats accounting for 24.1% (215.93 Mb) of the *O. alta* genome, markedly higher than the 17.55% in Nipponbare and 20.55% in R498 (Figure 2D). *Gypsy*-type LTRs were the most abundant transposable element subfamilies, covering 19.9% (177.60 Mb) of the *O. alta* genome (Figure 2D; Table S2). However, *Gypsy* and *Copia*-type LTRs were estimated to have a similar insertion burst time less than one MYA (Figure 2E). Repeats and *Gypsy*-type LTRs covered 62.6% and 21.6% of the D_t subgenome, respectively,
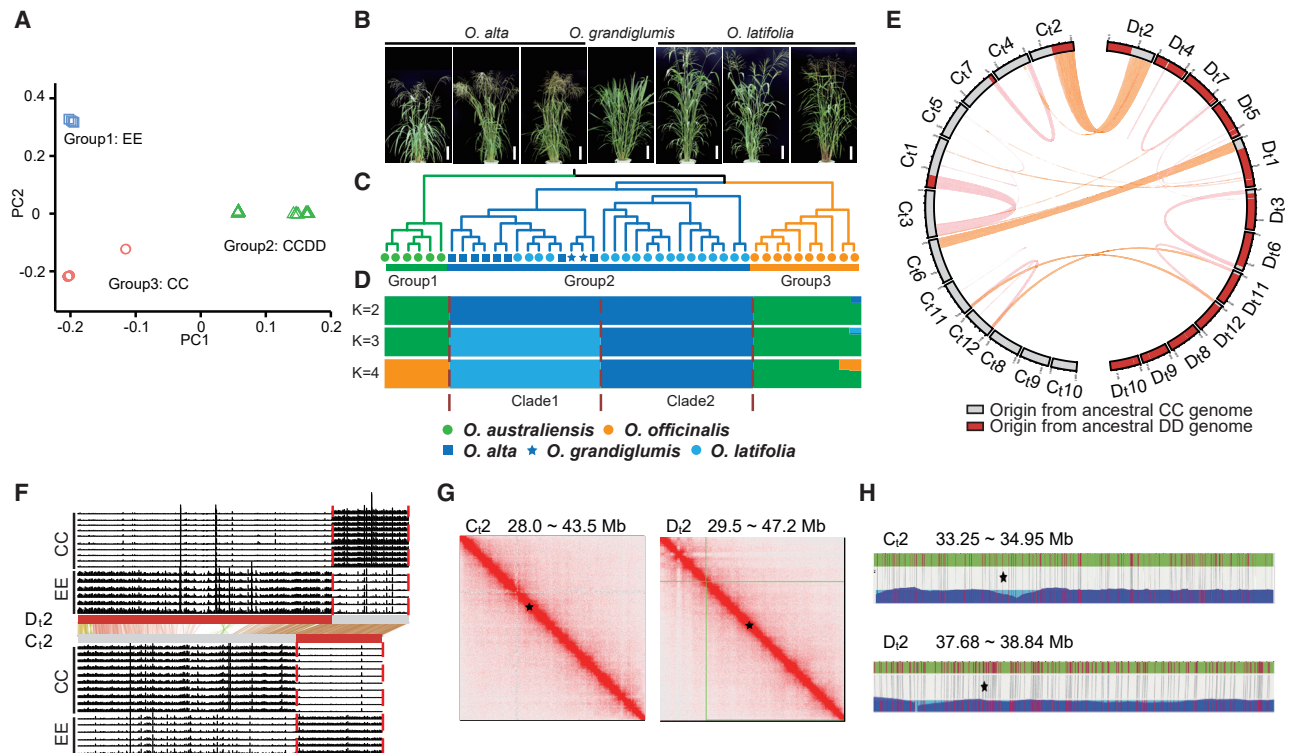
Figure 3. Genetic variation map of 44 CC, EE, and CCDD rice accessions and chromosomal rearrangements in the *O. alta* genome

(A) Principal component analysis (PCA) of 44 rice accessions. Different colors indicate accessions belonging to different groups: blue (group 1), diploid EE species; green (group 2), tetraploid CCDD species; red (group 3), diploid CC species.

(B) Representative mature plants of three CCDD *Oryza* species. Scale bars, 30 cm.

(C) Phylogenetic relationship of 44 rice accessions based on genetic distance.

(D) Structure plot for rice accessions with different numbers of clusters (K = 2, 3, and 4).

(E) Homologous exchanges and collinear blocks among non-homologous chromosomes between C_t and D_t subgenomes. Pink lines connect collinear blocks within C_t or D_t subgenomes, and orange lines connect collinear blocks between C_t and D_t subgenomes.

(F) Genome rearrangements between *O. alta* chromosome C_t2 and D_t2. Grey and red boxes represent the regions origin from ancestral CC and DD genome, respectively. Back lines represent the sequence depth of the resequencing data from ten CC and six EE diploid rice accessions. Color lines between two chromosomes represent collinear blocks between *O. alta* chromosomes C_t2 and D_t2.

(G) Chromatin interactions at 150 kb resolution by Hi-C data. Black pentagrams show the borders of the large segmental translocation between *O. alta* chromosomes C_t2 and D_t2.

(H) Collinearity between genome sequences by BioNano maps. Upper horizontal green bars represent the genome sequences and the below blue bars represent the BioNano maps. The lines between green bars and blue bars represent the collinear restriction sites. Black pentagrams in (G) and (H) show borders of the large segmental translocation between *O. alta* chromosomes C_t2 and D_t2.

See also Figure S5 and Tables S1 and S4.

Principal component analysis (PCA) showed that the 44 rice accessions were clearly clustered into three groups (Figure 3A), namely group 1 from the EE accessions, group 2 from the CCDD accessions, and group 3 from the CC accessions (Figure 3B). Neighbor-joining tree and model-based clustering analysis further divided group 2 into two clusters (Figures 3C and 3D), with all eight *O. alta*, two *O. grandiglumis*, and four *O. latifolia* accessions in cluster 1 and the other 13 *O. latifolia* accessions residing in cluster 2. Nucleotide diversity (π) in CCDD accessions was estimated to be 0.0049, which is lower than other polyploidy crops such as potato (0.0111) (Hardigan et al., 2017), but higher than in diploid rice (0.0024 in *O. sativa* and 0.0030 in *O. rufipogon*) (Huang et al., 2010), indicating that this population of allotetraploid rice can provide diverse genetic resources for future functional genomic study and breeding.

Chromosomal rearrangements in allotetraploid *O. alta*

By aligning the resequencing data of the diploid CC and EE *Oryza* species to the PPR1 genome, we were able to identify which parts of the current *O. alta* genome are derived from the ancestral CC or DD genomes. This revealed several translocations of large segments between the two subgenomes after tetraploidization (Figure 3E; Table S4). Chromosomes C_t2 and D_t2 were highly syntenic, but the aligning pattern of the reads from CC and EE genomes to PPR1 clearly showed the occurrence of a large homeologous exchange between them (Figure 3F). The continuous interaction signals on the Hi-C heatmap and the consistency with BioNano maps in these regions indicated that this was not due to assembly errors (Figures 3G and 3H).

In addition to the homeologous exchange, we found three segments on C_t7 matching to EE genome reads were collinear with

C_t4, suggesting that these segments were translocated from the originally ancestral chromosome D_t4 (Figures 3E and S5A). Again the correctness of the assembly was confirmed by Hi-C heatmap and BioNano data (Figures S5B and S5C). Similarly, we found several other segments, including one on D_t4, one on D_t3, and one on C_t1, were translocated from the ancestral chromosomes C_t7, C_t6, and D_t3, respectively. Additionally, we also identified 26 syntenic blocks located on non-homeologous chromosomes, apparently caused by ten intra-subgenome translocations, four intra-chromosomal duplications, and 12 inter-chromosomal duplications, respectively (Figure 3E; Table S4). C_t11, C_t12, and D_t11, D_t12 likely experienced multiple duplications and translocations (Figure 3E), but the other six chromosomes have experienced fewer chromosomal rearrangements. According to the matching reads of CC and EE species on each gene, we further identified all the genes originating from ancestral CC genome (ancestral CC genes) and ancestral DD genes in the PPR1 genome, and we found that 5,672 genes (13.04%) that were now in the C_t subgenome were of DD origin, while 5,132 genes (13.54%) on the D_t subgenome were CC ancestral. Meanwhile, we found that a large number of genes had been lost since tetraploidization; 5,516 ancestral CC genes had no ancestral DD meolgs in the current *O. alta* genome, and 3,954 ancestral DD genes had no ancestral CC homeologs. All these data suggest the highly dynamic evolution of the *O. alta* genome.

A route for rapid domestication of PPR1 by genome editing

During the domestication of Asian cultivated rice, traits beneficial for the farming rather than for natural growth, such as erect growth habit, hull color, shattering, awn, pericarp color, and grain size, were selected and improved, and their genetic mechanisms have been extensively studied (Chen et al., 2019). Wild rice *O. alta* shares some traits similar to those of the wild ancestors of the modern cultivars. To rapidly domesticate these traits, we first identified the *O. alta* homologs of the domestication-related genes in diploid rice (Figure 4A), and found that seven of ten important domestication-related genes (Chen et al., 2019), including homologs of *qSH1* for shattering, *An-1* and *An-2* for awn length, *Bh4* for hull color, *Rc* for pericarp color, *OsLG1* for panicle shape, and *GW5* for grain width, were present in the *O. alta* genome with >84% identity covering at least 87.5% of *O. sativa* proteins. *PROG1*, a key gene for the erect growth of cultivated rice appeared to have only low level homologs in the subgenomes of *O. alta*, possibly because *O. alta* already has an erect growth habit. These domestication-related genes have been exposed to a variety of evolutionary processes. *Sh4* experienced inter-subgenome translocation from the D_t to the C_t subgenome, leading to the shortening of its coding regions. Both of the homologs of *GAD1/RAE2* and *OsLG1* are present in the C_t subgenome, owing to the translocation of a segment from the D_t subgenome to the C_t subgenome.

As a proof-of-concept to show that modification of these homologs could truly achieve rapid domestication of wild *O. alta*, we used a CRISPR/Cas9-mutagenesis approach to edit the *qSH1* homologs for shattering and *An-1* homologs for awn length (Konishi et al., 2006; Luo et al., 2013). We first designed a single guide RNA (sgRNA) targeting the first exon

of both *OaqSH1-CC* (OalC01g168290) and *OaqSH1-DD* (OalD01g114050) and identified one mutant, $qsh1^{CR}$-1, which harbored biallelic frameshift mutations for *OaqSH1-CC* and heterozygous frameshift mutations for *OaqSH1-DD* (Figure 4B; Table S6). We found that wild-type PPR1 had a complete layer of abscission cells between the rice grain and the pedicel, forming a longitudinal continuous line similar to wild diploid rice, whereas in $qsh1^{CR}$-1, the line of abscission cells was absent, indicating that the editing of the *qSH1* homologs in *O. alta* can prevent seed shattering (Figure 4C). We further designed one sgRNA that targets the first exons of both *OaAn-1-CC* (OalC04g136090) and *OaAn-1-DD* (OalD04g130280) and another sgRNA that targets the first exon of *OaAn-1-DD*, and obtained two mutants of $an-1^{CR}$-1 and $an-1^{CR}$-2 harboring different homozygous frameshift mutations in both *OaAn1-CC* and *OaAn1-DD* (Figure 4D; Table S6). We found that the average awn length of $an-1^{CR}$-1 (1.61 cm) and $an-1^{CR}$-2 (2.63 cm) were both significantly shorter than that of PPR1 (4.11 cm) (Figures 4E and 4F). Taken together, these results provide a solid basis for the rapid domestication of *O. alta*.

Variable sequences of *O. alta* homologous to agronomically important *O. sativa* genes

In addition to the domestication-related genes, we also identified *O. alta* homologs of 113 agronomically important genes in modern diploid rice cultivars, which determine important traits such as grain yield, grain quality, fertility, heading date, biotic and abiotic resistance, and nutrient-use efficiency (Figure 5A; Table S7). We found that the genes affecting different features had different similarity levels. The genes affecting sterility in *O. sativa* were poorly conserved in *O. alta*, because the best-matched homologs of *qHMS7*, *S5*, *SaF*, *Sc-j*, and *S1TPR* had similarities less than 40%, and there were no detectable homologs of *SaM*, *WA352*, and *pms3* (Figure 5A; Table S7). The genes associated with biotic stress were also relatively poorly conserved, possibly because diseases and pests are different in Asia and America. In contrast, all genes associated with abiotic stress were highly conserved except for *COLD1* (Ma et al., 2015b) and *CAL1* (Luo et al., 2018), which are responsible for cold resistance and cadmium accumulation, respectively. Fortunately, we found that most of the genes related to yield, grain quality, and efficient use of nutrients were fairly conserved between the cultivated diploid and the tetraploid wild rice. Based on these findings, we further tested whether these homologs of modern breeding genes could be altered to improve wild allotetraploid rice.

The "Green Revolution" gene *sd1* is one of the most important genes in modern rice breeding, because its mutation resulted in a shortened culm with improved lodging resistance and a greater harvest index (Sasaki et al., 2002). We therefore designed two sgRNAs targeting the first exons of both *OaSD1-CC* (OalC01g172060) and *OaSD1-DD* (OalD01g109880), and obtained an $sd1^{CR}$-1 mutant that harbored biallelic frameshift mutations in both *OaSD1-CC* and *OaSD1-DD* (Figure 5B; Table S6). We found that its height was dramatically reduced compared with the wild-type PPR1 (Figure 5C). In contrast to *O. sativa*, which usually has 4–5 elongated internodes, *O. alta* has 10–12 elongated internodes. We carried out a detailed comparison of the lengths of the first eight internodes from the base
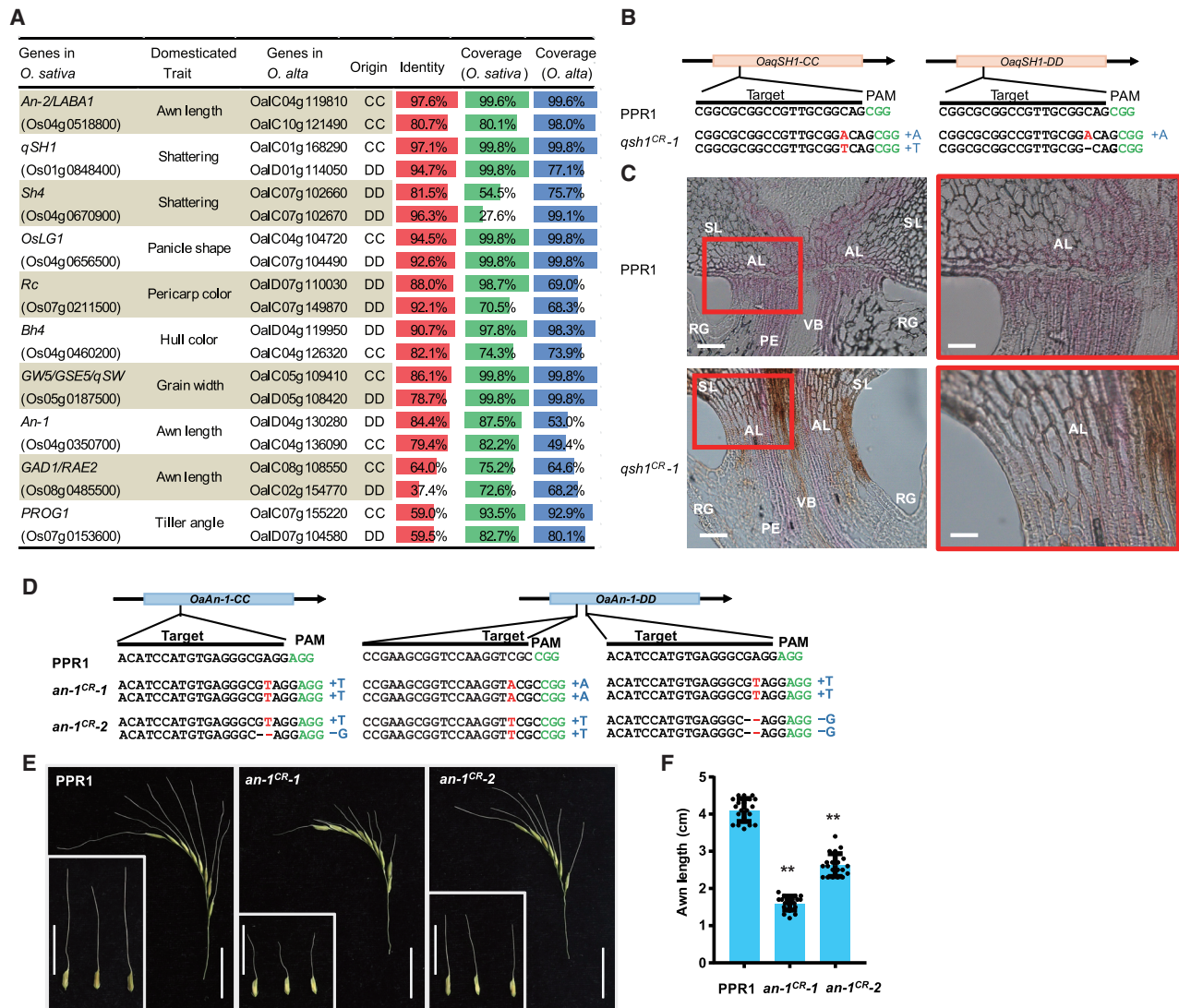
Figure 4. Rapid domestication of PPR1 by genome editing

(A) The *O. alta* homologs of 10 domestication-related genes of *O. sativa*. Coverage, the proportion of homologous alignment sequences to total protein length.

(B) CRISPR/Cas9-induced mutations in *OaqSH1-CC* and *OaqSH1-DD* in T_0 plant.

(C) Lignin of PPR1 and *qsh1^CR-1* stained with phloroglucinol. Right panels show enlarged images of boxed areas in left panels. Scale bars, 200 μm (left panels); 80 μm (right panels). AL, abscission layer; PE, pedicel; RG, rudimentary glume; SL, sterile lemma; VB, vascular bundle.

(D) CRISPR/Cas9-induced mutations in *OaAn-1-CC* and *OaAn-1-DD* in T_0 plants.

(E) Comparison of awn lengths among PPR1, *an-1^CR-1* and *an-1^CR-2*. Scale bars, 2 cm.

(F) Statistical analysis of the data in (E). Data are means ± SD (n = 20 awns). Asterisks represent significant differences between mutant and PPR1 determined by Student's t test. **p < 0.01.

See also in Tables S4 and S6.

of the stem and found that the length of the first and fifth internodes was relatively unchanged in *sd1^CR-1*, whereas all six other internodes were significantly shortened (Figures 5D and 5E). Similar results were obtained in the other five independently edited lines (Figures S6A, S6C, and S6D; Table S6). Furthermore, we also designed an sgRNA targeting the homologs of *GS3*, a key gene controlling seed morphology in diploid rice (Mao et al., 2010) and obtained six independent mutants, in which the grain lengths were significantly increased compared with

PPR1 (Figures 5F, 5G, S6B, S6E, and S6F; Table S6). Taken together, these results demonstrated that the knowledge gained in diploid rice can provide valuable information and directly facilitate the improvement of wild allotetraploid rice.

Improving PPR1 with base substitution editor and multiplex editing technologies

Ideal Plant Architecture 1 (IPA1) is considered as one of the new green revolution genes, because a point mutation in the gene

Figure 5. Sequence similarities and genome editing of O. alta homologs of agronomically important O. sativa genes

(A) Similarities of 113 agronomically important genes in O. sativa to their homologs in O. alta.

(B) CRISPR/Cas9-induced mutations in OaSD1-CC and OaSD1-DD in T_0 plants.

(C) Comparison of heights of PPR1 and $sd1^{CR}$-1 plants. Scale bars, 30 cm.

(D) Comparison of internode lengths between PPR1 and $sd1^{CR}$-1. Scale bars, 30 cm.

(E) Statistical analysis of the data in (D). Data are means ± SD (n = 10 tillers). Asterisks represent significant differences of each internode determined by Student's t test. **$p < 0.01$; *$p < 0.05$; ns, not significant.

(F) CRISPR/Cas9-induced mutations in OaGS3-CC and OaGS3-DD in T_0 plants.

(G) Comparison of grain length between PPR1 and $gs3^{CR}$-1 plants. Scale bars, 1 cm.

(H) Statistical analysis of the data in (G). Data are means ± SD (n = 20 grains). Asterisks represent significant differences determined by Student's t test. **$p < 0.01$.

See also Figure S6 and Tables S6 and S7.

perturbs OsmiR156 target sites to regulate its expression level, leading to an increased stem diameter, improved lodging resistance, large panicles, and dramatically enhanced grain yield (Jiao et al., 2010; Miura et al., 2010). We found that the sequence of OamiR156-DD was identical to OsmiR156, which matches its target site on OaIPA1-DD (OalD08g132520), whereas OamiR156-CC had a single base change and matches OaIPA1-CC (OalC08g106170) (Figures 6A and 6B). We therefore tested whether a base editor optimized for diploid rice could generate gain-of-function base substitution mutations in O. alta (Zong et al., 2018) and designed an sgRNA targeting the miRNA target sites in both OaIPA1-CC and OaIPA1-DD. We obtained a mutant, $ipa1^{CR}$-1, which contains one point mutation in OaIPA1-DD without affecting OaIPA1-CC (Figure 6B; Table S6). In-detail examination of this mutant revealed that the stem

diameter of $ipa1^{CR}$-1 was significantly greater than that of PPR1 (Figures 6C–6E). We further examined the transcriptional changes of OaIPA1-CC and OaIPA1-DD and found that the expression levels of OaIPA1-DD were 10-fold greater in the shoot bases of $ipa1^{CR}$-1, whereas OaIPA1-CC expression was unchanged (Figure 6F). Therefore, the use of base substitution editors can expand the strategies for gene improvement in O. alta.

Rice heading dates, unlike other traits that usually have changes in one direction, become more variable during breeding for expanded habitats, different cropping systems, and different day lengths. In diploid rice, varieties with weak alleles of Ghd7 can be grown for grain production in temperate zones (Weng et al., 2014; Xue et al., 2008), whereas the varieties with low DTH7 activities are less sensitive to increased

Figure 6. Improvement of PPR1 with base editor and multiplex editing systems

(A) Sequence alignment of OsmiRNA156, OamiRNA156-CC and OamiRNA156-DD.

(B) Base editor-induced point mutation in *OaIPA1-DD* in T_0 plant.

(C and D) Comparison of third internodes from the stem bases (C) and their cross sections (D) between PPR1 and *ipa1^CR^-1*. Scale bars, 5 cm (C); 0.5 cm (D).

(E) Statistical analysis of stem thickness of the third internodes of PPR1 and *ipa1^CR^-1*. Data are means ± SD (n = 10 tillers). Asterisks represent significant difference determined by Student's t test. **p < 0.01.

(F) Relative expression levels of *OaIPA1-CC* and *OaIPA1-DD* to *OaUBI* in the shoot bases of PPR1 and *ipa1^CR^-1*. Data are means ± SD (n = 4 technical replicates). Asterisks represent significant difference between *ipa1^CR^-1* and PPR1 determined by Student's t test. **p < 0.01; ns, not significant.

(G) Genotypes of *OaGhd7-CC*, *OaGhd7-DD*, *OaDTH7-CC*, and *OaDTH7-DD* and plant morphologies of eight edited T_0 plants at 110 days after transplantation in Beijing. Red triangles indicate booting young panicles. Scale bars, 30 cm.

(H) Structure of the multiplex genome editing system with eight sgRNAs targeting 16 genes in PPR1. NLS, nuclear localization sequence; Pubi, promoter of maize ubiquitin; Tnos, terminator of nopaline synthase gene.

See also Figure S7 and Table S6.

day length to flower earlier in high geographical regions (Gao et al., 2014; Yan et al., 2013). PPR1 originates from tropical regions of South America, displays strong long-day sensitivity, and fails to flower in Beijing (40°N). We used a multiplex CRISPR/Cas9 editing approach with three sgRNAs in one vector to edit *OaGhd7-CC* (OaIC07g145340), *OaGhd7-DD* (OaID07g113810), *OaDTH7-CC* (OaID04g100060, a CC ances-

tral gene locating on chromosome D_t4), and *OaDTH7-DD* (OaID07g145130), and generated eight T_0 independently transformed PPR1 lines. We found that the eight lines all contained different mutant alleles (Figure 6G; Table S6), and all four genes had been edited in six of the lines, indicating that this technique is highly effective in creating variants in PPR1. In consistence with their genotypes, heading dates varied extensively in these

edited lines. Line 1 with all four genes mutated had a dramatically shortened heading date of 82 days after transplantation in Beijing, whereas wild-type *PPR1* failed to flower after 150 days. Line 2 and line 3 headed 10 days later after line 1, and lines 4 to 8 gradually headed after 103 to 130 days. The seeds of lines 1 to 7 could be filled, while line 8 seeds failed, suggesting that lines 1 to 3 could even be planted in far northern regions. More importantly, we also developed a multiplex editing system that can target eight or sixteen genes each with a single vector with fairly high efficiency (Figures 6H and S7; Table S6). Taken together, these results showed that it is rational in theory and technology to achieve rapid *de novo* domestication and improvement of wild allotetraploid rice into a staple food crop in near future.

DISCUSSION

Polyploidization, one of the most important evolutionary events in plants, can increase genetic diversity, introduce new genetic combinations, foster adaptation to new environments, and create vigorousness effects. A great number of economically important crops are alloploids; these include food crops like wheat and oat, industrial crops like tobacco, cotton, and sugarcane, and fruit crops like strawberry, but most of them are natural allopolyploid species. Although the advantages of allopolyploidy are widely recognized, artificially induced alloploid crops are very few and usually result from a long-term breeding program to screen out frequently encountered undesirable features (e.g., triticale, an alloploid crop derived from a cross between wheat and rye). Instead of inducing alloploidy in the modern crops, we aimed at *de novo* domestication of a wild alloploid plant using genome editing. Starting from the selected genotype PPR1 of the wild allotetraploid rice *O. alta*, we established an efficient tissue culture and genome editing system, and generated a high-quality genome assembly. By taking advantage of the well-studied domestication-related genes in cultivated diploid rice, we identified their homologs in the *O. alta* genome, providing essential information for editing its potentially important agronomic genes. We then applied the CRISPR/Cas9, base editing, and multiplex editing technologies to improve PPR1 by editing a series of domestication-related and/or agronomically important genes in diploid rice, in which we obtained edited lines with desired characteristics including seed shattering, awn length, plant height, grain size, stem thickness, and heading date. The present work provides an effective solution to the technical difficulties in achieving rapid *de novo* domestication of wild allotetraploid rice. We present the use of wild germplasm, complete reference genomes, improved transformation protocol, and diverse genome editing tools to enable the creation of new crop species.

Hereafter, enormous efforts and investments are still needed to achieve the ultimate success in creating a new crop cultivar deployed in farmers' field. First, although the barrier of transformation and genome editing has been overcome in wild allotetraploid rice, many alleles require weak effects rather than complete knockouts to generate the most desirable traits. Toward this, high efficient genome editing tools in generating various kinds of modification are still urgently needed. Second, the genes implicated usually have pleotropic effects, and the outcome of

different combination of edited alleles is difficult to predict. Therefore, there need more genetic diversities to enable a breeding program that optimizes the engineered alleles. Third, wild rice has many advantages over cultivated rice for traits such as biotic and abiotic stress resistance, which are key factors toward combating the impending climate change challenges (Chen et al., 2019; Khush, 1997; Prusty et al., 2018; Song et al., 1995). Thus, there is huge interest in both characterizing these traits in wild rice (Huang et al., 2008; Stein et al., 2018; Zhou et al., 2011) and exploring the capability of hybridization with wild rice guided by molecular markers (Atwell et al., 2014; Menguer et al., 2017). Unfortunately, gene cloning in wild rice is much more difficult than in cultivated rice due to not only the lack of genomic information and transformation systems, but also difficulties in constructing new genetic populations and reflecting the complexity of these desirable traits. Therefore, classical breeding will be needed to maintain these traits without a full understanding of their mechanisms. Fourth, because the wild allotetraploid rice population has a higher genetic diversity, they offer natural genetic resources to be exploited in future functional genomic studies and serve as hybridization in new breeding programs.

Over the past decades, significant progress has been made in rice functional genomics studies with the cloning of many key genes in cultivated rice (Chen et al., 2019; Guo et al., 2019; Jiao et al., 2010; Kim and Buell, 2015; Li et al., 2003b; Miura et al., 2010; Wang et al., 2018a, 2018b). This has greatly benefited rice breeding by allowing the important alleles from different cultivated varieties to be combined by hybridization. One successful example of rice is the development of new elite rice varieties of both high yield and superior quality (Tian et al., 2009; Zeng et al., 2017). At present, three diploid (AA, BB, and FF) rice genomes have been assembled (Chen et al., 2013; Du et al., 2017; Goff et al., 2002; International Rice Genome Sequencing Project, 2005; Stein et al., 2018; Yu et al., 2002; Zhang et al., 2014; Zhao et al., 2018), but only AA species have high-quality pseudo-chromosome level genomes (Wing et al., 2018; Zhou et al., 2020). Generating high-quality genome assemblies for polyploid genomes represents a challenge, especially in discriminating between homeologous subgenomes (Kyriakidou et al., 2018). Many commercially important allopolyploid and autopolyploid crops have been sequenced and assembled, such as wheat (Brenchley et al., 2012; International Wheat Genome Sequencing Consortium, 2018), oilseed rape (Chalhoub et al., 2014), cotton (Li et al., 2015; Zhang et al., 2015; Huang et al., 2020), sweet potato (Yang et al., 2017), sugarcane (Zhang et al., 2018), and teff (VanBuren et al., 2020) with contig N50 ranging from ~480 bp to ~5 Mb. From our experiences in this study, excellent assemblers, sufficient extra-long-read sequencing data, a combination of different sequencing strategies, and selecting the good material with a low heterozygosity rate are critical for assembled polyploid genome. The high-quality *O. alta* genome will provide a powerful tool for molecular genetic studies of tetraploid rice, the evolution of *Oryza* species, and the molecular mechanisms underlying the polyploidization of higher plants.

In conclusion, the strategy of the *de novo* domestication of *O. alta*, a wild allotetraploid rice, demonstrates a clear path for

creating novel crops in the future by combining the advantages of (1) polyploidy, (2) functional genomics knowledge of cultivated crops, (3) desirable attributes of wild species, and (4) rapid genetic modification(s) through genome editing.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Assembly of the PPR1 genome
  - Genome annotation
  - Gene function annotation
  - Annotation of repetitive sequences
  - Annotation of non-coding RNAs
  - Annotation of resistance genes, transcription factors and transcription regulators
  - LTR-RT identification and insertion time estimation
  - Transcriptome data analysis
  - Gene family analysis
  - Population genetics analysis
  - Collinear analysis
  - Homologous genes of diploid rice in *O. alta*
  - Transgenic system of *O. alta*
  - CRISPR/Cas9 based genome editing of *O. alta*
  - RNA extraction and quantitative real-time PCR analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

Supplemental information
Supplemental information can be found online at https://doi.org/10.1016/j.cell.2021.01.013.

AUTHOR CONTRIBUTIONS

H.Y., T.L., X.M., H.D., J. Zhang, L.K., X. Li, Q.G., M.C., G.L., Y.J., and Yan Liang performed the experiments. H.Y., T.L., X.M., H.D., J. Zhang, Z.T., C.C., J. Zuo, Y.W., C.G., Q.Q., B.H., C.L., and J.L. analyzed the data. Z.F., Yuntao Liang, Z.C., M.C., X. Liu, and J.W. collected and provided germplasm. H.Y., C.G., C.L., and J.L. designed the experiments. J.L. conceived the project. H.Y., T.L., and J.L. wrote the manuscript with input from C.G., C.L., J. Zhang, J. Zuo, X.M., A.Z., and R.W. All authors have read, edited, and approved the content of the manuscript.

REFERENCES

Allen, G.C., Flores-Vergara, M.A., Krasynanski, S., Kumar, S., and Thompson, W.F. (2006). A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. Nat. Protoc. *1*, 2320–2325.

Ammiraju, J.S., Luo, M., Goicoechea, J.L., Wang, W., Kudrna, D., Mueller, C., Talag, J., Kim, H., Sisneros, N.B., Blackmon, B., et al. (2006). The *Oryza* bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus *Oryza*. Genome Res. *16*, 140–147.

Ammiraju, J.S., Fan, C., Yu, Y., Song, X., Cranston, K.A., Pontaroli, A.C., Lu, F., Sanyal, A., Jiang, N., Rambo, T., et al. (2010a). Spatio-temporal patterns of genome evolution in allotetraploid species of the genus *Oryza*. Plant J. *63*, 430–442.

Ammiraju, J.S., Song, X., Luo, M., Sisneros, N., Angelova, A., Kudrna, D., Kim, H., Yu, Y., Goicoechea, J.L., and Lorieux, M. (2010b). The *Oryza* BAC resource: a genus-wide and genome scale tool for exploring rice genome evolution and leveraging useful genetic diversity from wild relatives. Breed. Sci. *60*, 536–543.

International Wheat Genome Sequencing Consortium (IWGSC) (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science *361*, eaar7191.

Atwell, B.J., Wang, H., and Scafaro, A.P. (2014). Could abiotic stress tolerance in wild relatives of rice be used to improve *Oryza sativa*? Plant Sci. *215-216*, 48–58.

Bailey-Serres, J., Parker, J.E., Ainsworth, E.A., Oldroyd, G.E.D., and Schroeder, J.I. (2019). Genetic strategies for improving crop yields. Nature *575*, 109–118.

Bao, Z., and Eddy, S.R. (2002). Automated *de novo* identification of repeat sequence families in sequenced genomes. Genome Res. *12*, 1269–1276.

Bao, Y., and Ge, S. (2004). Origin and phylogeny of *Oryza* species with the CD genome based on multiple-gene sequence data. Plant Syst. Evol. *249*, 55–66.

Bennetzen, J.L., Schmutz, J., Wang, H., Percifield, R., Hawkins, J., Pontaroli, A.C., Estep, M., Feng, L., Vaughn, J.N., Grimwood, J., et al. (2012). Reference genome sequence of the model plant *Setaria*. Nat. Biotechnol. *30*, 555–561.

Birney, E., and Durbin, R. (2000). Using GeneWise in the *Drosophila* annotation experiment. Genome Res. *10*, 547–548.

Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G.L., D'Amore, R., Allen, A.M., McKenzie, N., Kramer, M., Kerhornou, A., Bolser, D., et al. (2012). Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature *491*, 705–710.

Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. *268*, 78–94.

Cai, D., Chen, J., Chen, D., Dai, B., Zhang, W., Song, Z., Yang, Z., Du, C., Tang, Z., He, Y., et al. (2007). The breeding of two polyploid rice lines with the characteristic of polyploid meiosis stability. Sci. China C Life Sci. *50*, 356–366.

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

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I.A., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., et al. (2014). Plant genetics. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science *345*, 950–953.

Chen, J., Huang, Q., Gao, D., Wang, J., Lang, Y., Liu, T., Li, B., Bai, Z., Luis Goicoechea, J., Liang, C., et al. (2013). Whole-genome sequencing of *Oryza*

brachyantha reveals mechanisms underlying *Oryza* genome evolution. Nat. Commun. *4*, 1595.

Chen, E., Huang, X., Tian, Z., Wing, R.A., and Han, B. (2019). The genomics of *Oryza* species provides insights into rice domestication and heterosis. Annu. Rev. Plant Biol. *70*, 639–665.

Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain *w1118*; *iso-2*; *iso-3*. Fly (Austin) *6*, 80–92.

Comai, L. (2005). The advantages and disadvantages of being polyploid. Nat. Rev. Genet. *6*, 836–846.

Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics *27*, 1164–1165.

De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. Bioinformatics *22*, 1269–1271.

DeHaan, L., Larson, S., López-Marqués, R.L., Wenkel, S., Gao, C., and Palmgren, M. (2020). Roadmap for accelerated domestication of an emerging perennial grain crop. Trends Plant Sci. *25*, 525–537.

Du, H., and Liang, C. (2019). Assembly of chromosome-scale contigs by efficiently resolving repetitive sequences with long reads. Nat. Commun. *10*, 5360.

Du, H., Yu, Y., Ma, Y., Gao, Q., Cao, Y., Chen, Z., Ma, B., Qi, M., Li, Y., Zhao, X., et al. (2017). Sequencing and *de novo* assembly of a near complete *indica* rice genome. Nat. Commun. *8*, 15324.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. *3*, 95–98.

Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for *de novo* detection of LTR retrotransposons. BMC Bioinformatics *9*, 18.

Eshed, Y., and Lippman, Z.B. (2019). Revolutions in agriculture chart a coursresee for targeted breeding of old and new crops. Science *366*, eaax0025.

Falush, D., Stephens, M., and Pritchard, J.K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics *164*, 1567–1587.

Fang, Z., and Morrell, P.L. (2016). Domestication: Polyploidy boosts domestication. Nat. Plants *2*, 16116.

Farris, J. (1977). Phylogenetic analysis under Dollos Law. Syst. Zool. *26*, 77–88.

Felsenstein, J. (1989). PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics *5*, 164–166.

Gao, H., Jin, M., Zheng, X.M., Chen, J., Yuan, D., Xin, Y., Wang, M., Huang, D., Zhang, Z., Zhou, K., et al. (2014). *Days to heading 7*, a major quantitative locus determining photoperiod sensitivity and regional adaptation in rice. Proc. Natl. Acad. Sci. USA *111*, 16337–16342.

Godfray, H.C., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., and Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people. Science *327*, 812–818.

Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science *296*, 92–100.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307–321.

Guo, T., Yu, H., Qiu, J., Li, J., Han, B., and Lin, H. (2019). Advances in rice genetics and breeding by molecular design in China. Sci. SINICA Vitae *49*, 1185–1212.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. *9*, R7.

Hardigan, M.A., Laimbeer, F.P.E., Newton, L., Crisovan, E., Hamilton, J.P., Vaillancourt, B., Wiegert-Rininger, K., Wood, J.C., Douches, D.S., Farré, E.M., et al. (2017). Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. Proc. Natl. Acad. Sci. USA *114*, E9999–E10008.

Hiei, Y., Ohta, S., Komari, T., and Kumashiro, T. (1994). Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. Plant J. *6*, 271–282.

Huang, C.L., Hwang, S.Y., Chiang, Y.C., and Lin, T.P. (2008). Molecular evolution of the *Pi-ta* gene resistant to rice blast in wild rice (*Oryza rufipogon*). Genetics *179*, 1527–1538.

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. Genet. *42*, 961–967.

Huang, X., Kurata, N., Wei, X., Wang, Z.X., Wang, A., Zhao, Q., Zhao, Y., Liu, K., Lu, H., Li, W., et al. (2012). A map of rice genome variation reveals the origin of cultivated rice. Nature *490*, 497–501.

Huang, G., Wu, Z., Percy, R.G., Bai, M., Li, Y., Frelichowski, J.E., Hu, J., Wang, K., Yu, J.Z., and Zhu, Y. (2020). Genome sequence of *Gossypium herbaceum* and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution. Nat. Genet. *52*, 516–524.

International Rice Genome Sequencing Project (2005). The map-based sequence of the rice genome. Nature *436*, 793–800.

Izawa, T., and Shimamoto, K. (1996). Becoming a model plant: The importance of rice to plant science. Trends Plant Sci. *1*, 95–99.

Jiao, Y., Wang, Y., Xue, D., Wang, J., Yan, M., Liu, G., Dong, G., Zeng, D., Lu, Z., Zhu, X., et al. (2010). Regulation of *OsSPL14* by OsmiR156 defines ideal plant architecture in rice. Nat. Genet. *42*, 541–544.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236–1240.

Khush, G.S. (1997). Origin, dispersal, cultivation and variation of rice. Plant Mol. Biol. *35*, 25–34.

Kim, J., and Buell, C.R. (2015). A revolution in plant metabolism: Genome-enabled pathway discovery. Plant Physiol. *169*, 1532–1539.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Konishi, S., Izawa, T., Lin, S.Y., Ebana, K., Fukuta, Y., Sasaki, T., and Yano, M. (2006). An SNP caused loss of seed shattering during rice domestication. Science *312*, 1392–1396.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res. *27*, 722–736.

Korf, I. (2004). Gene finding in novel genomes. BMC Bioinformatics *5*, 59.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645.

Kyriakidou, M., Tai, H.H., Anglin, N.L., Ellis, D., and Strömvik, M.V. (2018). Current strategies of polyploid plant genome sequence assembly. Front. Plant Sci. *9*, 1660.

Lacchini, E., Kiegle, E., Castellani, M., Adam, H., Jouannic, S., Gregis, V., and Kater, M.M. (2020). CRISPR-mediated accelerated domestication of African rice landraces. PLoS ONE *15*, e0229782.

Lagesen, K., Hallin, P., Rødland, E.A., Staerfeldt, H.H., Rognes, T., and Ussery, D.W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. *35*, 3100–3108.

Lemmon, Z.H., Reem, N.T., Dalrymple, J., Soyk, S., Swartwood, K.E., Rodriguez-Leal, D., Van Eck, J., and Lippman, Z.B. (2018). Rapid improvement of domestication traits in an orphan crop by genome editing. Nat. Plants *4*, 766–770.

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

Li, L., Stoeckert, C.J., Jr., and Roos, D.S. (2003a). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. *13*, 2178–2189.

Li, X., Qian, Q., Fu, Z., Wang, Y., Xiong, G., Zeng, D., Wang, X., Liu, X., Teng, S., Hiroshi, F., et al. (2003b). Control of tillering in rice. Nature *422*, 618–621.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R.J., Ma, Z., Shang, H., Ma, X., Wu, J., et al. (2015). Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nat. Biotechnol. *33*, 524–530.

Li, P., Quan, X., Jia, G., Xiao, J., Cloutier, S., and You, F.M. (2016). RGAugury: a pipeline for genome-wide prediction of resistance gene analogs (RGAs) in plants. BMC Genomics *17*, 852.

Li, T., Yang, X., Yu, Y., Si, X., Zhai, X., Zhang, H., Dong, W., Gao, C., and Xu, C. (2018). Domestication of wild tomato is accelerated by genome editing. Nat. Biotechnol. *36*, 1160–1163.

Lichman, B.R., Godden, G.T., and Buell, C.R. (2020). Gene and genome duplications in the evolution of chemodiversity: perspectives from studies of *Lamiaceae*. Curr. Opin. Plant Biol. *55*, 74–83.

Lin, Y.J., and Zhang, Q. (2005). Optimising the tissue culture conditions for high efficiency transformation of *indica* rice. Plant Cell Rep. *23*, 540–547.

Long, S.P., Marshall-Colon, A., and Zhu, X.G. (2015). Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. Cell *161*, 56–66.

Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. *25*, 955–964.

Lu, F., Ammiraju, J.S., Sanyal, A., Zhang, S., Song, R., Chen, J., Li, G., Sui, Y., Song, X., Cheng, Z., et al. (2009). Comparative sequence analysis of *MONOCULM1*-orthologous regions in 14 *Oryza* genomes. Proc. Natl. Acad. Sci. USA *106*, 2071–2076.

Luo, J., Liu, H., Zhou, T., Gu, B., Huang, X., Shangguan, Y., Zhu, J., Li, Y., Zhao, Y., Wang, Y., et al. (2013). *An-1* encodes a basic helix-loop-helix protein that regulates awn development, grain size, and grain number in rice. Plant Cell *25*, 3360–3376.

Luo, J.S., Huang, J., Zeng, D.L., Peng, J.S., Zhang, G.B., Ma, H.L., Guan, Y., Yi, H.Y., Fu, Y.L., Han, B., et al. (2018). A defensin-like protein drives cadmium efflux and allocation in rice. Nat. Commun. *9*, 645.

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

Ma, Y., Dai, X., Xu, Y., Luo, W., Zheng, X., Zeng, D., Pan, Y., Lin, X., Liu, H., Zhang, D., et al. (2015b). *COLD1* confers chilling tolerance in rice. Cell *160*, 1209–1221.

Majoros, W.H., Pertea, M., and Salzberg, S.L. (2004). TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. Bioinformatics *20*, 2878–2879.

Mao, H., Sun, S., Yao, J., Wang, C., Yu, S., Xu, C., Li, X., and Zhang, Q. (2010). Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. Proc. Natl. Acad. Sci. USA *107*, 19579–19584.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

Meng, X., Yu, H., Zhang, Y., Zhuang, F., Song, X., Gao, S., Gao, C., and Li, J. (2017). Construction of a genome-wide mutant library in rice using CRISPR/Cas9. Mol. Plant *10*, 1238–1241.

Menguer, P.K., Sperotto, R.A., and Ricachenevsky, F.K. (2017). A walk on the wild side: *Oryza* species as source for rice abiotic stress tolerance. Genet. Mol. Biol. *40* (1, suppl 1), 238–252.

Miura, K., Ikeda, M., Matsubara, A., Song, X.J., Ito, M., Asano, K., Matsuoka, M., Kitano, H., and Ashikari, M. (2010). OsSPL14 promotes panicle branching and higher grain productivity in rice. Nat. Genet. *42*, 545–549.

Mohammed, S., Samad, A.A., and Rahmat, Z. (2019). *Agrobacterium*-mediated transformation of rice: constraints and possible solutions. Rice Science *26*, 133–146.

Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics *29*, 2933–2935.

Nelson, G.C., Rosegrant, M.W., Koo, J., Robertson, R., Sulser, T., Zhu, T., Ringler, C., Msangi, S., Palazzo, A., and Batka, M. (2009). Climate Change: Impact on Agriculture and Costs of Adaptation*Volume 21* (Intl. Food Policy Res. Inst.).

Ou, S., and Jiang, N. (2018). LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. Plant Physiol. *176*, 1410–1422.

Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190.

Price, A.L., Jones, N.C., and Pevzner, P.A. (2005). *De novo* identification of repeat families in large genomes. Bioinformatics *21* (*Suppl 1*), i351–i358.

Prusty, M.R., Kim, S.R., Vinarao, R., Entila, F., Egdane, J., Diaz, M.G.Q., and Jena, K.K. (2018). Newly identified wild rice accessions conferring high salt tolerance might use a tissue tolerance mechanism in leaf. Front. Plant Sci. *9*, 417.

Qian, Q., Guo, L., Smith, S.M., and Li, J. (2016). Breeding high-yield superior quality hybrid super rice by rational design. Natl. Sci. Rev. *3*, 283–294.

Rosenzweig, C., and Parry, M. (1994). Potential impact of climate change on world food supply. Nature *367*, 133–138.

Sasaki, A., Ashikari, M., Ueguchi-Tanaka, M., Itoh, H., Nishimura, A., Swapan, D., Ishiyama, K., Saito, T., Kobayashi, M., Khush, G.S., et al. (2002). Green revolution: a mutant gibberellin-synthesis gene in rice. Nature *416*, 701–702.

Seppey, M., Manni, M., and Zdobnov, E.M. (2019). BUSCO: assessing genome assembly and annotation completeness. Methods Mol. Biol. *1962*, 227–245.

Shelton, J.M., Coleman, M.C., Herndon, N., Lu, N., Lam, E.T., Anantharaman, T., Sheth, P., and Brown, S.J. (2015). Tools and pipelines for BioNano data: molecule assembly pipeline and FASTA super scaffolding tool. BMC Genomics *16*, 734.

Shimamoto, K., Terada, R., Izawa, T., and Fujimoto, H. (1989). Fertile transgenic rice plants regenerated from transformed protoplasts. Nature *338*, 274–276.

Shimizu-Sato, S., Tsuda, K., Nosaka-Takahashi, M., Suzuki, T., Ono, S., Ta, K.N., Yoshida, Y., Nonomura, K.I., and Sato, Y. (2020). Agrobacterium-mediated genetic transformation of wild *oryza* species using immature embryos. Rice (N. Y.) *13*, 33.

Song, W.Y., Wang, G.L., Chen, L.L., Kim, H.S., Pi, L.Y., Holsten, T., Gardner, J., Wang, B., Zhai, W.X., Zhu, L.H., et al. (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science *270*, 1804–1806.

Stanke, M., Steinkamp, R., Waack, S., and Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. *32*, W309-12.

Stein, J.C., Yu, Y., Copetti, D., Zwickl, D.J., Zhang, L., Zhang, C., Chougule, K., Gao, D., Iwata, A., Goicoechea, J.L., et al. (2018). Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza*. Nat. Genet. *50*, 285–296.

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. Genetics *105*, 437–460.

Tang, H., Wang, X., Bowers, J.E., Ming, R., Alam, M., and Paterson, A.H. (2008). Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. Genome Res. *18*, 1944–1954.

Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinformatics *Chapter 4*, Unit 4.10.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673–4680.

Tian, Z., Qian, Q., Liu, Q., Yan, M., Liu, X., Yan, C., Liu, G., Gao, Z., Tang, S., Zeng, D., et al. (2009). Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities. Proc. Natl. Acad. Sci. USA *106*, 21760–21765.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Van de Peer, Y., Mizrachi, E., and Marchal, K. (2017). The evolutionary significance of polyploidy. Nat. Rev. Genet. *18*, 411–424.

Van Tassel, D.L., Tesdell, O., Schlautman, B., Rubin, M.J., DeHaan, L.R., Crews, T.E., and Streit Krug, A. (2020). New food crop domestication in the age of gene editing: genetic, agronomic and cultural change remain co-evolutionarily entangled. Front. Plant Sci. *11*, 789.

VanBuren, R., Man Wai, C., Wang, X., Pardo, J., Yocca, A.E., Wang, H., Chaluvadi, S.R., Han, G., Bryant, D., Edger, P.P., et al. (2020). Exceptional subgenome stability and functional divergence in the allotetraploid Ethiopian cereal teff. Nat. Commun. *11*, 884.

Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., and Schatz, M.C. (2017). GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics *33*, 2202–2204.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE *9*, e112963.

Wang, B., Smith, S.M., and Li, J. (2018a). Genetic regulation of shoot architecture. Annu. Rev. Plant Biol. *69*, 437–468.

Wang, J., Zhou, L., Shi, H., Chern, M., Yu, H., Yi, H., He, M., Yin, J., Zhu, X., Li, Y., et al. (2018b). A single transcription factor promotes both yield and immunity in rice. Science *361*, 1026–1028.

Weng, X., Wang, L., Wang, J., Hu, Y., Du, H., Xu, C., Xing, Y., Li, X., Xiao, J., and Zhang, Q. (2014). *Grain number, plant height, and heading date7* is a central regulator of growth, development, and stress response. Plant Physiol. *164*, 735–747.

Wheeler, T., and von Braun, J. (2013). Climate change impacts on global food security. Science *341*, 508–513.

Wing, R.A., Purugganan, M.D., and Zhang, Q. (2018). The rice genome revolution: from an ancient grain to Green Super Rice. Nat. Rev. Genet. *19*, 505–517.

Wu, J., Shahid, M.Q., Guo, H., Yin, W., Chen, Z., Wang, L., Liu, X., and Lu, Y. (2014). Comparative cytological and transcriptomic analysis of pollen development in autotetraploid and diploid rice. Plant Reprod. *27*, 181–196.

Xie, X., Ma, X., Zhu, Q., Zeng, D., Li, G., and Liu, Y.G. (2017). CRISPR-GE: a convenient software toolkit for crispr-based genome editing. Mol. Plant *10*, 1246–1249.

Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. *35*, W265-8.

Xu, C., Bai, Y., Lin, X., Zhao, N., Hu, L., Gong, Z., Wendel, J.F., and Liu, B. (2014). Genome-wide disruption of gene expression in allopolyploids but not hybrids of rice subspecies. Mol. Biol. Evol. *31*, 1066–1076.

Xue, W., Xing, Y., Weng, X., Zhao, Y., Tang, W., Wang, L., Zhou, H., Yu, S., Xu, C., Li, X., and Zhang, Q. (2008). Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. Nat. Genet. *40*, 761–767.

Yan, W., Liu, H., Zhou, X., Li, Q., Zhang, J., Lu, L., Liu, T., Liu, H., Zhang, C., Zhang, Z., et al. (2013). Natural variation in *Ghd7.1* plays an important role in grain yield and adaptation in rice. Cell Res. *23*, 969–971.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591.

Yang, J., Moeinzadeh, M.H., Kuhl, H., Helmuth, J., Xiao, P., Haas, S., Liu, G., Zheng, J., Sun, Z., Fan, W., et al. (2017). Haplotype-resolved sweet potato genome traces back its hexaploidization history. Nat. Plants *3*, 696–703.

Yoon, J., Cho, L.H., Kim, S.L., Choi, H., Koh, H.J., and An, G. (2014). The BEL1-type homeobox gene *SH5* induces seed shattering by enhancing abscission-zone development and inhibiting lignin biosynthesis. Plant J. *79*, 717–728.

Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science *296*, 79–92.

Zeng, D., Tian, Z., Rao, Y., Dong, G., Yang, Y., Huang, L., Leng, Y., Xu, J., Sun, C., Zhang, G., et al. (2017). Rational design of high-yield and superior-quality rice. Nat. Plants *3*, 17031.

Zhang, Q.J., Zhu, T., Xia, E.H., Shi, C., Liu, Y.L., Zhang, Y., Liu, Y., Jiang, W.K., Zhao, Y.J., Mao, S.Y., et al. (2014). Rapid diversification of five *Oryza* AA genomes associated with rice adaptation. Proc. Natl. Acad. Sci. USA *111*, E4954–E4962.

Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., Zhang, J., Saski, C.A., Scheffler, B.E., Stelly, D.M., et al. (2015). Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. Nat. Biotechnol. *33*, 531–537.

Zhang, Y., Li, J., and Gao, C. (2016). Generation of stable transgenic rice (*Oryza sativa* L.) by *agrobacterium*-mediated transformation. Curr. Protoc. Plant Biol. *1*, 235–246.

Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., Zhu, F., Jones, T., Zhu, X., Bowers, J., et al. (2018). Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. Nat. Genet. *50*, 1565–1573.

Zhang, L., Shivute, F.N., Shahid, M.Q., Kamara, N., Wu, J., and Liu, X. (2019). In vitro induction of auto-allotetraploid in a newly developed wild rice line from *Oryza alta* Swallen. Plant Cell Tissue Organ Cult. *139*, 577–587.

Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., Zhan, Q., Lu, Y., Zhang, L., Huang, T., et al. (2018). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. Nat. Genet. *50*, 278–284.

Zheng, Y., Jiao, C., Sun, H., Rosli, H.G., Pombo, M.A., Zhang, P., Banf, M., Dai, X., Martin, G.B., Giovannoni, J.J., et al. (2016). iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. Mol. Plant *9*, 1667–1670.

Zhou, Y., Uzokwe, V.N., Zhang, C., Cheng, L., Wang, L., Chen, K., Gao, X., Sun, Y., Chen, J., Zhu, L., et al. (2011). Improvement of bacterial blight resistance of hybrid rice in China using the *Xa23* gene derived from wild rice (*Oryza rufipogon*). Crop Prot. *30*, 637–644.

Zhou, Y., Chebotarov, D., Kudrna, D., Llaca, V., Lee, S., Rajasekar, S., Mohammed, N., Al-Bader, N., Sobel-Sorenson, C., Parakkal, P., et al. (2020). A platinum standard pan-genome resource that represents the population structure of Asian rice. Sci. Data *7*, 113.

Zong, Y., Song, Q., Li, C., Jin, S., Zhang, D., Wang, Y., Qiu, J.L., and Gao, C. (2018). Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. Nat. Biotechnol. *36*, 950–953.

Zou, X.H., Zhang, F.M., Zhang, J.G., Zang, L.L., Tang, L., Wang, J., Sang, T., and Ge, S. (2008). Analysis of 142 genes resolves the rapid diversification of the rice genus. Genome Biol. *9*, R49.

Zou, X.H., Yang, Z., Doyle, J.J., and Ge, S. (2013). Multilocus estimation of divergence times and ancestral effective population sizes of *Oryza* species and implications for the rapid diversification of the genus. New Phytol. *198*, 1155–1164.

Zsögön, A., Čermák, T., Naves, E.R., Notini, M.M., Edel, K.H., Weinl, S., Freschi, L., Voytas, D.F., Kudla, J., and Peres, L.E.P. (2018). *De novo* domestication of wild tomato using genome editing. Nat. Biotechnol. *36*, 1211–1216.

Zuccolo, A., Sebastian, A., Talag, J., Yu, Y., Kim, H., Collura, K., Kudrna, D., and Wing, R.A. (2007). Transposable element distribution, abundance and role in genome size variation in the genus *Oryza*. BMC Evol. Biol. *7*, 152.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Agrobacterium tumefaciens strain EHA 105	Meng et al., 2017	N/A
Chemicals, peptides, and recombinant proteins		
BsaI	NEB	Cat#R0535L
T4 DNA Ligase	NEB	Cat#M0202L
Phloroglucinol	SIGMA	Cat#108-73-6
Critical commercial assays		
Wizard SV Gel and PCR Clean-Up System	Promega	Cat#A9282
TRIzol Kit	Invitrogen	Cat#15596018
Maxima H Minus cDNA Synthesis Master Mix	Thermo Scientific	Cat#M1682
SsoFast EvaGreen Supermix	Bio-Rad	Cat#1725201
VK005-01	Viewsolid Biotech	Cat#VK005-01
Deposited data		
Allotetraploid rice: Pacbio, Bionano, Hi-C and DNA sequencing data of PPR1	This study	BioProject: PRJCA002366
Allotetraploid rice: *de novo* assembly genome data of PPR1	This study	BioProject: PRJCA002366
Allotetraploid rice: Transcriptome sequence data	This study	BioProject: PRJCA002366
Allotetraploid and diploid rice: DNA re-sequencing data of CCDD, CC and EE species	This study	BioProject: PRJCA002366
The original unprocessed data	This study	Mendeley Dataset: https://doi.org/10.17632/mk3ykkms44.1
Oligonucleotides		
Forward primer sequence for quantitative real-time PCR analysis for *OaUBI* (OalC01 g115320 and OalC03 g154610): AACCAGCTGAGGCCCAAGA	This study	N/A
Reverse primer sequences for quantitative real-time PCR analysis for *OaUBI* (OalC01 g115320 and OalC03 g154610): ACGATTGATTTAACCAGTCCATGA	This study	N/A
Forward primer sequences for quantitative real-time PCR analysis for *OalPA1-CC* (OalC08 g106170): GTGTGGGACCTTTCCTGCTA	This study	N/A
Reverse primer sequences for quantitative real-time PCR analysis for *OalPA1-CC* (OalC08 g106170): CATAAGCACCGCTCATCAAA	This study	N/A
Forward primer sequences for quantitative real-time PCR analysis for *OalPA1-DD* (OalD08 g132520): TTCAGTTTCGGTTGGGAGAG	This study	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Reverse primer sequences for quantitative real-time PCR analysis for *OalPA1-DD* (OalD08 g132520): TGGAGGGCTGGAAAGATACA	This study	N/A
Recombinant DNA		
pYLCRISPR/Cas9Pubi-H	Ma et al., 2015a	Addgene Plasmid #66187; RRID:Addgene_66187
A3A-PBE	Zong et al., 2018	Addgene Plasmid #119768; RRID:Addgene_119768
Software and algorithms		
Genomescope (version 1.0)	Vurture et al., 2017	http://www.genome.umd.edu/jellyfish.html#Release
CANU (version 1.3)	Koren et al., 2017	https://github.com/marbl/canu
Aigner and Assembler (version 2.0.6)	Shelton et al., 2015	https://bionanogenomics.com/
HERA (version 1.0)	Du and Liang, 2019	https://github.com/liangclab/HERA/
Juicer (version 1.5)	Durand et al., 2016	https://github.com/aidenlab/juicer/
Pilon (version 1.23)	Walker et al., 2014	https://github.com/broadinstitute/pilon/releases
BUSCO (version 3.0.2)	Seppey et al., 2019	https://busco.ezlab.org/
AUGUSTUS (version 3.3)	Stanke et al., 2004	https://sourceforge.net/projects/augustus/
GENSCAN(version 1.0)	Burge and Karlin, 1997	http://hollywood.mit.edu/gendist.html
Glimmer-HMM(version 3.0.4)	Majoros et al., 2004	http://www.cbcb.umd.edu/software/glimmerm/index.shtml
SNAP (version 2.3)	Korf, 2004	https://github.com/KorfLab/SNAP
NCBI BLAST (version 2.2.24)	Camacho et al., 2009	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
GeneWise (version 2.1.20 stable)	Birney and Durbin, 2000	https://www.ebi.ac.uk/~birney/wise2/
EVidenceModeler (version 1.1.1)	Haas et al., 2008	https://github.com/EVidenceModeler/EVidenceModeler/releases
InterProScan (V5.19-58.0)	Jones et al., 2014	https://www.ebi.ac.uk/interpro/
RepeatScout (version 1.0.5)	Price et al., 2005	http://bix.ucsd.edu/repeatscout/
RECON (version 1.0.8)	Bao and Eddy, 2002	http://www.repeatmasker.org/RepeatModeler/
RepeatMasker(version 4.0.7)	Tarailo-Graovac and Chen, 2009	http://repeatmasker.org/
RNAmmer (version 1.2)	Lagesen et al., 2007	http://www.cbs.dtu.dk/services/RNAmmer/
tRNAscan-SEM (version 1.23)	Lowe and Eddy, 1997	http://lowelab.ucsc.edu/tRNAscan-SE/
INFERNAL (version 1.1.3)	Nawrocki and Eddy, 2013	http://eddylab.org/infernal/
LTR-FINDER (version 1.07)	Xu and Wang, 2007	https://github.com/xzhub/LTR_Finder
LTR-Harvest (version 1.5.7)	Ellinghaus et al., 2008	https://github.com/Adamtaranto/LTR-Harvest-Reformat
LTR-retriever (version 2.8.5)	Ou and Jiang, 2018	https://github.com/oushujun/LTR_retriever
TopHat2 (version 2.1.0)	Kim et al., 2013	http://ccb.jhu.edu/software/tophat/index.shtml
Cufflinks (version 0.9.3)	Trapnell et al., 2010	https://github.com/cole-trapnell-lab/cufflinks
R (versio 3.5.0)	R Core Team	https://www.r-project.org/
OrthoMCL (version 2.0.9)	Li et al., 2003a	https://orthomcl.org/orthomcl/
CLUSTALW (version 2.0.10)	Thompson et al., 1994	https://www.ebi.ac.uk/Tools/msa/clustalw2/

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
MCMCTree (version 4.9)	Yang, 2007	http://web.mit.edu/6.891/www/lab/paml.html
DOLLOP (RV267)	Farris, 1977	https://sourceforge.net/projects/dollop/
CAFÉ (version 4.2.1)	De Bie et al., 2006	https://github.com/hahnlab/CAFE
BWA (version 0.7.17)	Li and Durbin, 2009	http://bio-bwa.sourceforge.net
SAMtools (version 1.5)	Li et al., 2009	http://samtools.sourceforge.net
MCScan (version 1.1)	Tang et al., 2008	https://github.com/tanghaibao/mcscan
Circos (version 0.69-9)	Krzywinski et al., 2009	http://www.circos.ca/software/
snpEFF (version 4.3)	Cingolani et al., 2012	http://snpeff.sourceforge.net
PHYLIP (version 3.698)	Felsenstein, 1989	https://evolution.genetics.washington.edu/phylip.html
STRUCTURE (version 2.3.4)	Falush et al., 2003	https://web.stanford.edu/group/pritchardlab/structure.html
EIG (version 6.0.1)	Patterson et al., 2006	https://github.com/DReichLab/EIG/tree/master/EIGENSTRAT
RGAugury	Li et al., 2016	https://bitbucket.org/yaanlpc/rgaugury
iTAK (version 18.12)	Zheng et al., 2016	http://itak.feilab.net/cgi-bin/itak/index.cgi
CRISPR-GE	Xie et al., 2017	http://skl.scau.edu.cn/

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be made available upon reasonable request by the Lead Contact, Jiayang Li (jyli@genetics.ac.cn).

Materials availability
The study did not generate new unique reagents.

Data and code availability
The sequence data reported in this study have been deposited in the Genome Sequence Archive in BIG Data Center (http://bigd.big.ac.cn/gsa) under accession number PRJCA002366.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

A total of 44 wild rice accessions were collected, including 28 CCDD species (eight *O. alta*, two *O. grandiglumis*, and 18 *O. latifolia*), ten CC species, six EE species (Table S1). Plants and the tissue culture plantlets were grown and cultivated in the experimental station and the greenhouses of Institute of Genetics and Developmental Biology, Chinese Academy of Sciences (IGDB-CAS), Beijing, China (40°22′N and 116°23′E) and/or Hainan (18°31′N and 109°62′E) in the growth season from 2016 to 2020.

METHOD DETAILS

Assembly of the PPR1 genome
Genomic DNA was extracted from young leaves using the cetyltrimethylammonium bromide (CTAB) method (Allen et al., 2006). At least 5 μg of genomic DNA was used for each accession to construct paired-end sequencing libraries with insert sizes of approximately 300 bp. For SMRT PacBio sequencing, a 20-kb insert SMRTbell library was generated using a 15-kb lower-end size selection protocol on a BluePippin (Sage Science). The PPR1 genome was sequenced on the PacBio RS II platform (Pacific Biosciences). PacBio raw reads were processed to filter low-quality reads (RQ > 0.75). The number of 19-mers for corrected PE150 reads was counted using the Genomescope program (Vurture et al., 2017). The PPR1 genome size was calculated by dividing the total number of k-mers times of the highest k-mer frequency in the whole genome. The following steps were used to assemble, improve and correct the PPR1 genome assembly. First, ~121.1 Gb of clean PacBio reads were pre-processed and assembled into contigs using the CANU software (Koren et al., 2017). Then 287.1 Gb of Bionano data were used for assisting the scaffold construction using Aigner and Assembler (Shelton et al., 2015). To correct and fill gaps for misassembly, the assembler HERA (Du and Liang, 2019) was used to

update the scaffolds to supercontigs. The resulting supercontigs and scaffolds were integrated using 103.2 Gb of Hi-C data using Juicer (Durand et al., 2016). In the final round, the 85.2 Gb of Illumina paired-end reads were aligned to PPR1 genome assembly to correct errors using Pilon (Walker et al., 2014).

Two types of data were retrieved for the identification of homeologous subgenomes in the PPR1 genome: 1) containing BAC sequences (Ammiraju et al., 2010b), and 2) 278 genes generated from CC and EE species (Zou et al., 2008). First, we aligned these sequences to the PPR1 genome using NCBI BLAST with an E-value cutoff of $1e^{-5}$. Next, the best hits were filtered based on the sequence length of 100 nucleotides and greater than 80% sequence identity. Then, the information of these aligned BAC sequences and genes were used to label the numbers of 24 chromosomes of PPR1 as C_t1-12 and D_t1-12. To obtain homeologous gene pairs between both subgenomes in the PPR1 genome, we first aligned proteins obtained from the C_t subgenome against the D_t gene set using NCBI BLAST version 2.2.24 (Camacho et al., 2009) with an E-value cutoff of $1e^{-5}$. Finally, the best hits were filtered based on 50% amino acid length and a score of more than 50.

Genome annotation
Gene prediction
Three types of gene prediction methods were combined to annotate the PPR1 genome, including *ab initio*, assembled transcripts and protein homolog prediction. Four programs, including AUGUSTUS (Stanke et al., 2004), GENSCAN (Burge and Karlin, 1997), Glimmer-HMM (Majoros et al., 2004), and SNAP (Korf, 2004) were performed for *ab initio* prediction. RNA-seq reads from multiple tissues were mapped to the PPR1 genome using TopHat2 (Kim et al., 2013), and assembled into transcripts using Cufflinks (Trapnell et al., 2010) for gene structures and new genes. Protein sequences from the *O. sativa* ssp. *indica* and *japonica* and uniprot database were aligned to the PPR1 genome using NCBI BLAST (default parameters except for an e-value $< 10^{-5}$) for homology-based prediction. Accurate spliced sites were identified using GeneWise (version 2.1.20 stable) (Birney and Durbin, 2000). All predicted gene structure evidence was integrated with EVidenceModeler (EVM) (Haas et al., 2008) to produce high-confidence gene models. Gene models were filtered with the parameters: codon length > 150 bp and transcripts supported by homology-base methods. Finally, the resulting gene sets were checked for completeness in comparison with the BUSCO gene set (Seppey et al., 2019). The predicted gene models with FPKM > 0.01 in any of RNA-seq samples were labeled as High-Confidence (HC) genes, and other gene models as Low-Confidence (LC) genes.

Gene function annotation
Gene functions were assigned according to the best match for each predicted gene in alignment to multiple databases including the NCBI non-redundant (nr) protein database, SWISS-PROT, and Clusters of Orthologous Groups (COG) using the NCBI BLAST program (E-value $< 10^{-5}$). Gene ontology (GO) analysis was performed with InterProScan V5.19-58.0 using the default setting (Jones et al., 2014).

Annotation of repetitive sequences
The assembled genome was searched for repeats with a combination of *de novo* and homolog based strategies. The initial repeat library was built using two *de novo* programs including RepeatScout (Price et al., 2005) and RECON (Bao and Eddy, 2002). The initial repeat library and the Repbase database for the homolog repeat annotation were classified into classes, subclasses, superfamilies and families by RepeatMasker (Tarailo-Graovac and Chen, 2009).

Annotation of non-coding RNAs
Ribosomal RNAs (rRNAs) were identified using RNAmmer with default parameters (Lagesen et al., 2007). Reliable tRNA structures were detected using tRNAscan-SEM (version 1.23) (Lowe and Eddy, 1997). Non-coding RNAs containing miRNA, snRNA, and snoRNA features were annotated using INFERNAL with default parameters (Nawrocki and Eddy, 2013). Noncoding RNAs were classified into different families through searching against the Rfam database.

Annotation of resistance genes, transcription factors and transcription regulators
The resistance genes were annotated using RGAugury (Li et al., 2016) with default parameters. The transcription factors and transcription regulators were annotated using iTAK (version 18.12) (Zheng et al., 2016) with default parameters.

LTR-RT identification and insertion time estimation
Long terminal repeat retrotransposons (LTR-RTs) were scanned in the PPR1 genome using LTR-FINDER (Xu and Wang, 2007) and LTR-Harvest (Ellinghaus et al., 2008). The identification of complete LTR-RTs was obtained by adjusting the ends of LTR-pair terminal sequences using LTR-retriever (Ou and Jiang, 2018). LTR index of PPR1 is 22.11. LTR insertion times were estimated according to the formula $T = d/2\mu$ (d, the nucleotide distance for each pair of LTRs; μ, the nucleotide substitution rate estimated as previously described). The orthologous LTR-RTs in *O. alta* and *Brachypodium distachyon* were identified to measure an appropriate LTR nucleotide substitution rate based on the divergence time ($\sim$52.0 MYA) of both species (Bennetzen et al., 2012). The mutation rate was estimated as 0.65×10^{-8} substitutions per site per year.

Transcriptome data analysis

For RNA-seq analysis, 10 representative tissues were collected in the greenhouse, including young tissues of seedling (leaf, stem and root), mature plant tissues (leaf, stem, root, seed, panicle, sheath), and callus. Three biological repeats were made for each tissue except seed and panicle. Total RNA was extracted with the Quick RNA isolation Kit according to the manufacturer's instructions (Huayueyang, China). RNA libraries (300 bp) were constructed and sequenced using the BGI platform. Raw RNA-seq reads (PE100) were processed to trim terminal low quality bases and adaptor sequences using a custom pipeline. High-quality RNA-seq reads were aligned to the assembled genome using TopHat2 (Kim et al., 2013) with default parameters. Transcript expression levels, in fragments per kilobase per million reads mapped (FPKM) for each sample, were determined using Cufflinks (Trapnell et al., 2010). Differentially expressed homologous pairs on the corresponding chromosomes of the two subgenomes were determined by Student's t test with Bonferroni correction using 28 samples of RNA-seq.

Gene family analysis

Ten species, including *O. alta*, *O. sativa*, *B. distachyon*, *Aegilops tauschii*, *Triticum urartu*, *T. aestivum*, *T. dicoccoides*, *Setaria italica*, *Sorghum bicolor*, and *Zea mays* were selected to construct gene families on the basis of protein homologs (sequence length > 30, e-value < 10^{-5}, and alignment length coverage > 50%) using the OrthoMCL (Li et al., 2003a). Single-copy families were used to build the phylogenetic tree. For each single-copy gene family, we aligned and clustered the protein sequences from the 10 species with CLUS-TALW (Thompson et al., 1994) and transformed the alignments to the phylip format. We determined the best fitting model for the construction of phylogenetic trees using Prottest (version 3.4) (Darriba et al., 2011). Phylogenetic trees were built with PhyML (Guindon et al., 2010) with 1,000 bootstrap replicates. The divergence time of each node was estimated based on the JC69 model using the MCMCTree program in the PAML package (Yang, 2007) with default parameters except for RooAge (< 2.0). Two reported divergence times were used as a calibration, including ~26 MYA for Sorghum and Maize, and ~52 MYA for *O. sativa* and *B. distachyon* (Falush et al., 2003). Gain and loss of gene families for the ten species were predicted using DOLLOP (Farris, 1977) from the PHYLIP package (Felsenstein, 1989). Expansion and contraction of gene families were determined using CAFÉ (De Bie et al., 2006) with the lambda value of 0.184609. A phylogenetic tree from the above study was used as a probabilistic model to infer family expansions and contractions.

Population genetics analysis

Genomic DNA from 44 wild *Oryza* rice accessions was extracted from fresh young leaves using the CTAB method (Allen et al., 2006). All accessions were re-sequenced on the Illumina NovoSeq 5000 or BGISEG-500 platform, generating 150-bp or 100-bp paired-end reads. All sequence reads from the 44 accessions were aligned to the PPR1 reference genome using BWA (Li and Durbin, 2009) and SAMtools (Li et al., 2009) with the default parameters. Variations (SNP and INDEL) were detected with GATK (McKenna et al., 2010) with base quality $\geq$ 30. The identified variations were further categorized as variations in intergenic regions, UTRs, coding sequences and introns according to the PPR1 genome annotation using snpEFF (Cingolani et al., 2012). To build a neighbor-joining tree, we screened a subset of 524,933 SNPs (MAF > 5% and missing data < 40%) at fourfold-degenerate sites in the 44 rice accessions from the entire SNP dataset. A phylogenetic tree was constructed using PHYLIP with 100 bootstrap replicates. Using the same dataset, we also investigated the population structure using STRUCTURE on the basis of allele frequencies (Falush et al., 2003). In addition, principal-component analysis (PCA) was generated using all SNPs across the genome (MAF > 10%, missing < 5%) (Patterson et al., 2006). Nucleotide diversity for each group was measured based on the degree of variability (Tajima, 1983). The level of genetic diversity was calculated using a 100-kb window with a step size of 10 kb in each group.

Collinear analysis

For each pairwise alignment, the coding sequences of predicted gene models were compared to each other using all-versus-all BLASTP (e-value < 10^{-5}) within the PPR1 genome. Syntenic blocks (with at least five genes per block) were identified by MCScan (Tang et al., 2008) with the default parameters. The collinear blocks represented pairs of genomic sequences containing at least five collinear genes. Whole-genome syntenic relationships between the species were visualized using Circos (Krzywinski et al., 2009).

Homologous genes of diploid rice in *O. alta*

To identify homologous genes related to domestication in *O. sativa* in the *O. alta* genome, the genes in diploid rice were first aligned against the PPR1 gene set using BLASTP with an E-value cutoff of 1e^{-5}. Next, the best hits were filtered based on 50% amino acid length and a score of more than 50. The top two best-matched results were assigned as the homologs for the following studies in the domestication-related and agronomically important genes.

Transgenic system of *O. alta*

The genetic transformation system of *O. alta* was established as previously described (Zhang et al., 2016) with modifications, which are briefly described as following.

Callus induction and proliferation. (1) The mature dehusked seeds without pathogen spots were surface-sterilized with ethanol and bleach solution. (2) The sterilized seeds were transferred to the M1 calli induction medium (Zhang et al., 2016) with the modification of replacing the sucrose with maltose and incubated in dark at 28°C. (3) After 15 days, calli were formed from the scutella of mature

seeds. Embryogenic calli with a granular structure, smooth surface and creamy/yellow color were selected and subcultured in fresh M1 medium in the dark at 28°C for 15 days after removing the seedling and seed structures. The selection of embryogenic calli from non-embryogenic calli is essential. (4) The subculturing process can be repeated a few times to get more calli, but should be performed for less than a three-month period.

Transformation and selection. (1) Activation of *Agrobacterium tumefaciens* strain EHA 105 containing constructs of interest was performed as previously described (Zhang et al., 2016). (2) The *Agrobacterium* was then resuspended in cooled (10°C) fresh infection medium (supplemented with freshly prepared 200 μM acetosyringone and 0.1% Pluronic F68) and adjusted to an optical density with OD_{600} 0.5 and placed on ice for 10-20 minutes. (3) For infection, the *Agrobacterium* cell suspension was poured into the flask immersing the calli, and the flask was then gently shaken at 40 rpm for 10 minutes. The calli were placed onto a stack of sterile filter paper in the dishes and co-cultivated in the dark at 22°C for 3 days. (4) After co-cultivation, the infected calli were transferred to freshly prepared solid M2 selection medium and were cultured in the dark at 28°C for 7 days. The calli were sub-cultured by changing the fresh selection medium every two weeks for three cycles. The newly formed hygromycin B resistant and actively growing micro-calli were subcultured under the same conditions.

Regeneration and hardening of transgenic plants. (1) Embryogenic resistant calli were transferred to solid M5 regeneration medium and incubated at 28°C in the dark for 1 week, and then place in a 16-hour light/8-hour dark photoperiod for 3 weeks under cold fluorescent light with an intensity 50 - 100 μE m^{-2} s^{-1}. Green spots in the calli and newly formed shoots could be seen in 2 and 3 weeks, respectively. (2) After 4 weeks, regenerated shoots were transferred to M6 rooting medium and incubated at 28°C until the shoots grew to 10 cm. (3) Plants were then hardened by opening the lid of the medium at 28°C for 5 days with watering. Finally, the plants were washed with warm water and transplanted to soil in the greenhouse or field.

CRISPR/Cas9 based genome editing of *O. alta*

Single-guiding RNAs (sgRNAs) were designed using the CRISPR-GE tool (http://skl.scau.edu.cn/) (Xie et al., 2017). The sgRNAs used in this study are listed in Table S6. The sgRNAs were cloned by enzyme digestion and ligation methods into the binary vector VK005-01 for single gene editing and A3A-PBE for base editing, or generated into sgRNAs expression cassettes using the overlapping PCR method and then cloned by the Golden Gate cloning method into pYLCRISPR/Cas9P$_{ubi}$-H for multiplex gene editing (Ma et al., 2015a; Zong et al., 2018). The final binary vector was transformed into PPR1 by the established *Agrobacterium*-mediated transformation methods. For genotyping of the T$_0$ transgenic lines, leaves of T$_0$ transgenic lines were collected from each plant and genomic DNA was isolated using the CTAB method (Allen et al., 2006). PCR amplification was performed using primers flanking the sgRNA target sites (Table S6) and Sanger sequencing was used to identify mutations in target genes of T$_0$ plants.

The T$_0$ seedlings of *OaGhd7-CC*, *OaGhd7-DD*, *OaDTH7-CC*, and *OaDTH7-DD* edited plants and the wild-type PPR1 were refined in the greenhouse of Institute of Genetics and Developmental Biology in Beijing, and then transplanted to the experimental fields of Institute of Genetics and Developmental Biology in Beijing when the plants grown up to 20 cm, and the picture of all lines was taken 110 days after transplantation. The T$_0$ seedlings of other edited plants and wild-type PPR1 were refined in the greenhouse of Institute of Genetics and Developmental Biology in Beijing, and then transplanted to the experimental fields of Institute of Genetics and Developmental Biology in Hainan.

The grain length was measure by the Scan Marker i560 (MICROTEK). The cross section photos of the stems were taken by stereomicroscope (OLYMPUS, SZX16) and the diameter of the third internodes of primary tillers were measured by vernier caliper. To observe abscission cells, the histochemical analysis was performed based on methods in Yoon et al. (2014) with some modifications. Samples were fixed in formaldehyde–acetic acid–alcohol solution. The tissue-paraffin block was cut to a thickness of 8 μm with a microtome (Leica Microsystems, RM2145). The cut samples were attached to a coated slide with sterile water on 42°C hot-plate, rehydrated with 100% xylene 10 min twice for clearing wax, and washed in 50% xylene (xylene in ethanol) and ethanol series (100, 100, 95, 85, 70, 50 and 30%). The samples were stained with phloroglucinol (saturated phloroglucinol in 20% HCl) until sections turned red. The sections were covered with acidic glycerol and then observed under a microscope (Leica, DMR).

RNA extraction and quantitative real-time PCR analysis

The total RNA of stem base was extracted using a TRIzol Kit by three biological repeats (Invitrogen, Cat#15596018) according to the user's manual. About 1.5 μg of each RNA sample was used for cDNA synthesis (Maxima H Minus cDNA Synthesis Master Mix, with dsDNase, Thermo Scientific, Cat#M1682). Quantitative real-time PCR was performed using primer pairs (see KEY RESOURCES TABLE) in the SsoFast EvaGreen Supermix (Bio-Rad, Cat#1725201) with the real-time PCR detection system (Bio-Rad, CFX96). PCR reactions were performed in quadruplicate for each sample, and expression levels were normalized to *OaUBI* for expression detection. The experiments were repeated independently three times.

Cell
Article

QUANTIFICATION AND STATISTICAL ANALYSIS

The statistical details of experiments can be found in the figure legends, including the statistical tests used, exact value of n, what n represents, definition of center, and dispersion measures. Comparisons for two groups were calculated using the unpaired two-tailed Student's t tests, and statistical analysis was performed using Microsoft Excel 2016. Results are represented as mean ± s.d.. P values < 0.01 were considered significant, with levels of significance as follows: **p < 0.01; ns, not significant. We observed normal distribution and no difference in variance between groups in individual comparisons. The selection of sample size was based on extensive experience. Further methods to determine whether the data met assumptions of the statistical approach were not relevant for these analyses.

Supplemental Figures

A

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Callus induction level	Good	Modest	Good	Low	Modest	Modest	Modest	Modest	Low	Unable	Modest	Modest	Low	Modest
Whiten shoot ratio	8/56	9/56	4/56	4/56	1/56	4/56	14/56	0/56	8/56		0/56	0/56	2/56	0/56
Green shoot ratio	14/56	0/56	10/56	1/56	1/56	1/56	3/56	0/56	1/56		0/56	0/56	2/56	0/56

ID	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Callus induction level	Unable	Unable	Modest	Unable	Unable	Low	Modest	Unable	Low	Unable	Modest	Low	Low	Modest
Whiten shoot ratio			0/56			2/56	6/56		2/56		14/56	0/56	4/56	4/56
Green shoot ratio			0/56			3/56	0/56		0/56		0/56	0/56	1/56	4/56

B

C

D

E

Experiment	1	2	3	4
Agrobacterium-infected calli	156	122	183	133
Hygromycin resistant calli	123	96	147	120
PCR-positive plants	50	41	56	49
Transformation efficiency (%)	78.8	78.7	80.3	90.2
Regeneration efficiency (%)	40.7	42.7	38.1	40.8

Figure S1. Transformation and regeneration efficiencies of allotetraploid rice species, related to Figure 1

(A) Callus induction and regeneration of 28 allotetraploid rice accessions. Detail information of each ID was listed in Table S1. The green shoot ratio or whiten shoot ratio is labeled as the number of calli producing green shoot or whiten shoot divided by number of individual callus transferred to regeneration medium. (B) Hygromycin resistant calli of PPR1. (C) Regenerated plants of PPR1. (D) Detection of the vector in transformed plants. (E) Transformation and regeneration efficiencies of PPR1 in four experiments. The transformation efficiencies are calculated as the number of hygromycin resistant calli divided by the number of *Agrobacterium*-infected calli. The regeneration efficiencies are calculated as the number of calli producing plants divided by the number of individual callus transferred to regeneration medium.

A

B

	Read insert size (bp)	Read length (bp)	Number of Raw bases (bp)	GC content (%)	Sequence Coverage (X)
Illumina reads	350	PE150	85,180,308,300	42.65	86.79
Pacbio reads	-	12717.45	121,126,051,769	44.64	123.42
Bionano reads	-	-	287,173,000,000	-	292.62
Hi-C reads	-	PE150	103,222,897,500	44.01	105.18
Total	-	-	596,702,257,569	-	608.01

C

D
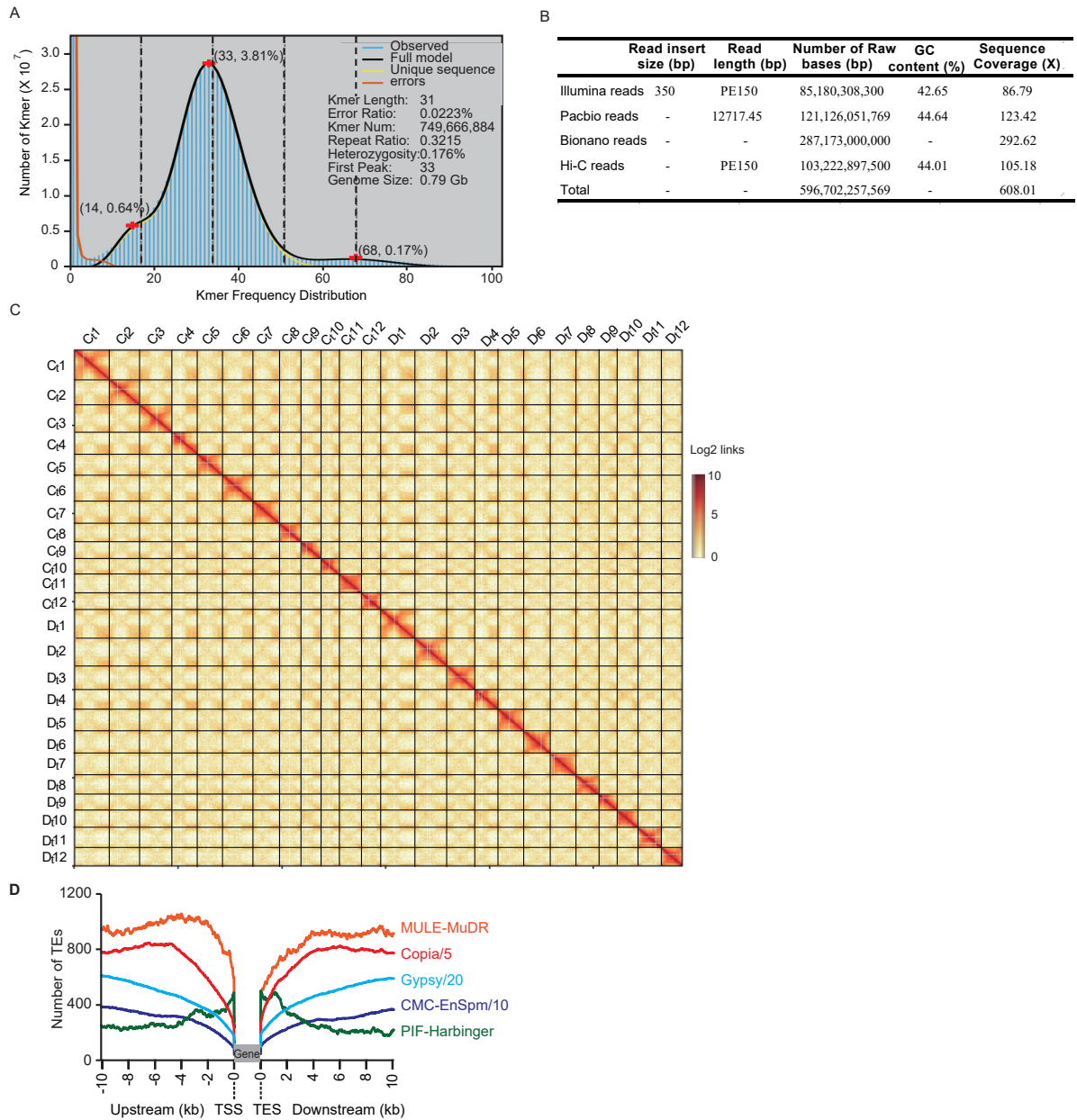
Figure S2. Genome estimation and assembly of *O. alta*, related to Table 1

(A) Distribution of 31-mer in clean sequence data of the *O. alta* genome. The x axis and y axis indicate the 31-mer frequency and number, respectively. (B) Reads information used in the *de novo* sequencing of *O. alta* genome. (C) Heatmap of the *O. alta* genome using Hi-C technology. The heatmap represents the normalized contact matrix. (D) Frequencies of transposable elements (TE) in the vicinity of genes. The plots are anchored around the transcription start sites (TSS) and end sites (TES) of protein coding genes.

A

Sample	Clean reads (Quality >= 30)	Mapped reads	Alignment rate (%)
Young Leaf Replicate 1	96,366,950	90,094,205	93.49
Young Leaf Replicate 2	74,684,778	69,861,718	93.54
Young Leaf Replicate 3	80,645,798	75,454,525	93.56
Young Root Replicate 1	97,824,372	91,922,778	93.97
Young Root Replicate 2	94,570,876	89,013,925	94.12
Young Root Replicate 3	96,308,924	90,397,032	93.86
Young Stem Replicate 1	85,846,702	79,982,113	93.17
Young Stem Replicate 2	84,979,396	79,154,609	93.15
Young Stem Replicate 3	93,218,874	87,092,067	93.43
Mature Leaf Replicate 1	85,284,758	79,204,353	92.87
Mature Leaf Replicate 2	94,615,010	88,459,004	93.49
Mature Leaf Replicate 3	85,192,094	79,016,391	92.75
Mature Root Replicate 1	91,275,322	85,094,901	93.23
Mature Root Replicate 2	101,542,764	93,639,110	92.22
Mature Root Replicate 3	97,920,270	90,436,508	92.36
Mature Shoot Replicate 1	98,911,458	92,450,609	93.47
Matur e Shoot Replicate 2	85,682,222	80,202,175	93.60
Mature Shoot Replicate 3	80,566,794	75,225,640	93.37
Seed Replicate 1	90,512,106	81,648,965	90.21
Panicle Replicate 1	69,380,680	63,646,339	91.73
Sheath Replicate 1	92,508,602	85,481,604	92.40
Sheath Replicate 2	90,250,422	83,732,580	92.78
Sheath Replicate 3	91,986,304	85,343,873	92.78
Callus Replicate 1	100,380,762	90,013,457	89.67
Callus Replicate 2	96,299,700	86,864,742	90.20
Callus Replicate 3	96,364,422	90,143,049	93.54

B

C

Length coverage (C)	Transcripts number	Percentage (%)
Mapped	435,449	90.16
C=100%	357,035	73.92
90%<=C<100%	62,296	12.90
80%<=C<90%	6,097	1.26
70%<=C<80%	3,346	0.69
60%<=C<70%	2,464	0.51
50%<=C<60%	2,331	0.48
0<C<50%	1,880	0.39
Unmapped	47,548	9.84
Total	482,997	100.00

Note: Sequence identity set as at least 99%.

D

E

Database	Total transcripts number	Percentage (%)
NR	75,233	75.75
Swissprot	60,036	60.45
InterPro	60,835	61.26
COG	33,429	33.66
GO	45,313	45.63
KEGG	77,863	78.40
Total*	81,421	81.99

* Number of transcripts annotated in the genome.

Figure S3. Transcriptome data and annotation of *O. alta*, related to Table 1
(A) Statistics of transcriptome data. (B) Principal component analysis of transcriptome data. These tissues were apparently separated and three repeats for each tissue have a relatively concentrated distribution. YL, young leaf; ML, mature leaf; YS, young stem; YR, young root; MR, mature root; MS, mature shoot; SE, seed; PA, panicle; SH, sheath; CA, callus. (C) *De novo* transcripts mapped to the genome assembly. (D) Summary of the genes predicted by three algorithms. (E) Statistics of transcripts annotated by different databases.
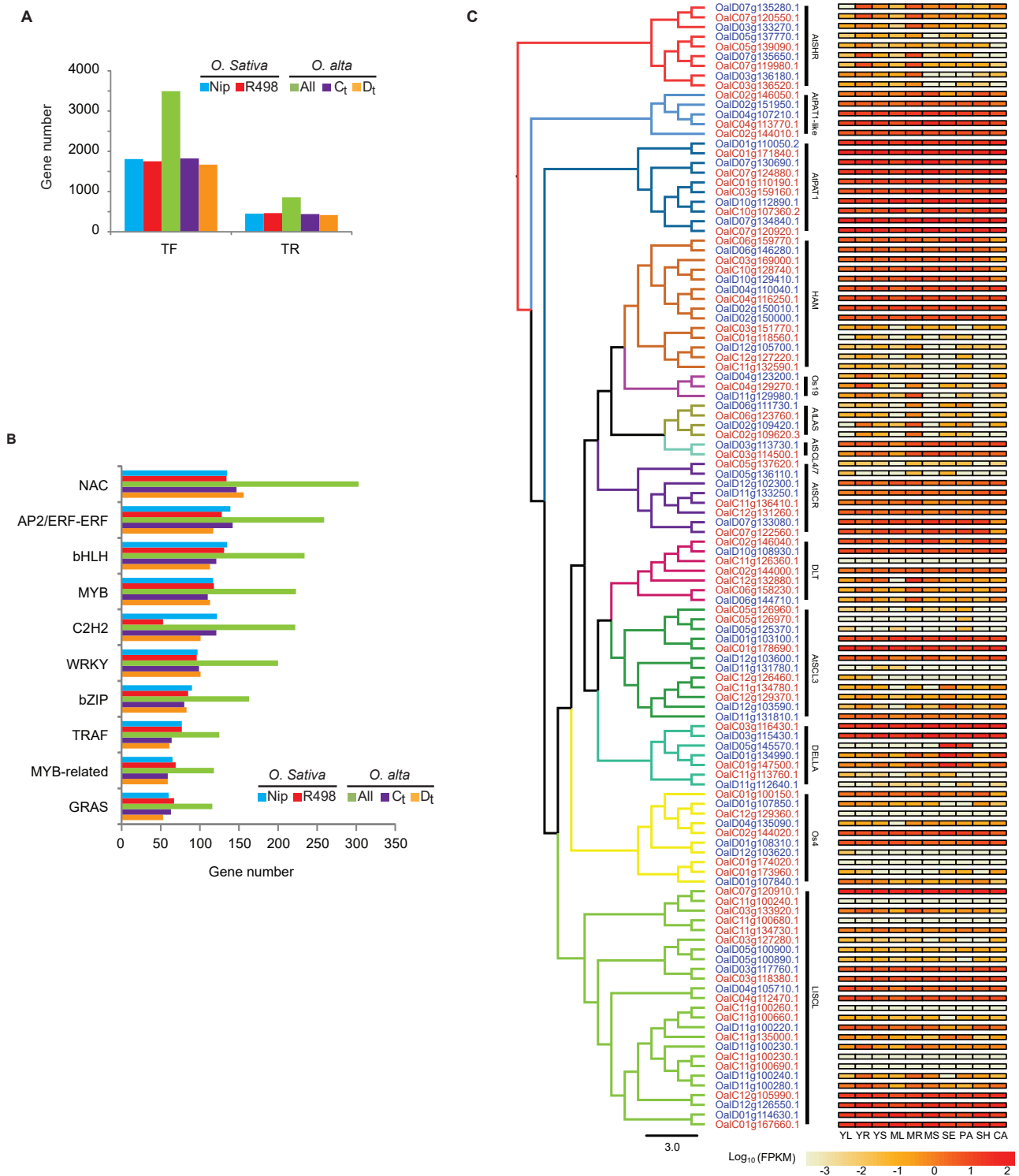
Figure S4. Transcription factors and regulators in *O. alta*, related to Figure 2

(A) Gene number of transcription factors and regulators in *O. alta* and *O. sativa* ssp. *indica* (R498) and *japonica* (Nip) genomes. (B) Gene number of ten families of transcription factors and regulators in *O. alta* and *O. sativa* ssp. *indica* (R498) and *japonica* (Nip) genomes. (C) GRAS family in *O. alta* and their expression in different tissues. Different colors on phylogenic tree indicate different GRAS families. Red and blue gene ids indicate the genes on C_t and D_t subgenomes, respectively. YL, young leaf; ML, mature leaf; YS, young stem; YR, young root; MR, mature root; MS, mature shoot; SE, seed; PA, panicle; SH, sheath; CA, callus.
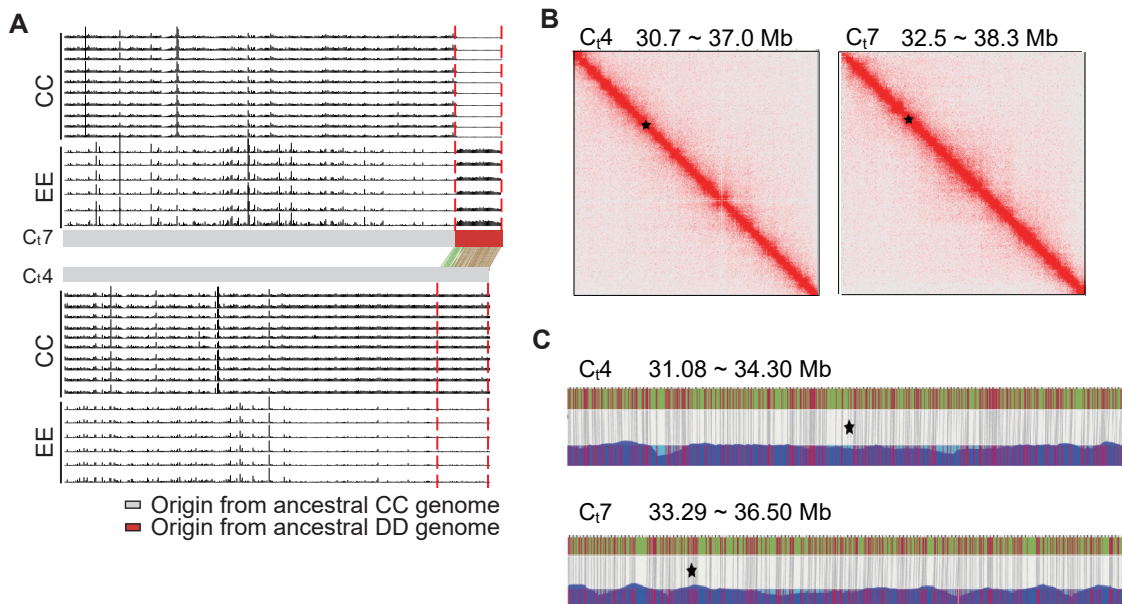
Figure S5. Chromosomal rearrangements in the *O. alta* genome, related to Figure 3

(A) Genome rearrangements between *O. alta* chromosome C_t4 and C_t7. Grey and red boxes represent the regions origin from the ancestral CC and DD genomes, respectively. Back lines represent the sequence depth of resequencing data from ten CC and six EE diploid rice accessions. Color lines between two chromosomes represent collinear blocks between *O. alta* chromosomes C_t4 and C_t7.

(B) Chromatin interactions at 150-kb resolution by Hi-C data. The black pentagrams show the borders of the large segmental translocation between *O. alta* chromosomes C_t4 and C_t7.

(C) Collinearity between genome sequences by BioNano maps. The upper horizontal green bars represent the genome sequences and the below blue bars represent the BioNano maps. The lines between green bars and blue bars represent the collinear restriction sites. Black pentagrams in (B) and (C) show the borders of the large segmental translocation between *O. alta* chromosomes C_t4 and C_t7.

A

	OaSD1-CC		OaSD1-DD	
	sgRNA-17	sgRNA-6	sgRNA-17	sgRNA-6
sd1^{CR}-2	+1/-128	-5/-128	WT/WT	+1/-1/+1-1
sd1^{CR}-3	+1/-128	+1/-128	WT/WT	+1/+1-2
sd1^{CR}-4	+1/-128	-2/-128	WT/WT	+1/-1/+1-1
sd1^{CR}-5	+1/+3-5	-1/+1	WT/WT	-1/+1-1
sd1^{CR}-6	+1/+1	+1/-2-1	WT/WT	-1/WT

B

	OaGS3-CC	OaGS3-DD
	sgRNA-7	sgRNA-7
gs3^{CR}-2	-1/-1	+1/-1
gs3^{CR}-3	+2/+2	+1/+1
gs3^{CR}-4	WT/+1	+1/+1
gs3^{CR}-5	+1/+1	+2/+1
gs3^{CR}-6	-1/-1	+1/-1

C

PPR1   sd1^{CR}-2   sd1^{CR}-3   sd1^{CR}-4   sd1^{CR}-5   sd1^{CR}-6

D

E

F
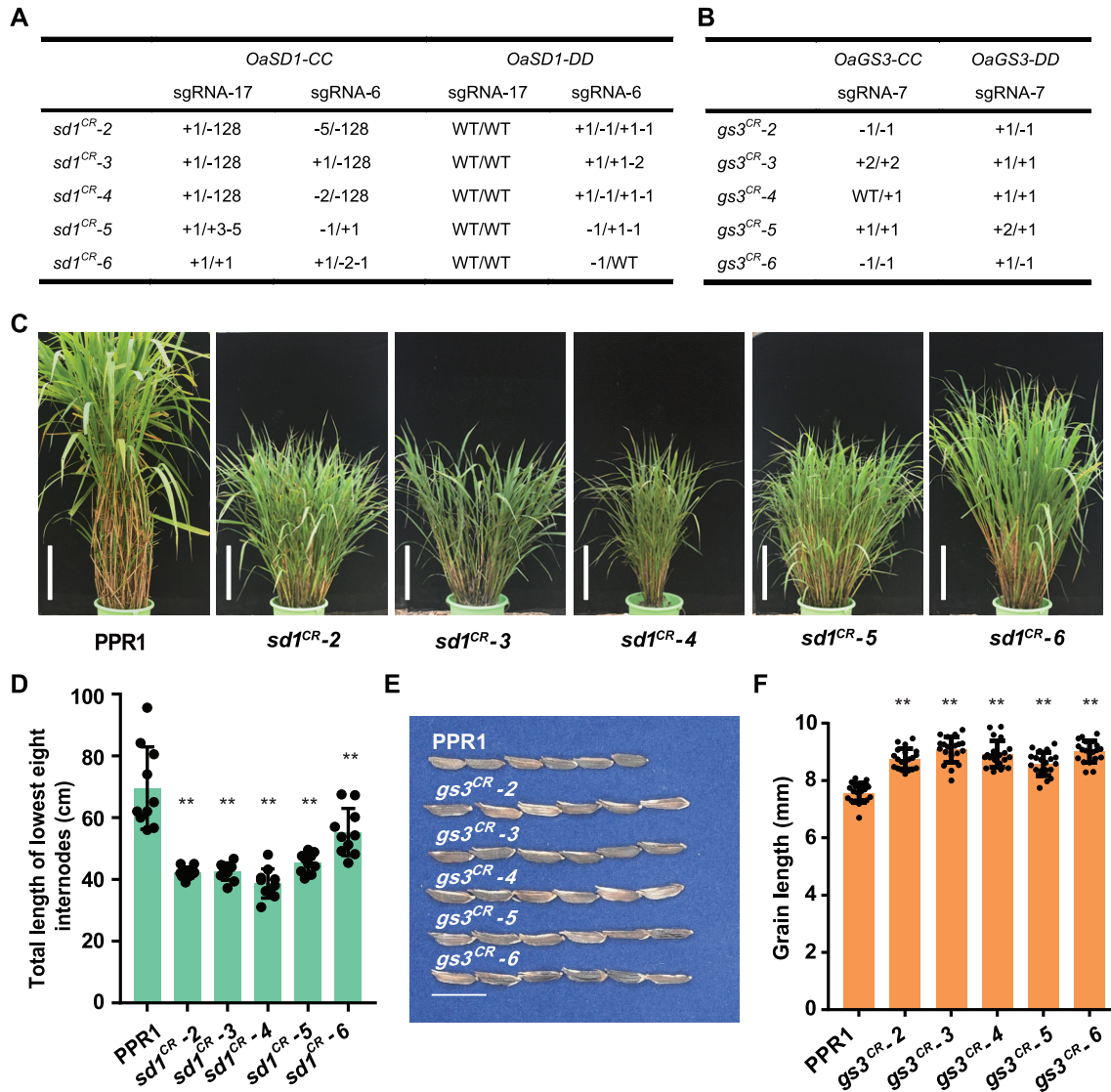
Figure S6. Genome editing of *O. alta* homologs of *SD1* and *GS3*, related to Figure 5

(A) CRISPR/Cas9-induced mutations in sd1^{CR}-2, sd1^{CR}-3, sd1^{CR}-4, sd1^{CR}-5, and sd1^{CR}-6 plants. (B) CRISPR/Cas9-induced mutations in gs3^{CR}-2, gs3^{CR}-3, gs3^{CR}-4, gs3^{CR}-5, and gs3^{CR}-6 plants. (C) Comparison of the total length of lowest eight internodes among PPR1, sd1^{CR}-2, sd1^{CR}-3, sd1^{CR}-4, sd1^{CR}-5, and sd1^{CR}-6 plants. Scale bars, 30 cm. (D) Statistical analysis of the data in (C). Data are means ± s.d. (n = 10 tillers). Asterisks represent significant differences for plant heights determined by Student's t test. **p < 0.01. (E) Comparison of grain lengths among PPR1, gs3^{CR}-2, gs3^{CR}-3, gs3^{CR}-4, gs3^{CR}-5, and gs3^{CR}-6 plants. Scale bars, 1 cm. (F) Statistical analysis of the data in (E). Data are means ± s.d. (n = 20 grains). Asterisks represent significant differences determined by Student's t test. **p < 0.01.

A
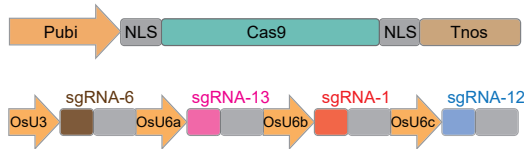
B

| | sgRNA-6 | | sgRNA-13 | | sgRNA-1 | | sgRNA-12 | |
	OaSD1-CC	OaSD1-DD	OaWX-CC	OaWX-DD	OaqSH1-CC	OaqSH1-DD	OaGn1a-CC	OaGn1a-DD
LHN2001	+2/+2	WT/WT	-1/WT	WT/WT	+1/WT	+1/WT	WT/WT	WT/WT
LHN2002	-1/+1	-2/+1	WT/WT	WT/WT	+1/WT	+1/WT	WT/WT	WT/WT
LHN2003	+2/-15	+1/+2-1	WT/WT	WT/WT	+1/-1	+1/+1	WT/WT	WT/WT
LHN2004	WT/+1	WT/-1	+1/WT	+1/WT	+1/WT	+1/WT	WT/WT	WT/WT
LHN2005	WT/WT	WT/-1	WT/WT	WT/WT	+1/WT	+1/WT	WT/WT	WT/WT
LHN2006	+1/+1	WT/+4-2	WT/WT	WT/WT	-3/WT	+1/WT	WT/WT	WT/WT
LHN2007	-1/-1	WT/+4-2	-6/WT	-1/-1	-3/WT	+1/WT	WT/WT	+1/WT
LHN2008	WT/WT	WT/+3	WT/WT	WT/WT	+1/+1	+1/WT	WT/WT	+1/WT
LHN2009	+1/+1	WT/+3	WT/WT	WT/WT	+1/WT	+1/WT	WT/WT	+1/WT
LHN2010	WT/WT	WT/WT	WT/WT	WT/WT	+1/+1	+1/+1	WT/WT	WT/WT
LHN2011	-1/+1-2	+1/-1	WT/WT	WT/WT	+1/+1	+1/WT	WT/WT	WT/WT
LHN2012	+1/+2-1	+1/-1	WT/WT	WT/WT	+1/+1	+1/WT	WT/WT	WT/WT
LHN2013	-1/-27	-1/-9	WT/WT	-1/-82	+1/-1	+1/WT	WT/WT	-1/-1
LHN2014	-37/-37	-31/-31	-1/WT	+1/WT	+1/+1	+1/WT	WT/WT	WT/WT
LHN2015	-1/-27	-1/-1	WT/WT	-1/WT	WT/+1	+1/WT	WT/WT	WT/WT
LHN2016	-1/-27	-1/WT	WT/WT	-1/WT	+1/WT	+1/WT	-1/WT	WT/WT
LHN2017	-1/-1	-1/-1	WT/WT	+4-3/WT	+1/+1	+1/WT	+1/+1	WT/WT
LHN2018	-1/WT/	-1/WT	WT/WT	WT/WT	+1/+1	+1/WT	WT/WT	WT/WT
LHN2019	WT	+1/+2	WT/WT	WT/WT	+1/WT	WT/WT	WT/WT	WT/WT
LHN2020	-1/-1	-1/+1	+1/+1	-1/WT	+1/+1	+1/WT	WT/WT	-1/-1

C

| | sgRNA-13 | | sgRNA-3 | | sgRNA-5 | | sgRNA-15 | | sgRNA-2 | | sgRNA-14 | | sgRNA-12 | | sgRNA-16 | |
	OaWx-CC	OaWx-DD	OaAn-1-CC	OaAn-1-DD	OaSD1-CC	OaSD1-DD	OaTAC1-CC	OaTAC1-DD	OaqSH1-CC	OaqSH1-DD	OaGW5-CC	OaGW5-DD	OaGn1a-CC	OaGn1a-DD	OaBh4-CC	OaBh4-DD
LHN2021	WT/WT	-1/WT	+1/+1	-1/WT	+1-2/+1-2	+1/+1	WT/WT	WT/WT	-1/-1	+1/WT	-18/-18	+5/WT	+1/WT	+5/+10-6	-7/+1	+1/+1
LHN2022	-1/WT	WT/WT	+1/+1	+1/WT	+1/+1	+1/+1	WT/WT	WT/WT	+1/+1-3	-1/+1	+1/+1	WT/WT	WT/WT	WT/WT	-4/+1	+1/+1
LHN2023	-1/WT	WT/WT	+1/+1	WT/+12-8	+1/+1	+1/+1	WT/WT	WT/WT	+1/+1-3	+1/+1	+1/+1	-27/-27	WT/WT	WT/WT	-4/+1	+1/+1
LHN2024	-1/WT	WT/WT	+1/+1	+1/-1	+1/+1	+1/+1	WT/WT	WT/WT	+1/+1-3	-1/+1	+1/+1	+1/WT	WT/WT	WT/WT	-4/+1	+1/+1
LHN2025	-1/WT	WT/WT	+1/+1	+1/+1	+1/+1	+1/+1	WT/WT	WT/WT	+1/+1-3	-1/+1	+1/+1	WT/WT	WT/WT	WT/WT	-4/+1	+1/+1
LHN2026	-1/WT	WT/WT	+1/+1	-8/-10	+1/+1	+1/+1	WT/WT	WT/WT	WT/WT	+1/+1	+1/+1	-27/-27	WT/WT	WT/+3-4	-4/+1	+1/+1
LHN2027	-1/WT	WT/WT	+1/+1	-8/-9	+1/+1	+1/+1	WT/WT	WT/WT	+1/+1-3	-1/+1	+1/+1	WT/WT	WT/WT	WT/WT	-4/+1	+1/+1
LHN2028	-1/WT	WT/WT	+1/+1	WT/WT	+1/+1	+1/+1	WT/WT	WT/WT	+1/+1-3	-1/+1	+1/+1	+1/WT	WT/WT	WT/WT	-4/+1	+1/+1
LHN2029	-1/WT	WT/WT	+1/+3-3	-8/-10	+1/+1	+1/+1	WT/WT	WT/WT	+1/+1-3	-1/+1	+1/+1	WT/WT	WT/WT	WT/WT	-4/+1	+1/+1
LHN2030	-1/WT	-1/WT	+2/+2	-8/-10	+1/+1	+1/+1	WT/WT	WT/WT	+1/+1-3	-1/+1	+1/+1	+1/-2	WT/WT	WT/WT	+1/+1	+1/+1
LHN2031	-1/WT	WT/WT	+1/+1	-8/-10	+1/+1	+1/+1	WT/WT	WT/WT	+1/+1-3	-1/+2	+1/+1	+1/+1-2	WT/WT	WT/WT	+1/+1	+1/+1
LHN2032	WT/WT	WT/WT	+1/+1	+1/+1	+1+1/+1+1	WT/+3-3	WT/WT	WT/WT	-1/-1	+1/+2	WT/WT	+2/WT	WT/WT	-1/-1	+1/-8	-2/+1
LHN2033	WT/WT	WT/WT	+1/+1	+1/+1	+1+1/+9-6	WT/+1	WT/WT	WT/WT	-1/-1	+1/+2	-18/-18	-5/WT	WT/WT	+5/+10-6	+1/-8	-2/+1
LHN2034	-1/WT	WT/WT	+1/+1	-8/-10	+1/+1	+1/+1	WT/WT	WT/WT	+1/+1-3	-1/+1	+1/+1	+1/WT	WT/WT	WT/WT	-5/+1	+1/+1
LHN2035	-1/WT	WT/WT	+1/+1	-8/-10	+1/+1	+1/+1	WT/WT	WT/WT	+1/+1-3	-1/+1	+1/+1	WT/WT	WT/WT	WT/WT	-5/+1	WT/WT
LHN2036	-1/WT	WT/WT	+1/+1	-8/-10	+1/+1	+1/+1	WT/WT	WT/WT	+1/+1-3	-1/+1	+1/+1	WT/WT	WT/WT	WT/WT	-5/+1	+1/+1
LHN2037	-1/WT	WT/WT	+4/+3	-8/-10	+1/+1	+1/+1	WT/WT	WT/WT	+1/+1-3	-1/+1	+1/+1	WT/WT	WT/WT	WT/WT	+1/+1	+1/+1
LHN2038	-1/WT	WT/WT	+1/+1	-8/-10	+1/+1	+1/+1	WT/WT	WT/WT	+1/+1-3	-1/+1	+1/+1	+1/WT	WT/WT	WT/WT	-5/+1	+1/+1
LHN2039	-1/WT	WT/WT	+1/+1	-8/-9	+1/+1	+1/+1	WT/WT	WT/WT	+1/+1-3	-1/+1	+1/+1	-1/-1	WT/WT	WT/WT	+1/-5	+1/+1
LHN2040	-1/WT	WT/WT	+1/+1	-8/-9	+1/+1	+1/+1	WT/WT	WT/WT	+1/+1-3	-1/+1	+1/+1	WT/WT	WT/WT	WT/WT	+1/-5	WT/WT
LHN2041	-8/+1	-3/+1	WT/WT	WT/+1	WT/WT	WT/WT	WT/WT	WT/WT	+1/WT	+3-2/+1	+1/+1	+1/+1	+1/WT	WT/WT	WT/WT	+1/WT
LHN2042	-1/+3-10	-2/WT	+2/+3-1	+1/+1	-27/-27	+1/-1	+1/+1	-1/+1	+2/-1	+1/WT	-3/-3	+1/+1	WT/WT	WT/WT	-36/WT	+1/-5
LHN2043	WT/WT	-1/WT	-1/+3-2	-1/+3-2	+1/+1	+1/+1	+1/WT	WT/WT	-1/-1	-1/+1	-42/-42	+1/WT	+1/WT	WT/WT	+1/+1	-1/-33
LHN2044	WT/WT	-1/WT	-1/-1	+1/+1	WT/-3	+1/WT	WT/WT	-1/+1	-1/-1	-1/+1	WT/WT	-5/WT	WT/WT	WT/WT	-1/-1	-1/WT
LHN2045	WT/WT	-1/WT	-1/-1	+1/+1	-3/-1	+1/WT	WT/WT	WT/WT	+2/-1	+1/WT	-1/WT	+1/WT	WT/WT	WT/WT	-1/-1	-1/WT
LHN2046	WT/WT	-1/+3-2	+1/-19	+1/+2-1	+1/-1	WT/WT	WT/WT	-1/-1	+1/WT	-42/-42	-1/WT	WT/WT	WT/WT	+1/+1	-1/-32	
LHN2047	WT/WT	-1/-2	-1/-1	+1/+1	-3/-1	+1/-1	WT/WT	WT/WT	+1/WT	+1/WT	-1/-1	-5/WT	WT/WT	WT/WT	-1/-1	-1/WT
LHN2048	-1/WT	-1/WT	-1/+3-2	+1/+1	+1/+2	+1/+1	WT/WT	WT/WT	+1/WT	-1/+2	-1/-1	+1/WT	+1/WT	WT/WT	+1/+1	-1/-32
LHN2049	WT/WT	WT/WT	+2/+2	+2/-7	+1/+1	+1/+1	WT/WT	WT/WT	+2/WT	-1/+1	+1/+1	-7/WT	+1/WT	+1/WT	+1/-1	+1/-1
LHN2050	WT/WT	WT/WT	+2/+2	+2/-7	+1/+1	+1/+1	WT/WT	WT/WT	+2/WT	-1/+1	+1/+1	-22/WT	WT/WT	WT/WT	+1/-1	+1/+1
LHN2051	-1/WT	WT/WT	+2/+2	+2/-7	+1/+1	+1/+1	WT/WT	WT/WT	+1/WT	-1/+1	+1/+1	-7/WT	WT/WT	WT/WT	+1/-1	+1/-1
LHN2052	WT/WT	WT/WT	+2/+2	+2/-7	+1/+1	+1/+1	+1/WT	WT/WT	WT/WT	-1/+1	+1/+1	+1/WT	WT/WT	WT/WT	+1/-1	+1/-1
LHN2053	WT/WT	+1/+2	-1/+2	+1/+1-2	+2/WT	+1/-1	WT/WT	WT/WT	-2/+1	-1/+1	WT/WT	+3-3/WT	WT/WT	WT/WT	+1/+1	+1/WT
LHN2054	WT/WT	-1/WT	+1/+1	+1/+1-2	+1/+1	+1/-15	WT/WT	WT/WT	-1/WT	-1/+1	+1/+1	+1/+1	WT/WT	WT/WT	-1/-1	+1/WT
LHN2055	-1/+2-6	-5/-5	WT/WT	+1/+1	-4/+15-8	+1/-2	+1/WT	WT/WT	+1/WT	-1/+1	+1/+1	WT/WT	WT/WT	WT/WT	WT/WT	WT/WT
LHN2056	-3/+3	-1/-1	WT/WT	+1/+1	-30/-30	-1/+1	WT/WT	WT/WT	WT/WT	+1/WT	WT/WT	WT/WT	+1/WT	WT/WT	WT/WT	WT/WT

Figure S7. Multiplex genome editing in *O. alta*, related to Figure 6

(A) Structure of the multiplex genome editing system with four sgRNAs targeting eight genes in PPR1. Pubi, the promoter of maize ubiquitin; NLS, nuclear localization sequence; Tnos, the terminator of nopaline synthase gene.

(B) Genotypes of 20 mutant lines transformed with the vector in (A).

(C) Genotypes of 36 mutant lines transformed with the vector in Figure 6(H).