

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:

<https://eprints.whiterose.ac.uk/168087/>

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](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

# 1 Selection of a subspecies-specific diterpene gene 2 cluster implicated in rice disease resistance

3

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

9

10 <sup>1</sup>National Key Laboratory of Crop Genetic Improvement and National Center of Plant Gene Research (Wuhan),  
11 Huazhong Agricultural University, Wuhan 430070, China.

12 <sup>2</sup>College of Tropical Crops, Hainan University, Haikou 572208, China.

13 <sup>3</sup>Centre for Novel Agricultural Products, Department of Biology, University of York, York, United Kingdom.

14 <sup>4</sup>College of Agriculture, Ibaraki University, 3-21-1 Chuo, Ami, Ibaraki 300-0393, Japan.

15 <sup>5</sup>Graduate School of Biological Sciences, Nara Institute of Science and Technology, Ikoma, Nara, 630-0192, Japan.

16 <sup>6</sup>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<sup>1,2</sup>. 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<sup>3</sup>. Functional**  
24 **characterization of pathway genes revealed *OsCYP71Z21* to encode a casbene C10**  
25 **oxidase, sought after for the biosynthesis of an array of medically important**

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

34

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

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

54 5,10-diketo-casbene (previously referred to as *ent*-10-oxodepressin)  
55 was the first casbene-type diterpene phytoalexin found in rice (*Oryza*  
56 *sativa*) that confers rice bacterial blight and rice blast fungus resistance<sup>13-</sup>  
57 <sup>16</sup>. However, surprisingly to date no study has yet assessed the natural  
58 variation in the ability to produce 5,10-diketo-casbene. More than 4,000  
59 diverse accessions of *O. sativa* (*indica* and *japonica*) and various wild rice  
60 relatives have been sequenced in recent years allowing the generation of a  
61 detailed genome-variation map<sup>17-19</sup> and the opportunity to perform Genome  
62 Wide Association Studies to locate the genetic basis of traits exhibiting  
63 natural variation.

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

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

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

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

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

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

114 synthases (Fig. 2a). The *OsTPS28* ORF minus the transit peptide was  
115 expressed in *Escherichia coli* BL21 and recombinant protein produced  
116 casbene in the presence of GGDP and  $Mg^{2+}$  (Extended Data Fig. 2a and  
117 Supplementary Fig. 5, a-d), with a  $K_m$  of 5.16 $\mu$ M and  $K_{cat}$  of 0.0236  $s^{-1}$   
118 (Supplementary Fig. 6). Stable transformation analysis in rice revealed that  
119 5,10-diketo-casbene was increased by ~1.9 fold in *OsTPS28*  
120 overexpression lines while reduced to non-detectable levels in the  
121 *OsTPS28* knockout plants (Supplementary Fig. 7).

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

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

155 To explore spatiotemporal expression of the members of *DGC7*, we  
156 collected samples at different stages from different parts of rice and carried  
157 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 (qRT-PCR) 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<sup>21,22</sup>. 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)<sup>21,23</sup>. 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)<sup>3</sup>. 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

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

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

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

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

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

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

246 3). Results from  $\pi$  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)<sup>29</sup>.

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<sup>30-33</sup>, *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<sup>34</sup>. Given that Southern China has  
258 been an endemic area of rice bacterial blight and 5,10-diketo-casbene  
259 confers rice blast resistance<sup>14,15,35</sup> 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*<sup>1,14</sup>. It can be induced by the rice blast fungus and furthermore  
265 inhibits rice blast fungus spore germination and germ tube growth<sup>13</sup>.  
266 Moreover, overexpression of *OsCYP71Z2* (one gene member of *DGC7*) in  
267 rice can enhance the resistance of rice to bacterial blight resistance<sup>15,16</sup>.

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

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

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

## 292 **Methods**

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

304 **Recombinant protein expression, purification and enzyme assay.** The 5' and 3'  
305 ends of the targeted TPSs were cloned by RACE (Takara, catalog number:  
306 634858) according to the manufacturer's directions. The full cDNAs of  
307 TPSs from Nipponbare (*O. sativa* L. spp. *japonica*) were cloned into the  
308 pGEX-6p-1 expression vector (Novagen) with a *Glutathione-S-transferase*  
309 (*GST*). The primers listed in Supplementary Table 9. Recombinant proteins  
310 were expressed in BL21 (Novagen) as previously described<sup>43</sup>.

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

312 in a total volume of 200  $\mu$ l containing 200  $\mu$ M substrates, 5 mM MgCl<sub>2</sub> 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  $\mu$ l of  
315 hexane and vortexing. The organic phase was then filtered through a 0.2  
316  $\mu$ m 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<sup>44</sup>. 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



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

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

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

356 600 of 2.0 in 10ml of Luria-Bertani (LB) medium containing 50µg/mL  
357 Kanamycin, washed with washing buffer (10 mM 2-(N-morpholino)  
358 ethanesulfonic acid [MES], pH = 5.6), and resuspended in MMA buffer (10  
359 mM MES [pH = 5.6], 10 mM MgCl<sub>2</sub>, 100 mM acetosyringone) to OD600  
360 of 1.0. The culture was incubated for 2 hours in room temperature and one  
361 milliliter of culture was used to infiltrate the underside of 6-week-old *N.*  
362 *benthamiana* leaves with a needleless 1 mL syringe<sup>46</sup>. Leaves were  
363 harvested 3 days post infiltration, flash frozen and stored at -80°C for later  
364 processing.

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

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

### 370 **Data availability**

371 The sequences data of 424 *O. sativa* accessions is available in NCBI  
372 under the BioProject PRJNA171289<sup>17</sup>. The single nucleotide  
373 polymorphisms (SNPs) information of 424 *O. sativa* accessions is  
374 available in RiceVarMap (<http://ricevarmap.ncpgr.cn/v1>). The pan-genome  
375 data were acquired from the pan-genome dataset  
376 ([https://figshare.com/collections/Novel\\_sequences\\_structural\\_variations\\_](https://figshare.com/collections/Novel_sequences_structural_variations_)

377 and\_gene\_presence\_variations\_of\_Asian\_cultivated\_rice/3876022/1 and  
378 <http://cgm.sjtu.edu.cn/3kricedb/>)<sup>18,19,24</sup>. 13 of *O. rufipogon* were selected  
379 from 446 diverse *O. rufipogon* accessions from Asia and Oceania, and  
380 represented all the major genetically distinct clusters in *O. rufipogon* and  
381 the other 10 wild rice are from *EnsemblPlants*  
382 (<http://plants.ensembl.org/index.html>) and National Genomics Data Center  
383 (<https://bigd.big.ac.cn/search?dbId=gwh&q=Oryza>), including *Oryza*  
384 *barthii* (AA), *Oryza glumipatula* (AA), *Oryza glaberrima* (AA), *Oryza*  
385 *meridionalis* (AA), *Oryza longistaminata* (AA), *Oryza nivara* (AA), *Oryza*  
386 *brachyantha* (FF), *Oryza punctata* (BB) and *Oryza brachyantha* (GG)<sup>24</sup>.  
387 Genes reported in the study are deposited in the National Center for  
388 Biotechnology Information (NCBI). The genes can be found in GenBank  
389 or Rice Genome Annotation Project database  
390 ([http://rice.plantbiology.msu.edu/analyses\\_search\\_locus.shtml](http://rice.plantbiology.msu.edu/analyses_search_locus.shtml)) under the  
391 following accession numbers: OsTPS28, MN833254; OsCYP71Z21,  
392 LOC\_Os07g11870; OsCYP71Z2, LOC\_Os07g11739; OsCYP71Z22,  
393 LOC\_Os07g11970; OsCYP71Z30, LOC\_Os07g11890.

### 394 **Acknowledgements**

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

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

403

#### 404 **Author contributions**

405 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  
412 the manuscript.

413

#### 414 **Competing interests**

415 The authors declare no conflict of interests.

#### 416 **References**

- 417 1. Schmelz, E.A. *et al.* Biosynthesis, elicitation and roles of monocot terpenoid phytoalexins. *Plant*  
418 *J.* **79**, 659-78 (2014).  
419 2. Lu, X. *et al.* Inferring roles in defense from metabolic allocation of rice diterpenoids. *Plant Cell*  
420 **30**, 1119-1131 (2018).  
421 3. Li, T. *et al.* Jumonji C domain protein JM1705-mediated removal of histone H3 lysine 27

- 422 trimethylation is involved in defense-related gene activation in rice. *Plant Cell* **25**, 4725-4736  
423 (2013).
- 424 4. Chen, X. *et al.* Biological activities and potential molecular targets of cucurbitacins: a focus on  
425 cancer. *Anti-Cancer Drug* **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, H.-W., Huang, A. & Osbourn, A. Plant metabolic clusters – from genetics to  
429 genomics. *New Phytol.* **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 *Res.* **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, H.-W. *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 *Cell* **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. *Plant J.* **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. *Plant J.* **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).

- 510 44. Ikezawa, N. *et al.* Lettuce costunolide synthase (CYP71BL2) and its homolog (CYP71BL1) from  
511 sunflower catalyze distinct regio- and stereoselective hydroxylations in sesquiterpene lactone  
512 metabolism. *J. Biol. Chem.* **286**, 21601-21611 (2011).
- 513 45. Sainsbury, F., Thuenemann, E.C. & Lomonosoff, 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).
- 516 46. Zeng, X. *et al.* Genome-wide dissection of co-selected UV-B responsive pathways in the UV-B  
517 adaptation of qingke. *Mol. Plant* **13**, 112-127 (2020).

518  
519  
520

521 **Fig. 1 | Identification the structural variation of a diterpene gene cluster in rice. a,**

522 Manhattan plot of 5,10-diketo-casbene trait across the 12 rice chromosomes. In  
523 *japonica* population, all metabolite-SNP associations with *P* values below 1.8819e-7  
524 (horizontal dashed lines in all Manhattan plots) are plotted against genome location in  
525 intervals of 1Mb. The two-tailed Student's *t* test is performed for this analysis and the  
526 Bonferroni correction is used for the multiple-comparison correction. The Manhattan  
527 plots from two individual replicate for each locus are provided in Supplementary Fig.2.

528 Chr., chromosome; TPS, terpene synthase; CYP450, cytochromes P450  
529 monooxygenases; TE, transposable element. **b,** A 150-kb insertion in Nipponbare (Nip)

530 contains *DGC7* not present at the syntenic location in Minghui63 (MH63), Zhenshan97  
531 (ZS97). **c,** The model of two types of *DGC7*. **d,** The relative content of 5,10-diketo-

532 casbene in different subspecies of rice. **e,** Relative content of 5,10-diketo-casbene  
533 subjected to 0.1mM methyl jasmonate treatment for 24 hours in Nip, MH63, ZS97. CK,  
534 control check; nd, not detected. The data are presented as mean  $\pm$  s.d., *n*=3 biologically  
535 independent replicates. The asterisks in Fig.1e indicate significant differences

536 compared with the CK: \*\*\*\**P*<0.0001 by unpaired two-tailed Student's *t* tests. **f,** The  
537 relative content of 5,10-diketo-casbene in the randomly selected varieties. Presence or

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

541

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

555

556 **Fig. 3 | The regulation of *DGC7*. a**, Gene expression levels of *OsCYP71Z2*, *OsTPS28*,  
557 *OsCYP71Z21* in MeJA treated and control plants. The data are presented as mean  $\pm$  sd,  
558  $n=3$  biologically independent replicates. **b**, H3K27me3 ChIP-on-chip data for the genes  
559 from *DGC7*. The data is extracted from<sup>3</sup>. **c**, H3K27me3 ChIP analysis for the genes  
560 from *DGC7* in seedlings. Transcript levels were analyzed by qPCR. The data are



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

569

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

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

585 **Extended Data Fig. 2 | Functional analyses of OsTPS28, OsCYP71Z2 and**

586 **OsCYP71Z21. a,** Gas chromatography of the reaction products of OsTPS28 with

587 GGDP. GGDP, geranylgeranyl diphosphate. Casbene and neocembrene reference

588 compounds were purified from infiltrated *N. benthamiana* leaves by the method

589 described previously<sup>26</sup>. Compound 1, casbene; Compound 2, neocembrene. **b,** Gas

590 chromatography of *in vitro* enzyme assays showing the 10-keto-casbene C5 oxidase

591 activity of yeast-expressed CYP71Z2 in the present of NADPH. Microsomes prepared

592 from yeast containing PESC-URA empty vector were used as a negative control. 10-

593 keto-casbene reference compound was purified from rice leaves by the method

594 described previously<sup>13,14</sup>. Compound 3, 10-keto-casbene; Compound 4, 5,10-diketo-

595 casbene. **c,** Gas chromatography of the extracts prepared from the leaves of *N.*

596 *benthamiana* infiltrated with OsTPS28 over-expressing vector.

597

598 **Extended Data Fig. 3 | Mass spectrum and structure of 5,10-diketo-casbene. a,**

599 Mass spectrum and structure of the product in *N. benthamiana* leaves simultaneously

600 overexpressing *OsTPS28*, *OsCYP71Z2* and *OsCYP71Z21*. **b,** Mass spectrum of 5,10-

601 diketo-casbene reference. LC-MS, liquid chromatography-mass spectrometry. **c,** <sup>1</sup>H

602 NMR (left) and <sup>13</sup>C NMR (right) results of 5,10-diketo-casbene.

603

604 **Extended Data Fig. 4 | The expression profiles of genes from DGC7.** The genes from

605 *DGC7* are indicated in bold. The transcript abundances of indicated genes in different

606 organs at different stages were shown: expression levels of *OsTPS28*, *OsCYP17Z2* and

607 *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.