UNIVERSITY of York

This is a repository copy of Selection of a subspecies-specific diterpene gene cluster implicated in rice disease resistance.

White Rose Research Online URL for this paper: <u>https://eprints.whiterose.ac.uk/id/eprint/168087/</u>

Version: Accepted Version

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

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



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

4	Chuansong Zhan ¹ , Long Lei ^{2,1} , Zixin Liu ¹ , Shen Zhou ¹ , Chenkun Yang ¹ , Xitong Zhu ¹ , Hao Guo ^{2,1} , Feng
5	Zhang ¹ , Meng Peng ¹ , Meng Zhang ¹ , Yufei Li ¹ , Zixin Yang ¹ , Yangyang Sun ¹ , Yuheng Shi ¹ , Kang Li ¹ , Ling
6	Liu ² , Shuangqian Shen ¹ , Xuyang Wang ¹ , Jiawen Shao ¹ , Xinyu Jing ¹ , Zixuan Wang ¹ , Yi Li ³ , Tomasz
7	Czechowski ³ , Morifumi Hasegewa ⁴ , Ian Graham ³ , Takayuki Tohge ⁵ , Lianghuan Qu ¹ , Xianqing Liu ^{2,1} ,
8	Alisdair R. Fernie ⁶ , Ling-Ling Chen ¹ , Meng Yuan ¹ , Jie Luo ^{2,1}
9	
10	¹ National Key Laboratory of Crop Genetic Improvement and National Center of Plant Gene Research (Wuhan),
11	Huazhong Agricultural University, Wuhan 430070, China.
12	² College of Tropical Crops, Hainan University, Haikou 572208, China.
13	³ Centre for Novel Agricultural Products, Department of Biology, University of York, York, United Kingdom.
14	⁴ College of Agriculture, Ibaraki University, 3-21-1 Chuo, Ami, Ibaraki 300-0393, Japan.
15	⁵ Graduate School of Biological Sciences, Nara Institute of Science and Technology, Ikoma, Nara, 630-0192, Japan.
16	⁶ Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany.
17	⊠e-mail: jie.luo@hainanu.edu.cn
18	
19	Diterpenoids are the major group of antimicrobial phytoalexins in rice ^{1,2} . Here we
20	report the discovery of a rice diterpenoid gene cluster on chromosome 7 (DGC7)
21	encoding the entire biosynthetic pathway to 5,10-diketo-casbene, a member of the
22	mono-cyclic casbene-derived diterpenoids. We revealed that DGC7 is regulated
23	through MeJA mediated epigenetic control directly by JMJ705 ³ . Functional
24	characterization of pathway genes revealed OsCYP71Z21 to encode a casbene C10
25	oxidase, sought after for the biosynthesis of an array of medicinally important

diterpenoids. We further reveal that DGC7 arose relatively recently in the Oryza 26 genus, that it was partly formed in O. rufipogon and positively selected for in 27 *japonica* during domestication. Casbene synthesizing enzymes that are 28 functionally equivalent to OsTPS28 are present in several species of 29 Euphorbiaceae but gene tree analysis shows that these and other casbene-30 modifying enzymes have evolved independently. As such, combining casbene-31 modifying enzymes from these different families of plants may prove effective in 32 producing a diverse array of bioactive diterpenoid natural products. 33

34

The terpenoid class of specialized metabolites are important in the 35 adaptation of plants to their ecological niches, as well as serving as a 36 valuable medicinal resource^{2,4,5}. Enzymes involved in the metabolic 37 pathways of distinct terpenoid classes are encoded by gene clusters in a 38 range of plant species⁶. For example, the phytocassane metabolic gene 39 cluster in rice confers resistance to the fungal pathogen Magnaporthe 40 orvzae (M. orvzae) and the bacterial pathogen Xanthomonas orvzae pv 41 oryzae (Xoo), the cucurbitacins of Cucumis sativus can be used as 42 traditional medicines and the thalianol gene cluster in Arabidopsis thaliana 43 may modulate Arabidopsis root microbiota^{2,4,5}. The C-20 diterpene class of 44 terpenoids can be further subdivided into a large superfamily of labdane-45 related diterpenoids which include the gibberellins and are defined by an 46 initial dual cyclisation of geranylgeranyl diphosphate (GGDP) and others 47

including casbene-type diterpenoids which are formed by monocyclisation of GGDP⁷. Casbene-type diterpenoids are found predominantly
in the Euphorbiaceae family and are recognized as being rich in a range of
pharmacological activities⁸⁻¹¹, consistent with their widespread use in
traditional medicine around the world¹². 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) 54 was the first casbene-type diterpene phytoalexin found in rice (Oryza 55 sativa) that confers rice bacterial blight and rice blast fungus resistance¹³⁻ 56 ¹⁶. However, surprisingly to date no study has yet assessed the natural 57 variation in the ability to produce 5,10-diketo-casbene. More than 4,000 58 diverse accessions of O. sativa (indica and japonica) and various wild rice 59 relatives have been sequenced in recent years allowing the generation of a 60 detailed genome-variation map¹⁷⁻¹⁹ and the opportunity to perform Genome 61 Wide Association Studies to locate the genetic basis of traits exhibiting 62 natural variation. 63

Here, we report that the locus responsible for the biosynthesis of 5,10diketo-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, 73 we collected leaf samples of 424 rice (O. sativa) accessions from a diverse 74 worldwide resource panel (Extended Data Fig. 1 and Supplementary Table 75 1)²⁰. A metabolite-based genome-wide association study (mGWAS) was 76 performed for both the full population (all 424 accessions) and each of the 77 two subpopulations, *indica* (271 accessions) and *japonica* (132 accessions), 78 independently (Fig. 1a, Supplementary Fig. 1 and Supplementary Table 2). 79 The association results showed that natural variation in 5,10-diketo-80 casbene of *japonica* rice was mainly controlled by a locus on chromosome 81 7 (Fig. 1a and Supplementary Fig. 1a, b). OsTPS28 (Os07g11790), the only 82 terpene gene within this locus, was chosen as a candidate for the diterpene 83 synthase and four putative CYP450 genes (OsCYP71Z2, Os07g11739; 84 OsCYP71Z21, Os07g11870; OsCYP71Z30, Os07g11890; OsCYP71Z22, 85 Os07g11970) were candidates for the oxidation of casbene to produce 5,10-86 diketo-casbene (Fig. 1a and Supplementary Fig. 1 a-d). Together these gene 87 candidates represented a putative diterpene gene cluster across a 140kb 88 region hereafter referred to as Diterpene Gene Cluster on chromosome 7, 89 DGC7. 90

91

The absence of the association signal on chromosome 7 in indica

panel compared to the *japonica* panel (Fig. 1a and Supplementary Fig. 1), 92 led us to perform an in-depth analysis of the corresponding region using 93 reference genomes three high-quality (Nipponbare, Minghui63, 94 Zhenshan97). This revealed that OsTPS28 and OsCYP71Z21 present in 95 Nipponbare (*japonica* rice), but absent from Minghui63 and Zhenshan97 96 (indica rice) (Fig. 1b and Supplementary fig. 2). Analysis of the pan-97 genome data^{18,19} identified two major types of DGC7 (DGC7^{present}, 98 DGC7^{absent or incomplete}) (Fig. 1c and Supplementary Tables 3, 4). Consistent 99 with the presence/absence of intact DGC7, we observed the accumulation 100 of 5,10-diketo-casbene in most japonica varieties (131/132) while this 101 metabolite is absent in most *indica* varieties (266/271) both under control 102 conditions and following MeJA-treatment, indicating that the 103 presence/absence of DGC7 determines the natural variation of 5,10-diketo-104 casbene in rice (Fig. 1d, e and Supplementary Table 2). Furthermore, the 105 varieties lacking intact DGC7 do not produce 5,10-diketo-casbene while 106 varieties with intact DGC7 accumulate 5,10-diketo-casbene (Fig. 1f and 107 Supplementary Fig. 3). 108

To characterize the putative gene cluster, we first cloned the openreading-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 114 expressed in Escherichia coli BL21 and recombinant protein produced 115 casbene in the presence of GGDP and Mg²⁺ (Extended Data Fig. 2a and 116 Supplementary Fig. 5, a-d), with a $K_{\rm m}$ of 5.16µM and $K_{\rm cat}$ of 0.0236 s⁻¹ 117 (Supplementary Fig. 6). Stable transformation analysis in rice revealed that 118 increased by ~1.9 5,10-diketo-casbene was fold in OsTPS28 119 overexpression lines while reduced to non-detectable levels in the 120 OsTPS28 knockout plants (Supplementary Fig. 7). 121

Recombinant OsCYP71Z2 protein produced in Saccharomyces 122 cerevisiae oxidized the C5 position of 10-keto-casbene to produce 5,10-123 diketo-casbene (Extended Data Fig. 2b and Supplementary Fig. 5, e-f). To 124 125 further dissect the biosynthetic pathway to 5,10-diketo-casbene we used Agrobacterium-mediated transient expression in N. benthamiana using 126 OsTPS28 in combination with different CYP450s present in DGC7. Over-127 expression of OsTPS28 alone led to the formation of casbene (major 128 product) and neocembrene (minor products) (Extended Data Fig. 2c and 129 Supplementary Fig. 5, a-d); combined expression of OsTPS28 with 130 OsCYP71Z21 resulted in production of 10-keto-casbene (Fig. 2b and 131 Supplementary Fig. 5g, h) while combined expression of OsTPS28, 132 OsCYP71Z21 and OsCYP71Z2 produced 5,10-diketo-casbene (Fig. 2b, 133 Extended Data Fig. 3, a-c and Supplementary table 5). To further verify the 134 function of OsCYP71Z21, plant microsomes were isolated from the N. 135

benthamiana leaves that infiltrated OsCYP71Z21. In vitro enzyme assay 136 using isolated microsomes showed that OsCYP71Z21 was able to 137 converted casbene to 10-keto-casbene in the presence of NADPH 138 (Supplementary Fig. 8a), which further supported the notion that 139 OsCYP71Z21 encoded a casbene C10 oxidase. These results are also 140 consistent with those obtained for the recombinant OsTPS28 produced in 141 E. coli and OsCYP71Z2 produced in S. cerevisiae and lead us to conclude 142 that OsTPS28 is a casbene synthase, OsCYP71Z2 is a casbene C5 oxidase 143 and OsCYP71Z21 is a casbene C10 oxidase. Together these three enzymes 144 produce 5,10-diketo-casbene from GGDP. However, the expression of 145 OsCYP71Z21 in yeast (WAT11) did not lead to production of 10-keto-146 casbene or any other metabolite (Supplementary Fig. 8b). It is thus possible 147 that OsCYP71Z21 is not expressed in an active form in yeast. Sequence 148 similarity analysis was performed for OsCYP71Z2, OsCYP71Z21, 149 OsCYP71Z22 and OsCYP71Z30. CYP71Z2 revealing 73.38%, 76.15% 150 and 70.42% identities to OsCYP71Z21, OsCYP71Z22 and OsCYP71Z30 151 (Supplementary Fig. 9). Although the sequence similarities are all above 152 70%, the OsCYP71Z22 and OsCYP71Z30 still failed to exhibit activity 153 with casbene as substrate (Supplementary Fig. 8c). 154

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

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

from *DGC7* member genes in addition to defense-related genes³. 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 185 family, a phylogenomic approach has been used to investigate the 186 evolutionary origins of all three DGC7 members: OsTPS28, OsCYP71Z2 187 and OsCYP71Z21. BLAST searches have identified all homologous 188 sequences from the GenBank non-redundant protein database as well as 189 additional annotation datasets from 29 draft whole genome assemblies of 190 representative grass species including Oryza species of AA, BB, FF and 191 GG genome types (Supplementary Table 7). Subsequent progressive gene 192 tree analyses suggest all three DGC7 genes are members of subfamilies 193 specific to Oryza genus within their respective grass-specific gene families 194 (Supplementary Figs. 12 and 13). Furthermore, the latest gene 195 duplications giving rise to the closest paralogue pairs of OsTPS28/OsTPS2 196 and OsCYP71Z2/OsCYP71Z1/OsCYP71Z21-OsCYP71Z22 are likely to 197 have occurred prior to divergence of the BB (O. punctata) and AA genome 198 types (O. sativa) about 7 Mya (Fig. 4a, b)²⁴. In the latter case, the gene 199 duplication events that led to OsCYP71Z2, OsCYP71Z1, OsCYP71Z21, 200 and *OsCYP71Z22* appear to have happened after the AA/BB genome types 201

diverged from the GG genome type (O. granulate) approximately 15 Mya 202 (Fig. 4b)²⁵. In addition, the latest gene duplication leading to the youngest 203 paralogue pair OsCYP71Z21/OsCYP71Z22 might be within the AA 204 genome types before the African wild rice (O. longistaminata) diverged 205 from O. sativa approximately 2 Mya (Fig. 4b)²⁴. These results have led us 206 to conclude that all three functionally characterized members of DGC7, 207 OsTPS28, OsCYP71Z2 and OsCYP71Z22 have a recent origin (~2-7 Mya) 208 in the Oryza genus and the gene clustering would have happened following 209 these gene duplication events. Therefore, we have seen the convergent 210 evolution of casbene synthases to produce the casbene backbone and 211 subsequent independent evolution of the P450 oxidases which gave rise to 212 213 a range of different casbene-derived diterpenes in the Poaceae and Euphorbiaceae families (Supplementary Figs. 14 and Extended Data Fig. 214 5^{26} . Similar cases have been reported for cyanogenic glycosides and 215 triterpenes^{27,28}. 216

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 229 mainly restricted to cultivated rice, especially among varieties of *japonica* 230 subspecies (Fig. 4a, b), even though individual members may be present in 231 other wild rice species. Apart from the aforementioned sole orthologue of 232 OsTPS28 among 15 representative whole genome assemblies among rice 233 species, orthologues of OsCYP71Z2 can only be identified in the genomes 234 235 of O. punctata and O. rufipogon whereas none can be found corresponding to OsCYP71Z21. However, intact DGC7 with all three members present 236 can only be identified in O. sativa (Fig. 4a, b). This is further demonstrated 237 in a haplotypes survey of all three members of DGC7 in the combined total 238 of 435 varieties from the *japonica* and *indica* subspecies of O. sativa as 239 well as O. rufipogon. No intact DGC7 has been found in the 13 O. 240 rufipogon varieties, even though individual components are present 241 corresponding to all three components (Fig. 4c and Supplementary Tables 242 3, 4). Furthermore, intact DGC7 is highly enriched in japonica varieties 243 (102/109) compared to the *indica* varieties (13/313), suggesting the 244 selection of DGC7 during domestication (Fig. 4c and Supplementary Table 245

3). Results from π and F_{ST} revealed that DGC7 was located in a selective 246 sweep region (selective sweep defined as top 5% of the length of the whole 247 genome sequence) in *japonica* but not in *indica* (Supplementary Fig. 15)²⁹. 248 Unlike the momilactone and phytocassane gene clusters that 249 biosynthesize common labdane-related diterpenoids and are found in both 250 *indica* and *japonica* varieties³⁰⁻³³, *DGC7* biosynthesizes casbene-type 251 diterpenoids that are almost exclusively restricted to *japonica* varieties. In 252 summary, we have shown all members of DGC7 originated in the Oryza 253 genus; and DGC7 is at least partly formed in the wild rice ancestor O. 254 rufipogon and has been positively selected for in *japonica* rather than 255 indica during domestication. Japonica and indica rice originated in 256 Southern China and India respectively³⁴. Given that Southern China has 257 been an endemic area of rice bacterial blight and 5,10-diketo-casbene 258 confers rice blast resistance^{14,15,35} it can be speculated that this provided 259 the selection pressure for DCG7 to predominate in a subspecies-specific 260 manner. 261

Considerable evidence suggests that the end-product of DGC7 - 5,10diketo-casbene is a rice phytoalexin which has antifungal activity against $M. \ oryzae^{1,14}$. It can be induced by the rice blast fungus and furthermore inhibits rice blast fungus spore germination and germ tube growth¹³. Moreover, overexpression of OsCYP71Z2 (one gene member of DGC7) in 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 268 immunity. To further validate this suggestion, OsTPS28-OE, OsTPS28-KO, 269 and wild-type plants (Zhonghua 11) were separately inoculated with a 270 highly virulent Chinese Xoo strain FuJ23 via a leaf chipping method³⁶⁻³⁸. 271 The results showed that the disease areas caused by *Xoo* in the OsTPS28-272 OE lines were much smaller than those in the wild type plants and the 273 OsTPS28-KO lines exhibited larger disease areas relative to their wild type 274 counterparts (Supplementary Fig. 16 and Supplementary Table 8). 275 However, this does not mean all the *japonica* varieties which contain the 276 DGC7 cluster are more resistant to disease than *indica* varieties. Indeed, 277 some *indica* varieties also have high levels of disease resistance and 278 specific genes that confer rice disease resistance in *indica* varieties have 279 been identified. For example, WRKY45 plays a positive role in regulating 280 disease resistance in the *indica* varieties while playing a negative role in 281 *japonica* varieties^{37,39}. 282

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 casbenederived diterpenoids such as tiglanes, ingenanes and daphnanes⁸⁻¹¹. 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 metabolicengineering and production in heterologous hosts.

292 Methods

Plant materials. All plants used in this study were grown in Huazhong 293 Agricultural University, Wuhan, Hubei Province of China. The germplasm 294 set of 424 O. sativa accessions consisted of both elite and landraces 295 varieties (Supplementary Table 1). All samples were collected and flash-296 freezing in liquid N₂. Later, all samples were stored at -80°C until vacuum 297 freeze-drying. Samples were collected with two biological replicate sets at 298 different places and the data collected from them were used to calculate H^2 . 299 The samples were then ground in a ball mill (MM 04, Retsch, GmbH, Haan, 300 Germany) into to a fine powder. The freeze-dried samples were extracted 301 as previously described before analysis using an LC-ESI-QQQ-MS/MS 302 system⁴³. 303

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 described⁴³.

311

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

in a total volume of 200 μ l containing 200 μ M substrates, 5 mM MgCl₂ and totally 500 ng purified protein in Tris-HCl buffer (100 mM, pH = 7.4). After incubating for 15 mins, the reaction was stopped by adding 300 μ l of hexane and vortexing. The organic phase was then filtered through a 0.2 μ m filter (Millipore) before being used for GC-MS analysis. Peak identification of each component was confirmed using authentic samples analysis.

Gene expression in yeast and enzyme assay. Purified PCR products were cloned 319 into the *pEASY*-Blunt Cloning Vector (Transgen, catalog number: CB101) 320 and sequenced for errors. The full cDNAs of CYP450s from Nipponbare 321 (O. sativa L. spp. japonica) were cloned into the pESC-URA vector 322 323 (Stratagene, Accession #AF063585) expression vector. The primers listed in Supplementary Table 9. The constructed vectors were transformed into 324 the yeast strain WAT11 using the lithium acetate method following the 325 manufacturer described protocol (ZYMO RESEARCH, catalog number: 326 T2001). Yeast cultures were grown and microsomes were prepared as 327 previously described with some modification⁴⁴. Briefly, the recombinant 328 cells were first cultured in SC minimal medium containing 2% glucose at 329 30°C. For protein induction, cells were collected and resuspended in 330 Synthetic Complete Medium yeast minimal medium) containing 2% 331 galactose instead of glucose (http://fungenome.bioon.com.cn), and 332 cultured 30°C for 2 days. Cells were harvested by centrifugation and 333

broken with glass beads (0.45 mm in diameter, SIGMA) in 50 mM Tri-HCl 334 buffer, PH = 7.5, containing 1 mM EDTA and 600 mM sorbitol. The cells 335 were broken using a mix mill (Model MM 400, Retsch, Haan, Germany). 336 The homogenate was centrifuged for 60 min at 12,000g and the resulting 337 supernatant was centrifuged for 90 min at 120,000g. The pellet consisting 338 of microsomal membranes was resuspended in 100 mM Tris-HCl, PH = 339 7.5, 1 mM EDTA, and 20% (v/v) glycerol and stored at -80°C for long term 340 storage. 341

In vitro enzymatic activity assays were performed on a shaking 342 incubator (120 rpm), at 30°C for 4 h in 500 μ l of 100 mM Tris-HCl, PH = 343 7.5, containing 1 mg total microsomal proteins, 500 mM NADPH, 200 µM 344 345 substrate. Reactions were stopped by addition of 500 µl of hexane and vortexing. Negative control reactions by were carried out with microsomal 346 preparations from recombinant yeast transformed with 'empty' pESC-347 URA. Total protein content was estimated by measuring UV absorbance at 348 280 nm on NanoDrop ND-1000 spectrophometer. 349

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)

16

600 of 2.0 in 10ml of Luria-Bertani (LB) medium containing 50µg/mL 356 Kanamycin, washed with washing buffer (10 mM 2-(N-morpholino)) 357 ethanesulfonic acid [MES], pH = 5.6), and resuspended in MMA buffer (10 358 mM MES [pH = 5.6], 10 mM MgCl₂, 100 mM acetosyringone) to OD600 359 of 1.0. The culture was incubated for 2 hours in room temperature and one 360 milliliter of culture was used to infiltrate the underside of 6-week-old N. 361 *benthamiana* leaves with a needleless 1 mL syringe⁴⁶. Leaves were 362 harvested 3 days post infiltration, flash frozen and stored at -80°C for later 363 processing. 364

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.

370 **Data availability**

The sequences data of 424 O. sativa accessions is available in NCBI 371 BioProject PRJNA171289¹⁷. The under the single nucleotide 372 polymorphisms (SNPs) information of 424 O. sativa accessions is 373 available in RiceVarMap (http://ricevarmap.ncpgr.cn/v1). The pan-genome 374 data were acquired from the pan-genome dataset 375 (https://figshare.com/collections/Novel sequences structural variations 376

17

and gene presence variations of Asian cultivated rice/3876022/1 and 377 http://cgm.sjtu.edu.cn/3kricedb/)^{18,19,24}. 13 of O. rufipogon were selected 378 from 446 diverse O. rufipogon accessions from Asia and Oceania, and 379 represented all the major genetically distinct clusters in O. rufipogon and 380 the other 10 wild *EnsemblPlants* rice are from 381 (http://plants.ensembl.org/index.html) and National Genomics Data Center 382 (https://bigd.big.ac.cn/search?dbId=gwh&q=Oryza), including Oryza 383 barthii (AA), Oryza glumipatula (AA), Oryza glaberrima (AA), Oryza 384 meridionalis (AA), Oryza longistaminata (AA), Oryza nivara (AA), Oryza 385 brachyantha (FF), Oryza punctata (BB) and Oryza brachyantha (GG)²⁴. 386 Genes reported in the study are deposited in the National Center for 387 Biotechnology Information (NCBI). The genes can be found in GenBank 388 Rice Genome Project database Annotation or 389 (http://rice.plantbiology.msu.edu/analyses search locus.shtml) under the 390 following accession numbers: OsTPS28, MN833254; OsCYP71Z21, 391 LOC Os07g11870; OsCYP71Z2, LOC Os07g11739; OsCYP71Z22, 392 LOC Os07g11970; OsCYP71Z30, LOC Os07g11890. 393

394 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.

403

404 **Author contributions**

- J.L. designed the research. J.L., L.-L.C., L.Q., M.Y. and X.L. supervised
- 406 this study. C.Z., Long L., S.Z., Z.L., F.Z., M.Z., Y.S., Yuheng S., K.L., T.C.,

407 M.H., I.G., Z.Y. and T.T. participated in the material preparation. C.Z., C.Y.,

408 Y.L., X.W. and J.S. carried out the metabolite analyses. C.Z., Z.L., S.Z.,

409 C.Y., X.Z., H.G., M.P., M.Z., Yufei L., Z.Y., Ling L., S.S., J.S., X.J., Y.L.,

410 T.T. and Z.W. performed the data analyses. C.Z., Long L., Z.L., S.Z. and

- 411 C.Y. performed most of the experiments; J.L., C.Z., I.G. and A.R.F. wrote
- the manuscript.

413

414 **Competing interests**

The authors declare no conflict of interests.

416 **References**

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

422		trimethylation is involved in defense-related gene activation in rice. <i>Plant Cell</i> 25 , 4725-4736
423		(2013).
424 425	4.	Chen, X. <i>et al.</i> Biological activities and potential molecular targets of cucurbitacins: a focus on cancer. <i>Anti-Cancer Drug</i> 23 , 777-787 (2012).
426	5.	Huang, A.C. et al. A specialized metabolic network selectively modulates Arabidopsis root
427		microbiota. Science 364, eaau6389 (2019).
428	6.	Nützmann, HW., Huang, A. & Osbourn, A. Plant metabolic clusters - from genetics to
429		genomics. <i>New Phytol.</i> 211 , 771-789 (2016).
430	7.	Zi, J., Mafu, S. & Peters, R.J. To gibberellins and beyond! surveying the evolution of
431		(di)terpenoid metabolism. Annu. Rev. Plant. Biol. 65, 259-286 (2014).
432	8.	Panizza, B.J. et al. Phase I dose-escalation study to determine the safety, tolerability,
433		preliminary efficacy and pharmacokinetics of an intratumoral injection of tigilanol tiglate (EBC-
434		46). Ebiomedicine 50 , 433-441 (2019).
435	9.	Hezareh, M. Prostratin as a new therapeutic agent targeting HIV viral reservoirs. Drug News
436		Perspect. 18, 496-500 (2005).
437	10.	Johnson, H.E., Banack, S.A. & Cox, P.A. Variability in content of the anti-AIDS drug candidate
438		prostratin in samoan populations of homalanthus nutans. J. Nat. Prod. 71, 2041-2044 (2008).
439	11.	Lebwohl, M. et al. Ingenol mebutate gel for actinic keratosis. N. Engl. J. Med. 366, 1010-1019
440		(2012).
441	12.	Sabandar, C.W., Ahmat, N., Jaafar, F.M. & Sahidin, I. Medicinal property, phytochemistry and
442		pharmacology of several Jatropha species (Euphorbiaceae): A review. Phytochemistry 85, 7-29
443		(2013).
444	13.	Inoue, Y. et al. Identification of a novel casbane-type diterpene phytoalexin, ent-10-
445		oxodepressin, from rice leaves. Biosci. Biotech. Bioch. 77, 760-765 (2013).
446	14.	Horie, K., Sakai, K., Okugi, M., Toshima, H. & Hasegawa, M. Ultraviolet-induced amides and
447		casbene diterpenoids from rice leaves. Phytochem. Lett. 15, 57-62 (2016).
448	15.	Li, W. et al. OsCYP71Z2 involves diterpenoid phytoalexin biosynthesis that contributes to
449		bacterial blight resistance in rice. Plant Sci. 207, 98-107 (2013).
450	16.	Li, W. et al. Overexpressing CYP71Z2 enhances resistance to bacterial blight by suppressing
451		auxin biosynthesis in rice. Plos One 10, e0119867 (2015).
452	17.	Zhao, H. et al. RiceVarMap: a comprehensive database of rice genomic variations. Nucleic Acids
453		<i>Res.</i> 43 , D1018-D1022 (2015).
454	18.	Wang, W. et al. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature
455		557 , 497-501 (2018).
456	19.	Huang, X. et al. A map of rice genome variation reveals the origin of cultivated rice. Nature 490,
457		497-501 (2012).
458	20.	Chen, W. et al. Genome-wide association analyses provide genetic and biochemical insights
459		into natural variation in rice metabolism. Nat. Genet. 46, 714-721 (2014).
460	21.	Yu, N. et al. Delineation of metabolic gene clusters in plant genomes by chromatin signatures.
461		Nucleic Acids Res. 44 , 2255-2265 (2016).
462	22.	Zhou, S. et al. Cooperation between the H3K27me3 chromatin mark and non-CG methylation
463		in epigenetic regulation. Plant Physiol. 172, 1131-1141 (2016).
464	23.	Nützmann, HW. et al. Active and repressed biosynthetic gene clusters have spatially distinct
465		chromosome states. Proc. Natl. Acad.Sci.U.S.A. 24, 13800-13809 (2020).

466	24.	Stein, J.C. et al. Genomes of 13 domesticated and wild rice relatives highlight genetic
467		conservation, turnover and innovation across the genus Oryza. Nat. Genet. 50, 285-296 (2018).
468	25.	Wing, R.A., Purugganan, M.D. & Zhang, Q. The rice genome revolution: from an ancient grain
469		to green super rice. Nat. Rev. Genet. 19 , 505-517 (2018).
470	26.	King, A.J., Brown, G.D., Gilday, A.D., Larson, T.R. & Graham, I.A. Production of bioactive
471		diterpenoids in the Euphorbiaceae depends on evolutionarily conserved gene clusters. Plant
472		<i>Cell</i> 26 , 3286-3298 (2014).
473	27.	Takos, A.M. et al. Genomic clustering of cyanogenic glucoside biosynthetic genes aids their
474		identification in Lotus japonicus and suggests the repeated evolution of this chemical defence
475		pathway. <i>Plant J.</i> 68 , 273-286 (2011).
476	28.	Field, B. & Osbourn, A.E. Metabolic diversification-independent assembly of operon-like gene
477		clusters in different plants. Science 320, 543-547 (2008).
478	29.	M, W. et al. Parallel selection on a dormancy gene during domestication of crops from multiple
479		families. Nat. Genet. 50, 1435-1441 (2018).
480	30.	Swaminathan, S., Morrone, D., Wang, Q., Fulton, D.B. & Peters, R.J. CYP76M7 Is an ent-
481		cassadiene C11-hydroxylase defining a second multifunctional diterpenoid biosynthetic gene
482		cluster in rice. Plant Cell 21 , 3315-3325 (2009).
483	31.	Wang, Q., Hillwig, M.L. & Peters, R.J. CYP99A3: functional identification of a diterpene oxidase
484		from the momilactone biosynthetic gene cluster in rice. <i>Plant J.</i> 65 , 87-95 (2011).
485	32.	Wang, Q. et al. Characterization of CYP76M5-8 indicates metabolic plasticity within a plant
486		biosynthetic gene cluster. J. Biol. Chem. 287, 6159-68 (2012).
487	33.	Miyamoto, K. et al. Evolutionary trajectory of phytoalexin biosynthetic gene clusters in rice.
488		Plant J. 87 , 293-304 (2016).
489	34.	Gross, B.L. & Zhao, Z. Archaeological and genetic insights into the origins of domesticated rice.
490		Proc. Natl. Acad.Sci.U.S.A. 111, 6190-6197 (2014).
491	35.	Qi, Z. Genetics and improvement of bacterial blight resistance of hybrid rice in China. Rice Sci.
492		23 , 111-119 (2009).
493	36.	Kauffman, H.E., Reddy, A.P.K., Hsieh, S.P.Y. & Merca, S.D. An improved technique for evaluating
494		resistance of rice varieties to Xanthomonas oryzae. Plant dis. rep. 57, 537-541 (1973).
495	37.	Zhang, H. et al. Transposon-derived small RNA is responsible for modified function of WRKY45
496		locus. Nat. Plants 2 , 16016 (2016).
497	38.	Zhang, B. et al. Multiple alleles encoding atypical NLRs with unique central tandem repeats in
498		rice confer resistance to Xanthomonas oryzae pv. oryzae. Plant Commun. 1, 100088 (2020).
499	39.	Tao, Z. et al. A pair of allelic WRKY genes play opposite roles in rice-bacteria interactions. Plant
500		physiol. 151 , 936-948 (2009).
501	40.	King, A.J. et al. A cytochrome P450-mediated intramolecular carbon-carbon ring closure in the
502		biosynthesis of multidrug-resistance-reversing lathyrane diterpenoids. Chembiochem 17,
503		1593-1597 (2016).
504	41.	Boutanaev, A.M. et al. Investigation of terpene diversification across multiple sequenced plant
505		genomes. Proc. Natl. Acad.Sci.U.S.A. 112, E81-E88 (2015).
506	42.	Luo, D. et al. Oxidation and cyclization of casbene in the biosynthesis of Euphorbia factors from
507		mature seeds of Euphorbia lathyris L. Proc. Natl. Acad.Sci.U.S.A. 113, E5082-E5089 (2016).
508	43.	Peng, M. et al. Differentially evolved glucosyltransferases determine natural variation of rice
509		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).

513 45. Sainsbury, F., Thuenemann, E.C. & Lomonossoff, G.P. pEAQ: versatile expression vectors for
514 easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol. J.* 7,
515 682-693 (2009).

adaptation of qingke. Mol. Plant 13, 112-127 (2020).

Zeng, X. et al. Genome-wide dissection of co-selected UV-B responsive pathways in the UV-B

516 517 46.

518 519

520

Fig. 1 | Identification the structural variation of a diterpene gene cluster in rice. a, 521 Manhattan plot of 5,10-diketo-casbene trait across the 12 rice chromosomes. In 522 japonica population, all metabolite-SNP associations with P values below 1.8819e-7 523 (horizontal dashed lines in all Manhattan plots) are plotted against genome location in 524 intervals of 1Mb. The two-tailed Student's t test is performed for this analysis and the 525 Bonferroni correction is used for the multiple-comparison correction. The Manhattan 526 plots from two individual replicate for each locus are provided in Supplementary Fig.2. 527 TPS, synthase; CYP450, Chr. chromosome; terpene cytochromes P450 528 monooxygenases; TE, transposable element. **b**, A 150-kb insertion in Nipponbare (Nip) 529 contains DGC7 not present at the syntenic location in Minghui63 (MH63), Zhenshan97 530 (ZS97). c, The model of two types of DGC7. d, The relative content of 5,10-diketo-531 casbene in different subspecies of rice. e, Relative content of 5,10-diketo-casbene 532 subjected to 0.1mM methyl jasmonate treatment for 24 hours in Nip, MH63, ZS97. CK, 533 control check; nd, not detected. The data are presented as mean \pm s.d., n=3 biologically 534 independent replicates. The asterisks in Fig.1e indicate significant differences 535 compared with the CK: ****P<0.0001 by unpaired two-tailed Student's *t* tests. **f**, The 536 relative content of 5,10-diketo-casbene in the randomly selected varieties. Presence or 537

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

541

Fig. 2 | Identification of a diterpene gene cluster. a, Subcellular localization pattern 542 of the confirmed OsTPS28. Transient expression of confirmed OsTPS28 fused to GFP 543 in rice leaf protoplasts showing chloroplast localization. Bar=10µm. All experiments 544 were repeated three times with similar results. **b**, Metabolic profiling of *N*. benthamiana 545 546 leaves using ultra-performance liquid chromatography coupled with QQQ mass spectrometry (LC-ESI-QQQ-MS/MS) with and without the infiltration of the 547 corresponding candidates. 10-keto-casbene and 5,10-diketo-casbene reference 548 compounds were purified from rice leaves by the method described previously^{13,14}. GFP, 549 green fluorescent protein. c, Hierarchical clustering of RNA-Seq expression data. Color 550 key: known diterpene biosynthesis genes (gray), genes identified in this report 551 552 (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 553 methyl jasmonate treatment are indicated. 554

555

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 \pm sd, *n*=3 biologically independent replicates. **b**, H3K27me3 ChIP-on-chip data for the genes from *DGC7*. The data is extracted from³. **c**, H3K27me3 ChIP analysis for the genes

560 from *DGC7* in seedlings. Transcript levels were analyzed by qPCR. The data are

presented as mean \pm s.d., n=3 biologically independent replicates. d, The relative 561 content of 5,10-diketo-casbene in the JMJ705 overexpression line. The data are 562 presented as mean \pm s.d., n=3 biologically independent replicates. e, The relative 563 content of 5,10-diketo-casbene in the JMJ705 RNAi line. The data are presented as 564 mean \pm s.d., n=3 biologically independent replicates. The asterisks in Fig. 3a, 3c-e 565 indicate significant differences compared with the CK or ZH11: *P<0.05, **P<0.01, 566 ***P<0.001, ****P<0.0001 by unpaired two-tailed Student's t tests. Source data are 567 provided as a Source Data file. 568

569

Fig. 4 | The origin of DGC7. a, Phylogenetic analysis show an TPS-II clade across in 570 the Oryza species. b, The OsCYP71Z2/OsCYP71Z1/OsCYP71Z21-OsCYP71Z22 tree 571 shows the latest duplications are likely to have occurred prior to divergence of O. 572 punctata (BB genome type) and O. sativa (AA genome type). Leersia perrieri is the 573 evolutionally closest outgroup species for Oryza. c, The selection of DGC7. The 574 relative proportion of six types of gene modules. The intact DGC7 is absent in O. 575 rufipogon, highly enriched in japonica varieties but not in indica varieties. The data 576 extracted from^{18,19}. 577

578

579 Extended Data Fig. 1 | The distribution of the world-wide collection of rice 580 accessions in this study. The core collection of 424 cultivated rice accessions in this 581 study has a wide geographic distribution. Color dots indicate different subspecies/type 582 of cultivated rice. The map is draw by R 3.1 and the information of Latitude and 583 Longitude of the rice varieties have also been shown in the Supplementary Table 1.

584

Extended Data Fig. 2 | Functional analyses of OsTPS28, OsCYP71Z2 and 585 OsCYP71Z21. a, Gas chromatography of the reaction products of OsTPS28 with 586 GGDP. GGDP, geranylgeranyl diphosphate. Casbene and neocembrene reference 587 compounds were purified from infiltrated N. benthamiana leaves by the method 588 described previously²⁶. Compound 1, casbene; Compound 2, neocembrene. **b**, Gas 589 chromatography of in vitro enzyme assays showing the 10-keto-casbene C5 oxidase 590 activity of yeast-expressed CYP71Z2 in the present of NADPH. Microsomes prepared 591 from yeast containing PESC-URA empty vector were used as a negative control. 10-592 keto-casbene reference compound was purified from rice leaves by the method 593 described previously^{13,14}. Compound 3, 10-keto-casbene; Compound 4, 5,10-diketo-594 casbene. c, Gas chromatography of the extracts prepared from the leaves of N. 595 benthamiana infiltrated with OsTPS28 over-expressing vector. 596 597

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 5,10diketo-casbene reference. LC-MS, liquid chromatography-mass spectrometry. c, ¹H
NMR (left) and ¹³C 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
- **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.