Article:
Zhan, C, Lei, L, Liu, Z et al. (26 more authors) (2020) Selection of a subspecies-specific diterpene gene cluster implicated in rice disease resistance. Nature Plants. 1447–1454. ISSN 2055-026X

https://doi.org/10.1038/s41477-020-00816-7

Selection of a subspecies-specific diterpene gene cluster implicated in rice disease resistance

Chuansong Zhan[1], Long Lei[2,1], Zixin Liu [1],Shen Zhou[1], Chenkun Yang[1], Xitong Zhu[1], Hao Guo[2,1], Feng Zhang[1], Meng Peng[1], Meng Zhang[1], Yufei Li[1], Zixin Yang[1], Yangyang Sun[1], Yuheng Shi[1], Kang Li[1], Ling Liu[2], Shuangqian Shen[1], Xuyang Wang[1], Jiawen Shao[1], Xinyu Jing[1], Zixuan Wang[1], Yi Li[3], Tomasz Czechowski[3], Morifumi Hasegewa[4], Ian Graham[3], Takayuki Tohge[5], Lianghuan Qu[1], Xianqing Liu[2,1], Alisdair R. Fernie[6], Ling-Ling Chen[1], Meng Yuan[1], Jie Luo[2,1]✉

[1]National Key Laboratory of Crop Genetic Improvement and National Center of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan 430070, China.

[2]College of Tropical Crops, Hainan University, Haikou 572208, China.

[3]Centre for Novel Agricultural Products, Department of Biology, University of York, York, United Kingdom.

[4]College of Agriculture, Ibaraki University, 3-21-1 Chuo, Ami, Ibaraki 300-0393, Japan.

[5]Graduate School of Biological Sciences, Nara Institute of Science and Technology, Ikoma, Nara, 630-0192, Japan.

[6]Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany.

✉e-mail: jie.luo@hainanu.edu.cn

Diterpenoids are the major group of antimicrobial phytoalexins in rice[1,2]. Here we report the discovery of a *rice diterpenoid gene cluster on chromosome* 7 (*DGC7*) encoding the entire biosynthetic pathway to 5,10-diketo-casbene, a member of the mono-cyclic casbene-derived diterpenoids. We revealed that *DGC7* is regulated through MeJA mediated epigenetic control directly by JMJ705[3]. Functional characterization of pathway genes revealed *OsCYP71Z21* to encode a casbene C10 oxidase, sought after for the biosynthesis of an array of medicinally important

diterpenoids. We further reveal that *DGC7* arose relatively recently in the *Oryza* genus, that it was partly formed in *O. rufipogon* and positively selected for in *japonica* during domestication. Casbene synthesizing enzymes that are functionally equivalent to OsTPS28 are present in several species of Euphorbiaceae but gene tree analysis shows that these and other casbene-modifying enzymes have evolved independently. As such, combining casbene-modifying enzymes from these different families of plants may prove effective in producing a diverse array of bioactive diterpenoid natural products.

The terpenoid class of specialized metabolites are important in the adaptation of plants to their ecological niches, as well as serving as a valuable medicinal resource[2,4,5]. Enzymes involved in the metabolic pathways of distinct terpenoid classes are encoded by gene clusters in a range of plant species[6]. For example, the phytocassane metabolic gene cluster in rice confers resistance to the fungal pathogen *Magnaporthe oryzae* (*M. oryzae*) and the bacterial pathogen *Xanthomonas oryzae* pv *oryzae* (*Xoo*), the cucurbitacins of *Cucumis sativus* can be used as traditional medicines and the thalianol gene cluster in *Arabidopsis thaliana* may modulate Arabidopsis root microbiota[2,4,5]. The C-20 diterpene class of terpenoids can be further subdivided into a large superfamily of labdane-related diterpenoids which include the gibberellins and are defined by an initial dual cyclisation of geranylgeranyl diphosphate (GGDP) and others

including casbene-type diterpenoids which are formed by mono-cyclisation of GGDP[7]. Casbene-type diterpenoids are found predominantly in the Euphorbiaceae family and are recognized as being rich in a range of pharmacological activities[8-11], consistent with their widespread use in traditional medicine around the world[12]. By contrast in the Poaceae, to date, this type of diterpenoid has only been reported in rice[13,14].

5,10-diketo-casbene (previously referred to as *ent*-10-oxodepressin) was the first casbene-type diterpene phytoalexin found in rice (*Oryza sativa*) that confers rice bacterial blight and rice blast fungus resistance[13-16]. However, surprisingly to date no study has yet assessed the natural variation in the ability to produce 5,10-diketo-casbene. More than 4,000 diverse accessions of *O. sativa (indica and japonica)* and various wild rice relatives have been sequenced in recent years allowing the generation of a detailed genome-variation map[17-19] and the opportunity to perform Genome Wide Association Studies to locate the genetic basis of traits exhibiting natural variation.

Here, we report that the locus responsible for the biosynthesis of 5,10-diketo-casbene from GGDP encodes an epigenetically regulated gene cluster that includes *Oryza* genus-specific terpene synthase and cytochrome P450 oxidases (CYP450) that have been specifically selected in *japonica* during domestication. Combining biochemical analyses with rice population and evolutionary genetics, we have provided insights into

the epigenetic regulation, structural variation, and origin of eukaryotic metabolic gene cluster and clarified its evolutionary history from a systematic analysis of population.

To determine the extent of variation of casbene-type diterpenes in rice, we collected leaf samples of 424 rice (*O. sativa*) accessions from a diverse worldwide resource panel **(**Extended Data Fig. 1 and Supplementary Table 1)[20]. A metabolite-based genome-wide association study (mGWAS) was performed for both the full population (all 424 accessions) and each of the two subpopulations, *indica* (271 accessions) and *japonica* (132 accessions), independently (Fig. 1a, Supplementary Fig. 1 and Supplementary Table 2). The association results showed that natural variation in 5,10-diketo-casbene of *japonica* rice was mainly controlled by a locus on chromosome 7 (Fig. 1a and Supplementary Fig. 1a, b). *OsTPS28* (*Os07g11790*), the only terpene gene within this locus, was chosen as a candidate for the diterpene synthase and four putative CYP450 genes (*OsCYP71Z2*, *Os07g11739*; *OsCYP71Z21*, *Os07g11870*; *OsCYP71Z30*, *Os07g11890*; *OsCYP71Z22*, *Os07g11970*) were candidates for the oxidation of casbene to produce 5,10-diketo-casbene (Fig. 1a and Supplementary Fig. 1 a-d). Together these gene candidates represented a putative diterpene gene cluster across a 140kb region hereafter referred to as *Diterpene Gene Cluster on chromosome 7*, *DGC7*.

The absence of the association signal on chromosome 7 in *indica*

panel compared to the *japonica* panel (Fig. 1a and Supplementary Fig. 1), led us to perform an in-depth analysis of the corresponding region using three high-quality reference genomes (Nipponbare, Minghui63, Zhenshan97). This revealed that *OsTPS28* and *OsCYP71Z21* present in Nipponbare (*japonica* rice), but absent from Minghui63 and Zhenshan97 (*indica* rice) (Fig. 1b and Supplementary fig. 2). Analysis of the pan-genome data[18,19] identified two major types of *DGC7* (*DGC7*^{present}, *DGC7*$^{absent\ or\ incomplete}$) (Fig. 1c and Supplementary Tables 3, 4). Consistent with the presence/absence of intact *DGC7*, we observed the accumulation of 5,10-diketo-casbene in most *japonica* varieties (131/132) while this metabolite is absent in most *indica* varieties (266/271) both under control conditions and following MeJA-treatment, indicating that the presence/absence of *DGC7* determines the natural variation of 5,10-diketo-casbene in rice (Fig. 1d, e and Supplementary Table 2). Furthermore, the varieties lacking intact *DGC7* do not produce 5,10-diketo-casbene while varieties with intact *DGC7* accumulate 5,10-diketo-casbene (Fig. 1f and Supplementary Fig. 3).

To characterize the putative gene cluster, we first cloned the open-reading-frames (ORFs) of the candidate genes. The *OsTPS28* ORF was amplified by RACE included a 183bp plastid-localization transit peptide **(**Supplementary fig. 4**)** and the corresponding protein localized to plastids when transiently expressed in rice protoplasts as expected for diterpene

synthases (Fig. 2a). The *OsTPS28* ORF minus the transit peptide was expressed in *Escherichia coli* BL21 and recombinant protein produced casbene in the presence of GGDP and Mg^{2+} (Extended Data Fig. 2a and Supplementary Fig. 5, a-d), with a K_m of 5.16μM and K_{cat} of 0.0236 s^{-1} (Supplementary Fig. 6). Stable transformation analysis in rice revealed that 5,10-diketo-casbene was increased by ~1.9 fold in *OsTPS28* overexpression lines while reduced to non-detectable levels in the *OsTPS28* knockout plants (Supplementary Fig. 7).

Recombinant OsCYP71Z2 protein produced in *Saccharomyces cerevisiae* oxidized the C5 position of 10-keto-casbene to produce 5,10-diketo-casbene (Extended Data Fig. 2b and Supplementary Fig. 5, e-f). To further dissect the biosynthetic pathway to 5,10-diketo-casbene we used *Agrobacterium*-mediated transient expression in *N. benthamiana* using OsTPS28 in combination with different CYP450s present in *DGC7*. Over-expression of OsTPS28 alone led to the formation of casbene (major product) and neocembrene (minor products) (Extended Data Fig. 2c and Supplementary Fig. 5, a-d); combined expression of OsTPS28 with OsCYP71Z21 resulted in production of 10-keto-casbene (Fig. 2b and Supplementary Fig. 5g, h) while combined expression of OsTPS28, OsCYP71Z21 and OsCYP71Z2 produced 5,10-diketo-casbene (Fig. 2b, Extended Data Fig. 3, a-c and Supplementary table 5). To further verify the function of OsCYP71Z21, plant microsomes were isolated from the *N.*

benthamiana leaves that infiltrated OsCYP71Z21. *In vitro* enzyme assay using isolated microsomes showed that OsCYP71Z21 was able to converted casbene to 10-keto-casbene in the presence of NADPH (Supplementary Fig. 8a), which further supported the notion that OsCYP71Z21 encoded a casbene C10 oxidase. These results are also consistent with those obtained for the recombinant OsTPS28 produced in *E. coli* and OsCYP71Z2 produced in *S. cerevisiae* and lead us to conclude that OsTPS28 is a casbene synthase, OsCYP71Z2 is a casbene C5 oxidase and OsCYP71Z21 is a casbene C10 oxidase. Together these three enzymes produce 5,10-diketo-casbene from GGDP. However, the expression of OsCYP71Z21 in yeast (WAT11) did not lead to production of 10-keto-casbene or any other metabolite (Supplementary Fig. 8b). It is thus possible that OsCYP71Z21 is not expressed in an active form in yeast. Sequence similarity analysis was performed for OsCYP71Z2, OsCYP71Z21, OsCYP71Z22 and OsCYP71Z30. CYP71Z2 revealing 73.38%, 76.15% and 70.42% identities to OsCYP71Z21, OsCYP71Z22 and OsCYP71Z30 (Supplementary Fig. 9). Although the sequence similarities are all above 70%, the OsCYP71Z22 and OsCYP71Z30 still failed to exhibit activity with casbene as substrate (Supplementary Fig. 8c).

To explore spatiotemporal expression of the members of *DGC7*, we collected samples at different stages from different parts of rice and carried out RT-PCR analyses. The results show that genes of *DGC7* shared a very

158 similar expression patterns in rice (Extended Data Fig. 4). We therefore
159 conclude that *DGC7* is a new gene cluster that catalyzes the complete
160 biosynthesis of the casbene-type diterpene phytoalexin - 5,10-diketo-
161 casbene from the common precursor GGDP.

162 To further investigate the regulation of 5,10-diketo-casbene
163 biosynthesis we treated the aerial part of 12-day-old seedling with methyl
164 jasmonate (MeJA), a potent inducer of certain defense responses in plants.
165 RNA-Seq together with quantitative real-time PCR (qrtPCR) analyses
166 demonstrated that *OsTPS28*, *OsCYP71Z21*, *OsCYP71Z2* increased over
167 100-fold following 24h of the treatment (Figs. 2c and 3a and
168 Supplementary Table 6). H3K27me3 is an important histone modification
169 chromatin mark that is inversely correlated with gene silencing[21,22]. There
170 is evidence suggesting that repression of expression of metabolic gene
171 clusters in plants, such as in rice, Arabidopsis, maize and oat, is associated
172 with trimethylation of histone H3 lysine 27 (H3k27me3)[21,23]. Here, all three
173 genes of *DGC7* also show peaks of H3K27me3 in genome-wide
174 H3K27me3 ChIP-seq maps (Fig. 3b) and MeJA treatment resulted in
175 decreased H3K27me3 levels but increased transcript-levels of the *DGC7*
176 member genes compared to the control plants (Figs. 3a and 3c)[3]. These
177 results suggest that *DGC7* is regulated by chromatin decondensation and
178 this regulation is mediated by MeJA. Interestingly, JMJ705, a reported
179 histone demethylase is also induced by MeJA and can remove H3K27me3

from *DGC7* member genes in addition to defense-related genes[3]. Further analysis showed that the content of 5,10-diketo-casbene was increased in JMJ705 overexpression lines while decreased in RNAi plants, suggesting strongly that *DGC7* is regulated through MeJA mediated epigenetic control directly by JMJ705 (Fig. 3d, e and Supplementary Figs. 10 and 11).

Since casbene-type diterpenoids are rarely found in the Poaceae family, a phylogenomic approach has been used to investigate the evolutionary origins of all three *DGC7* members: *OsTPS28*, *OsCYP71Z2* and *OsCYP71Z21*. BLAST searches have identified all homologous sequences from the GenBank non-redundant protein database as well as additional annotation datasets from 29 draft whole genome assemblies of representative grass species including *Oryza* species of AA, BB, FF and GG genome types (Supplementary Table 7). Subsequent progressive gene tree analyses suggest all three *DGC7* genes are members of subfamilies specific to *Oryza* genus within their respective grass-specific gene families (Supplementary Figs. 12 and 13). Furthermore, the latest gene duplications giving rise to the closest paralogue pairs of OsTPS28/OsTPS2 and OsCYP71Z2/OsCYP71Z1/OsCYP71Z21-OsCYP71Z22 are likely to have occurred prior to divergence of the BB (*O. punctata*) and AA genome types (*O. sativa*) about 7 Mya (Fig. 4a, b)[24]. In the latter case, the gene duplication events that led to *OsCYP71Z2*, *OsCYP71Z1*, *OsCYP71Z21*, and *OsCYP71Z22* appear to have happened after the AA/BB genome types

diverged from the GG genome type (*O. granulate*) approximately 15 Mya (Fig. 4b)[25]. In addition, the latest gene duplication leading to the youngest paralogue pair OsCYP71Z21/OsCYP71Z22 might be within the AA genome types before the African wild rice (*O. longistaminata*) diverged from *O. sativa* approximately 2 Mya (Fig. 4b)[24]. These results have led us to conclude that all three functionally characterized members of *DGC7*, OsTPS28, OsCYP71Z2 and OsCYP71Z22 have a recent origin (~2-7 Mya) in the *Oryza* genus and the gene clustering would have happened following these gene duplication events. Therefore, we have seen the convergent evolution of casbene synthases to produce the casbene backbone and subsequent independent evolution of the P450 oxidases which gave rise to a range of different casbene-derived diterpenes in the Poaceae and Euphorbiaceae families (Supplementary Figs. 14 and Extended Data Fig. 5)[26]. Similar cases have been reported for cyanogenic glycosides and triterpenes[27,28].

Phylogenomic analyses of the CYP71Z and TPSs (terpene synthases) have also shown that OsCYP71Z2 and OsCYP71Z21-OsCYP71Z22 are products of localized gene duplications on chromosome 7, whereas OsTPS28 and its closest paralogue OsTPS2 are located on chromosome 7 and 1 respectively (Fig. 4a, b). This is very conserved in the *Oryza* genomes of all AA and BB genome types where chromosomal locations are available. Unfortunately, the chromosomal position is undefined for the

only orthologue of OsTPS28 from *O. longistaminata* as this would provide useful insight into the formation of *DGC7* gene cluster by comparing the relative position of orthologue of OsTPS28 to those of the OsCYP71Z2/OsCYP71Z21-OsCYP71Z22 genes on chromosome 7 in this species.

Finally, gene tree analyses have indicated that the intact *DGC7* is mainly restricted to cultivated rice, especially among varieties of *japonica* subspecies (Fig. 4a, b), even though individual members may be present in other wild rice species. Apart from the aforementioned sole orthologue of OsTPS28 among 15 representative whole genome assemblies among rice species, orthologues of OsCYP71Z2 can only be identified in the genomes of *O. punctata* and *O. rufipogon* whereas none can be found corresponding to OsCYP71Z21. However, intact *DGC7* with all three members present can only be identified in *O. sativa* (Fig. 4a, b). This is further demonstrated in a haplotypes survey of all three members of *DGC7* in the combined total of 435 varieties from the *japonica* and *indica* subspecies of *O. sativa* as well as *O. rufipogon*. No intact *DGC7* has been found in the 13 *O. rufipogon* varieties, even though individual components are present corresponding to all three components (Fig. 4c and Supplementary Tables 3, 4). Furthermore, intact *DGC7* is highly enriched in *japonica* varieties (102/109) compared to the *indica* varieties (13/313), suggesting the selection of *DGC7* during domestication (Fig. 4c and Supplementary Table

246 3). Results from π and F_{ST} revealed that *DGC7* was located in a selective
247 sweep region (selective sweep defined as top 5% of the length of the whole
248 genome sequence) in *japonica* but not in *indica* (Supplementary Fig. 15)[29].

249 Unlike the momilactone and phytocassane gene clusters that
250 biosynthesize common labdane-related diterpenoids and are found in both
251 *indica* and *japonica* varieties[30-33], *DGC7* biosynthesizes casbene-type
252 diterpenoids that are almost exclusively restricted to *japonica* varieties. In
253 summary, we have shown all members of *DGC7* originated in the *Oryza*
254 genus; and *DGC7* is at least partly formed in the wild rice ancestor *O.*
255 *rufipogon* and has been positively selected for in *japonica* rather than
256 *indica* during domestication. *Japonica* and *indica* rice originated in
257 Southern China and India respectively[34]. Given that Southern China has
258 been an endemic area of rice bacterial blight and 5,10-diketo-casbene
259 confers rice blast resistance[14,15,35] it can be speculated that this provided
260 the selection pressure for *DCG7* to predominate in a subspecies-specific
261 manner.

262 Considerable evidence suggests that the end-product of *DGC7* – 5,10-
263 diketo-casbene is a rice phytoalexin which has antifungal activity against
264 *M. oryzae*[1,14]. It can be induced by the rice blast fungus and furthermore
265 inhibits rice blast fungus spore germination and germ tube growth[13].
266 Moreover, overexpression of *OsCYP71Z2* (one gene member of *DGC7*) in
267 rice can enhance the resistance of rice to bacterial blight resistance[15,16].

Taken together, we suggest that *DGC7* is a gene cluster involved in rice immunity. To further validate this suggestion, *OsTPS28-OE*, *OsTPS28-KO*, and wild-type plants (Zhonghua 11) were separately inoculated with a highly virulent Chinese *Xoo* strain FuJ23 via a leaf chipping method[36-38]. The results showed that the disease areas caused by *Xoo* in the OsTPS28-OE lines were much smaller than those in the wild type plants and the *OsTPS28-KO* lines exhibited larger disease areas relative to their wild type counterparts (Supplementary Fig. 16 and Supplementary Table 8). However, this does not mean all the *japonica* varieties which contain the *DGC7* cluster are more resistant to disease than *indica* varieties. Indeed, some *indica* varieties also have high levels of disease resistance and specific genes that confer rice disease resistance in *indica* varieties have been identified. For example, *WRKY45* plays a positive role in regulating disease resistance in the *indica* varieties while playing a negative role in *japonica* varieties[37,39].

For the oxidases, casbene oxidases identified to date from the Euphorbiaceae encode C5, C5,6, C7,8 and C9-oxidases[26,40-42]. We show that OsCYP71Z21 encodes a C10 casbene oxidase activity and such represents an important step in the biosynthesis of medicinal casbene-derived diterpenoids such as tiglanes, ingenanes and daphnanes[8-11]. This discovery represents a breakthrough in the elucidation of the biosynthetic pathways to a number of drug molecules derived from the tiglane, ingenane

and daphnane classes of diterpenoids and open a door for metabolic engineering and production in heterologous hosts.

Methods

Plant materials. All plants used in this study were grown in Huazhong Agricultural University, Wuhan, Hubei Province of China. The germplasm set of 424 *O. sativa* accessions consisted of both elite and landraces varieties (Supplementary Table 1). All samples were collected and flash-freezing in liquid N_2. Later, all samples were stored at -80°C until vacuum freeze-drying. Samples were collected with two biological replicate sets at different places and the data collected from them were used to calculate H^2. The samples were then ground in a ball mill (MM 04, Retsch, GmbH, Haan, Germany) into to a fine powder. The freeze-dried samples were extracted as previously described before analysis using an LC-ESI-QQQ-MS/MS system[43].

Recombinant protein expression, purification and enzyme assay. The 5' and 3' ends of the targeted TPSs were cloned by RACE (Takara, catalog number: 634858) according to the manufacturer's directions. The full cDNAs of TPSs from Nipponbare (*O. sativa* L. spp. *japonica*) were cloned into the pGEX-*6p-1* expression vector (Novagen) with a *Glutathione*-S-transferase (*GST*). The primers listed in Supplementary Table 9. Recombinant proteins were expressed in BL21 (Novagen) as previously described[43].

The enzyme reactions in vitro assay for TPSs were performed at 37°C

312 in a total volume of 200 μl containing 200 μM substrates, 5 mM $MgCl_2$ and
313 totally 500 ng purified protein in Tris-HCl buffer (100 mM, pH = 7.4).
314 After incubating for 15 mins, the reaction was stopped by adding 300 μl of
315 hexane and vortexing. The organic phase was then filtered through a 0.2
316 μm filter (Millipore) before being used for GC-MS analysis. Peak
317 identification of each component was confirmed using authentic samples
318 analysis.

319 **Gene expression in yeast and enzyme assay.** Purified PCR products were cloned
320 into the *pEASY*-Blunt Cloning Vector (Transgen, catalog number: CB101)
321 and sequenced for errors. The full cDNAs of CYP450s from Nipponbare
322 (*O. sativa* L. spp. *japonica*) were cloned into the pESC-URA vector
323 (Stratagene, Accession #AF063585) expression vector. The primers listed
324 in Supplementary Table 9. The constructed vectors were transformed into
325 the yeast strain WAT11 using the lithium acetate method following the
326 manufacturer described protocol (ZYMO RESEARCH, catalog number:
327 T2001). Yeast cultures were grown and microsomes were prepared as
328 previously described with some modification[44]. Briefly, the recombinant
329 cells were first cultured in SC minimal medium containing 2% glucose at
330 30°C. For protein induction, cells were collected and resuspended in
331 Synthetic Complete Medium yeast minimal medium) containing 2%
332 galactose instead of glucose (http://fungenome.bioon.com.cn), and
333 cultured 30°C for 2 days. Cells were harvested by centrifugation and

broken with glass beads (0.45 mm in diameter, *SIGMA*) in 50 mM Tri-HCl buffer, PH = 7.5, containing 1 mM EDTA and 600 mM sorbitol. The cells were broken using a mix mill (Model MM 400, Retsch, Haan, Germany). The homogenate was centrifuged for 60 min at 12,000g and the resulting supernatant was centrifuged for 90 min at 120,000g. The pellet consisting of microsomal membranes was resuspended in 100 mM Tris-HCl, PH = 7.5, 1 mM EDTA, and 20% (v/v) glycerol and stored at -80°C for long term storage.

In vitro enzymatic activity assays were performed on a shaking incubator (120 rpm), at 30°C for 4 h in 500 μl of 100 mM Tris-HCl, PH = 7.5, containing 1 mg total microsomal proteins, 500 mM NADPH, 200 μM substrate. Reactions were stopped by addition of 500 μl of hexane and vortexing. Negative control reactions by were carried out with microsomal preparations from recombinant yeast transformed with 'empty' pESC-URA. Total protein content was estimated by measuring UV absorbance at 280 nm on NanoDrop ND-1000 spectrophometer.

Enzyme pathway reconstitution in *N. benthamiana.* Transient expression construct of candidate genes was generated by directionally inserting the full cDNAs first into the pDONR207 (Gen^R) entry vector and then into the destination vector pEAQ-HT using the Gateway recombination reaction (Invitrogen)[45], followed by transformed into *Agrobacterium tumefaciens* (EHA105). Positive clones were selected and grown to optical density (OD)

600 of 2.0 in 10ml of Luria-Bertani (LB) medium containing 50μg/mL Kanamycin, washed with washing buffer (10 mM 2-(N-morpholino) ethanesulfonic acid [MES], pH = 5.6), and resuspended in MMA buffer (10 mM MES [pH = 5.6], 10 mM $MgCl_2$, 100 mM acetosyringone) to OD600 of 1.0. The culture was incubated for 2 hours in room temperature and one milliliter of culture was used to infiltrate the underside of 6-week-old *N. benthamiana* leaves with a needleless 1 mL syringe[46]. Leaves were harvested 3 days post infiltration, flash frozen and stored at -80°C for later processing.

Statistics and reproducibility. The statistical analyses were performed using GraphPad Prism 8 and OriginPro 8. Each experiment was repeated at least twice, and similar results were obtained.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The sequences data of 424 *O. sativa* accessions is available in NCBI under the BioProject PRJNA171289[17]. The single nucleotide polymorphisms (SNPs) information of 424 *O. sativa* accessions is available in RiceVarMap (http://ricevarmap.ncpgr.cn/v1). The pan-genome data were acquired from the pan-genome dataset (https://figshare.com/collections/Novel_sequences_structural_variations_

and_gene_presence_variations_of_Asian_cultivated_rice/3876022/1 and http://cgm.sjtu.edu.cn/3kricedb/)[18,19,24]. 13 of *O. rufipogon* were selected from 446 diverse *O. rufipogon* accessions from Asia and Oceania, and represented all the major genetically distinct clusters in *O. rufipogon* and the other 10 wild rice are from *EnsemblPlants* (http://plants.ensembl.org/index.html) and National Genomics Data Center (https://bigd.big.ac.cn/search?dbId=gwh&q=*Oryza*), including *Oryza barthii* (AA), *Oryza glumipatula* (AA), *Oryza glaberrima* (AA), *Oryza meridionalis* (AA), *Oryza longistaminata* (AA), *Oryza nivara* (AA), *Oryza brachyantha* (FF), *Oryza punctata* (BB) and *Oryza brachyantha* (GG)[24]. Genes reported in the study are deposited in the National Center for Biotechnology Information (NCBI). The genes can be found in GenBank or Rice Genome Annotation Project database (http://rice.plantbiology.msu.edu/analyses_search_locus.shtml) under the following accession numbers: OsTPS28, MN833254; OsCYP71Z21, LOC_Os07g11870; OsCYP71Z2, LOC_Os07g11739; OsCYP71Z22, LOC_Os07g11970; OsCYP71Z30, LOC_Os07g11890.

Acknowledgements

We thank Prof. Jay D. Keasling, Prof. George P. Lomonossoff and Prof. Zongbao Zhao for their advice and their gift of the expression vectors and strains. We also thank Dr. David R Nelson, University of Tennessee, for the help in naming the OsCYP71Z30. This work was supported by the

National Science Fund for Distinguished Young Scholars (No. 31625021), the State Key Program of National Natural Science Foundation of China (No. 31530052), and the Hainan University Startup Fund KYQD(ZR)1866 to J.L.

Author contributions

J.L. designed the research. J.L., L.-L.C., L.Q., M.Y. and X.L. supervised this study. C.Z., Long L., S.Z., Z.L., F.Z., M.Z., Y.S., Yuheng S., K.L., T.C., M.H., I.G., Z.Y. and T.T. participated in the material preparation. C.Z., C.Y., Y.L., X.W. and J.S. carried out the metabolite analyses. C.Z., Z.L., S.Z., C.Y., X.Z., H.G., M.P., M.Z., Yufei L., Z.Y., Ling L., S.S., J.S., X.J., Y.L., T.T. and Z.W. performed the data analyses. C.Z., Long L., Z.L., S.Z. and C.Y. performed most of the experiments; J.L., C.Z., I.G. and A.R.F. wrote the manuscript.

Competing interests

The authors declare no conflict of interests.

References

1. Schmelz, E.A. *et al.* Biosynthesis, elicitation and roles of monocot terpenoid phytoalexins. *Plant J.* **79**, 659-78 (2014).
2. Lu, X. *et al.* Inferring roles in defense from metabolic allocation of rice diterpenoids. *Plant Cell* **30**, 1119-1131 (2018).
3. Li, T. *et al.* Jumonji C domain protein JMJ705-mediated removal of histone H3 lysine 27

trimethylation is involved in defense-related gene activation in rice. *Plant Cell* **25**, 4725-4736 (2013).
4. Chen, X. *et al.* Biological activities and potential molecular targets of cucurbitacins: a focus on cancer. *Anti-Cancer Drug* **23**, 777-787 (2012).
5. Huang, A.C. *et al.* A specialized metabolic network selectively modulates Arabidopsis root microbiota. *Science* **364**, eaau6389 (2019).
6. Nützmann, H.-W., Huang, A. & Osbourn, A. Plant metabolic clusters – from genetics to genomics. *New Phytol.* **211**, 771-789 (2016).
7. Zi, J., Mafu, S. & Peters, R.J. To gibberellins and beyond! surveying the evolution of (di)terpenoid metabolism. *Annu. Rev. Plant. Biol.* **65**, 259-286 (2014).
8. Panizza, B.J. *et al.* Phase I dose-escalation study to determine the safety, tolerability, preliminary efficacy and pharmacokinetics of an intratumoral injection of tigilanol tiglate (EBC-46). *Ebiomedicine* **50**, 433-441 (2019).
9. Hezareh, M. Prostratin as a new therapeutic agent targeting HIV viral reservoirs. *Drug News Perspect.* **18**, 496-500 (2005).
10. Johnson, H.E., Banack, S.A. & Cox, P.A. Variability in content of the anti-AIDS drug candidate prostratin in samoan populations of homalanthus nutans. *J. Nat. Prod.* **71**, 2041-2044 (2008).
11. Lebwohl, M. *et al.* Ingenol mebutate gel for actinic keratosis. *N. Engl. J. Med.* **366**, 1010-1019 (2012).
12. Sabandar, C.W., Ahmat, N., Jaafar, F.M. & Sahidin, I. Medicinal property, phytochemistry and pharmacology of several Jatropha species (Euphorbiaceae): A review. *Phytochemistry* **85**, 7-29 (2013).
13. Inoue, Y. *et al.* Identification of a novel casbane-type diterpene phytoalexin, ent-10-oxodepressin, from rice leaves. *Biosci. Biotech. Bioch.* **77**, 760-765 (2013).
14. Horie, K., Sakai, K., Okugi, M., Toshima, H. & Hasegawa, M. Ultraviolet-induced amides and casbene diterpenoids from rice leaves. *Phytochem. Lett.* **15**, 57-62 (2016).
15. Li, W. *et al.* OsCYP71Z2 involves diterpenoid phytoalexin biosynthesis that contributes to bacterial blight resistance in rice. *Plant Sci.* **207**, 98-107 (2013).
16. Li, W. *et al.* Overexpressing CYP71Z2 enhances resistance to bacterial blight by suppressing auxin biosynthesis in rice. *Plos One* **10**, e0119867 (2015).
17. Zhao, H. *et al.* RiceVarMap: a comprehensive database of rice genomic variations. *Nucleic Acids Res.* **43**, D1018-D1022 (2015).
18. Wang, W. *et al.* Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* **557**, 497-501 (2018).
19. Huang, X. *et al.* A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**, 497-501 (2012).
20. Chen, W. *et al.* Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat. Genet.* **46**, 714-721 (2014).
21. Yu, N. *et al.* Delineation of metabolic gene clusters in plant genomes by chromatin signatures. *Nucleic Acids Res.* **44**, 2255-2265 (2016).
22. Zhou, S. *et al.* Cooperation between the H3K27me3 chromatin mark and non-CG methylation in epigenetic regulation. *Plant Physiol.* **172**, 1131-1141 (2016).
23. Nützmann, H.-W. *et al.* Active and repressed biosynthetic gene clusters have spatially distinct chromosome states. *Proc. Natl. Acad.Sci.U.S.A.* **24**, 13800-13809 (2020).

24. Stein, J.C. *et al.* Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus Oryza. *Nat. Genet.* **50**, 285-296 (2018).
25. Wing, R.A., Purugganan, M.D. & Zhang, Q. The rice genome revolution: from an ancient grain to green super rice. *Nat. Rev. Genet.* **19**, 505-517 (2018).
26. King, A.J., Brown, G.D., Gilday, A.D., Larson, T.R. & Graham, I.A. Production of bioactive diterpenoids in the Euphorbiaceae depends on evolutionarily conserved gene clusters. *Plant Cell* **26**, 3286-3298 (2014).
27. Takos, A.M. *et al.* Genomic clustering of cyanogenic glucoside biosynthetic genes aids their identification in Lotus japonicus and suggests the repeated evolution of this chemical defence pathway. *Plant J.* **68**, 273-286 (2011).
28. Field, B. & Osbourn, A.E. Metabolic diversification—independent assembly of operon-like gene clusters in different plants. *Science* **320**, 543-547 (2008).
29. M, W. *et al.* Parallel selection on a dormancy gene during domestication of crops from multiple families. *Nat. Genet.* **50**, 1435-1441 (2018).
30. Swaminathan, S., Morrone, D., Wang, Q., Fulton, D.B. & Peters, R.J. CYP76M7 Is an ent-cassadiene C11-hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice. *Plant Cell* **21**, 3315-3325 (2009).
31. Wang, Q., Hillwig, M.L. & Peters, R.J. CYP99A3: functional identification of a diterpene oxidase from the momilactone biosynthetic gene cluster in rice. *Plant J.* **65**, 87-95 (2011).
32. Wang, Q. *et al.* Characterization of CYP76M5-8 indicates metabolic plasticity within a plant biosynthetic gene cluster. *J. Biol. Chem.* **287**, 6159-68 (2012).
33. Miyamoto, K. *et al.* Evolutionary trajectory of phytoalexin biosynthetic gene clusters in rice. *Plant J.* **87**, 293-304 (2016).
34. Gross, B.L. & Zhao, Z. Archaeological and genetic insights into the origins of domesticated rice. *Proc. Natl. Acad.Sci.U.S.A.* **111**, 6190-6197 (2014).
35. Qi, Z. Genetics and improvement of bacterial blight resistance of hybrid rice in China. *Rice Sci.* **23**, 111-119 (2009).
36. Kauffman, H.E., Reddy, A.P.K., Hsieh, S.P.Y. & Merca, S.D. An improved technique for evaluating resistance of rice varieties to Xanthomonas oryzae. *Plant dis. rep.* **57**, 537-541 (1973).
37. Zhang, H. *et al.* Transposon-derived small RNA is responsible for modified function of WRKY45 locus. *Nat. Plants* **2**, 16016 (2016).
38. Zhang, B. *et al.* Multiple alleles encoding atypical NLRs with unique central tandem repeats in rice confer resistance to Xanthomonas oryzae pv. oryzae. *Plant Commun.* **1**, 100088 (2020).
39. Tao, Z. *et al.* A pair of allelic WRKY genes play opposite roles in rice-bacteria interactions. *Plant physiol.* **151**, 936-948 (2009).
40. King, A.J. *et al.* A cytochrome P450-mediated intramolecular carbon–carbon ring closure in the biosynthesis of multidrug-resistance-reversing lathyrane diterpenoids. *Chembiochem* **17**, 1593-1597 (2016).
41. Boutanaev, A.M. *et al.* Investigation of terpene diversification across multiple sequenced plant genomes. *Proc. Natl. Acad.Sci.U.S.A.* **112**, E81-E88 (2015).
42. Luo, D. *et al.* Oxidation and cyclization of casbene in the biosynthesis of Euphorbia factors from mature seeds of Euphorbia lathyris L. *Proc. Natl. Acad.Sci.U.S.A.* **113**, E5082-E5089 (2016).
43. Peng, M. *et al.* Differentially evolved glucosyltransferases determine natural variation of rice flavone accumulation and UV-tolerance. *Nat. Commun.* **8**, 1975 (2017).

44. Ikezawa, N. *et al.* Lettuce costunolide synthase (CYP71BL2) and its homolog (CYP71BL1) from sunflower catalyze distinct regio- and stereoselective hydroxylations in sesquiterpene lactone metabolism. *J. Biol. Chem.* **286**, 21601-21611 (2011).
45. Sainsbury, F., Thuenemann, E.C. & Lomonossoff, G.P. pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol. J.* **7**, 682-693 (2009).
46. Zeng, X. *et al.* Genome-wide dissection of co-selected UV-B responsive pathways in the UV-B adaptation of qingke. *Mol. Plant* **13**, 112-127 (2020).

Fig. 1 | Identification the structural variation of a diterpene gene cluster in rice. a, Manhattan plot of 5,10-diketo-casbene trait across the 12 rice chromosomes. In *japonica* population, all metabolite-SNP associations with *P* values below 1.8819e-7 (horizontal dashed lines in all Manhattan plots) are plotted against genome location in intervals of 1Mb. The two-tailed Student's *t* test is performed for this analysis and the Bonferroni correction is used for the multiple-comparison correction. The Manhattan plots from two individual replicate for each locus are provided in Supplementary Fig.2. Chr., chromosome; TPS, terpene synthase; CYP450, cytochromes P450 monooxygenases; TE, transposable element. **b**, A 150-kb insertion in Nipponbare (Nip) contains *DGC7* not present at the syntenic location in Minghui63 (MH63), Zhenshan97 (ZS97). **c**, The model of two types of *DGC7*. **d**, The relative content of 5,10-diketo-casbene in different subspecies of rice. **e**, Relative content of 5,10-diketo-casbene subjected to 0.1mM methyl jasmonate treatment for 24 hours in Nip, MH63, ZS97. CK, control check; nd, not detected. The data are presented as mean ± s.d., n=3 biologically independent replicates. The asterisks in Fig.1e indicate significant differences compared with the CK: ****P<0.0001 by unpaired two-tailed Student's *t* tests. **f**, The relative content of 5,10-diketo-casbene in the randomly selected varieties. Presence or

absence of the *DGC7* genome fragment indicated by +/-. The data are presented as mean ± s.d., *n*=3 biologically independent replicates. Source data are provided as a Source Data file.

Fig. 2 | Identification of a diterpene gene cluster. a, Subcellular localization pattern of the confirmed OsTPS28. Transient expression of confirmed OsTPS28 fused to GFP in rice leaf protoplasts showing chloroplast localization. Bar=10μm. All experiments were repeated three times with similar results. **b**, Metabolic profiling of *N. benthamiana* leaves using ultra-performance liquid chromatography coupled with QQQ mass spectrometry (LC-ESI-QQQ-MS/MS) with and without the infiltration of the corresponding candidates. 10-keto-casbene and 5,10-diketo-casbene reference compounds were purified from rice leaves by the method described previously[13,14]. GFP, green fluorescent protein. **c**, Hierarchical clustering of RNA-Seq expression data. Color key: known diterpene biosynthesis genes (gray), genes identified in this report (*OsCYP71Z2*, *OsTPS28*, *OsCYP71Z21*, *OsCYP71Z22* and *OsCYP71Z30*) are red. The aerial part of 12-day-old seedling were used for the treatment. Hours (h) post 0.1mM methyl jasmonate treatment are indicated.

Fig. 3 | The regulation of *DGC7*. a, Gene expression levels of *OsCYP71Z2*, *OsTPS28*, *OsCYP71Z21* in MeJA treated and control plants. The data are presented as mean ± sd, *n*=3 biologically independent replicates. **b**, H3K27me3 ChIP-on-chip data for the genes from *DGC7*. The data is extracted from[3]. **c**, H3K27me3 ChIP analysis for the genes from *DGC7* in seedlings. Transcript levels were analyzed by qPCR. The data are

presented as mean ± s.d., *n*=3 biologically independent replicates. **d**, The relative content of 5,10-diketo-casbene in the JMJ705 overexpression line. The data are presented as mean ± s.d., *n*=3 biologically independent replicates. **e**, The relative content of 5,10-diketo-casbene in the JMJ705 RNAi line. The data are presented as mean ± s.d., *n*=3 biologically independent replicates. The asterisks in Fig. 3a, 3c-e indicate significant differences compared with the CK or ZH11: **P*<0.05, ***P*<0.01, ****P*<0.001, *****P*<0.0001 by unpaired two-tailed Student's *t* tests. Source data are provided as a Source Data file.

Fig. 4 | The origin of *DGC7*. a, Phylogenetic analysis show an TPS-II clade across in the *Oryza* species. **b**, The OsCYP71Z2/OsCYP71Z1/OsCYP71Z21-OsCYP71Z22 tree shows the latest duplications are likely to have occurred prior to divergence of *O. punctata* (BB genome type) and *O. sativa* (AA genome type). *Leersia perrieri* is the evolutionally closest outgroup species for *Oryza*. **c**, The selection of *DGC7*. The relative proportion of six types of gene modules. The intact *DGC7* is absent in *O. rufipogon*, highly enriched in *japonica* varieties but not in *indica* varieties. The data extracted from[18,19].

Extended Data Fig. 1 | The distribution of the world-wide collection of rice accessions in this study. The core collection of 424 cultivated rice accessions in this study has a wide geographic distribution. Color dots indicate different subspecies/type of cultivated rice. The map is draw by R 3.1 and the information of Latitude and Longitude of the rice varieties have also been shown in the Supplementary Table 1.

Extended Data Fig. 2 | Functional analyses of OsTPS28, OsCYP71Z2 and OsCYP71Z21. a, Gas chromatography of the reaction products of OsTPS28 with GGDP. GGDP, geranylgeranyl diphosphate. Casbene and neocembrene reference compounds were purified from infiltrated *N. benthamiana* leaves by the method described previously[26]. Compound 1, casbene; Compound 2, neocembrene. **b**, Gas chromatography of *in vitro* enzyme assays showing the 10-keto-casbene C5 oxidase activity of yeast-expressed CYP71Z2 in the present of NADPH. Microsomes prepared from yeast containing PESC-URA empty vector were used as a negative control. 10-keto-casbene reference compound was purified from rice leaves by the method described previously[13,14]. Compound 3, 10-keto-casbene; Compound 4, 5,10-diketo-casbene. **c**, Gas chromatography of the extracts prepared from the leaves of *N. benthamiana* infiltrated with OsTPS28 over-expressing vector.

Extended Data Fig. 3 | Mass spectrum and structure of 5,10-diketo-casbene. a, Mass spectrum and structure of the product in *N. benthamiana* leaves simultaneously overexpressing *OsTPS28*, *OsCYP71Z2* and *OsCYP71Z21*. **b**, Mass spectrum of 5,10-diketo-casbene reference. LC-MS, liquid chromatography-mass spectrometry. **c**, ^{1}H NMR (left) and ^{13}C NMR (right) results of 5,10-diketo-casbene.

Extended Data Fig. 4 | The expression profiles of genes from *DGC7*. The genes from *DGC7* are indicated in bold. The transcript abundances of indicated genes in different organs at different stages were shown: expression levels of *OsTPS28*, *OsCYP17Z2* and *OsCYP71Z21* is correlated at different developmental stages. The numerical values for

608 blue-to-red gradient represent normalized expression levels from quantitative real-time
609 PCR (qRT-PCR) analysis.
610
611 **Extended Data Fig. 5 | The casbene-type diterpene biosynthesis via distinct**
612 **biosynthetic routes in rice and castor.** The casbene-type diterpene biosynthetic
613 pathways in rice and castor. Chr.7, chromosome 7; GGDP, geranylgeranyl
614 diphosphate.