nature genetics

Control of rice grain-filling and yield by a gene with a potential signature of domestication

Ertao Wang[1,2], Jianjun Wang[3], Xudong Zhu[2], Wei Hao[1], Linyou Wang[3], Qun Li[1], Lixia Zhang[3], Wei He[1], Baorong Lu[4], Hongxuan Lin[1], Hong Ma[5], Guiquan Zhang[6] & Zuhua He[1]

Grain-filling, an important trait that contributes greatly to grain weight, is regulated by quantitative trait loci and is associated with crop domestication syndrome[1–4]. However, the genes and underlying molecular mechanisms controlling crop grain-filling remain elusive. Here we report the isolation and functional analysis of the rice *GIF1* (*GRAIN INCOMPLETE FILLING 1*) gene that encodes a cell-wall invertase required for carbon partitioning during early grain-filling. The cultivated *GIF1* gene shows a restricted expression pattern during grain-filling compared to the wild rice allele, probably a result of accumulated mutations in the gene's regulatory sequence through domestication. Fine mapping with introgression lines revealed that the wild rice *GIF1* is responsible for grain weight reduction. Ectopic expression of the cultivated *GIF1* gene with the *35S* or rice *Waxy* promoter resulted in smaller grains, whereas overexpression of *GIF1* driven by its native promoter increased grain production. These findings, together with the domestication signature that we identified by comparing nucleotide diversity of the *GIF1* loci between cultivated and wild rice, strongly suggest that *GIF1* is a potential domestication gene and that such a domestication-selected gene can be used for further crop improvement.

High yield has been a major breeding target in cereals, including rice (*Oryza sativa* L.), a staple food crop and a model monocot with the smallest genome of major cereals[5]. The rice yield trait consists of several key components, including grain number and grain weight, and is regulated by a number of quantitative trait loci (QTLs) derived from natural allelic variations[3]. A few rice genes corresponding to some yield QTLs, such as *Gn1a*, *GW2*, *GS3* and *Ghd7*, were recently isolated[6–10]. The duration and rate of grain-filling, which determine final grain weight and thereby contribute greatly to grain productivity, are also controlled by QTLs[1,11]. Because of the difficulty in measuring natural variations in grain-filling or weight, map-based cloning of the genes controlling grain-filling has been a major challenge.

We screened for mutants with grain-filling defects in our mutant population (*O. sativa japonica* Zhonghua 11)[12]. One mutant, *gif1* (*grain incomplete filling 1*), showed slower grain-filling than wild-type rice (**Fig. 1** and **Supplementary Fig. 1** online). The *gif1* mutant also showed markedly more grain chalkiness as a result of abnormally developed and loosely packed starch granules (**Fig. 1c–f**). The mutant was morphologically normal, with normal seed setting (**Supplementary Table 1** online). The reduced filling rate resulted in reduced weight of *gif1* grains starting 3 d after pollination (DAP); the final grain weight of the *gif1* mutant was ~24% lower than that of wild-type rice at 30 DAP (**Fig. 1g** and **Supplementary Table 1**). Consistently, amylose and amylopectin levels were significantly lower in *gif1* than in wild-type rice (**Supplementary Fig. 2** online).

We initially mapped the *GIF1* locus on chromosome 4 and then further narrowed it to a 32-kb fragment with three putative genes (**Supplementary Fig. 3** online). Sequencing of the entire region in the *gif1* mutant revealed a 1-nt deletion in the coding region of the *Os04g33740* gene, causing premature termination of its predicted open reading frame. RT-PCR analysis showed that the *Os04g33740* transcript level was greatly reduced in *gif1* grains (**Supplementary Fig. 3**). The *Os04g33740* gene contains seven exons and encodes a protein with 598 amino acids (**Supplementary Figs. 3** and **4** online), and its identity as *GIF1* was verified by functional complementation (**Supplementary Fig. 5** online). Database searches indicated that the GIF1 protein is a putative invertase with conserved motifs and a cysteine catalytic site, sharing high sequence similarity with the known maize Mn1 (ref. 13) and tomato LIN5 (ref. 14) invertases (**Supplementary Fig. 4**). Invertases constitute a large family in plants[15]; GIF1 (also known as OsCIN2) is a member of the cell-wall invertase subfamily with eight members in the rice genome[16].

To test the invertase activity of GIF1, we first measured cell-wall invertase activity in developing wild-type and *gif1* mutant grains. The cell-wall invertase activity of developing *gif1* grains at 7 DAP was only 17% of that in the wild-type grains (**Supplementary Fig. 6** online). This was consistent with the lower glucose and fructose content of *gif1*

[1]Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 300 Fenglin Road, Shanghai 200032, China. [2]China National Rice Research Institute, 359 TiyuChang Road, Hangzhou 31006, China. [3]Zhejiang Academy of Agricultural Sciences, 198 Shiqiao Road, Hangzhou 310021, China. [4]School of Life Sciences, Fudan University, 220 Handan Road, Shanghai 200433, China. [5]Department of Biology and the Huck Institutes of Life Sciences, Pennsylvania State University, University Park, Pennsylvania 16802, USA. [6]College of Agriculture, South China Agricultural University, Guangzhou 510642, China. Correspondence should be addressed to Z.H. (zhhe@sibs.ac.cn).

Received 17 June; accepted 9 July; published online 28 September 2008; doi:10.1038/ng.220

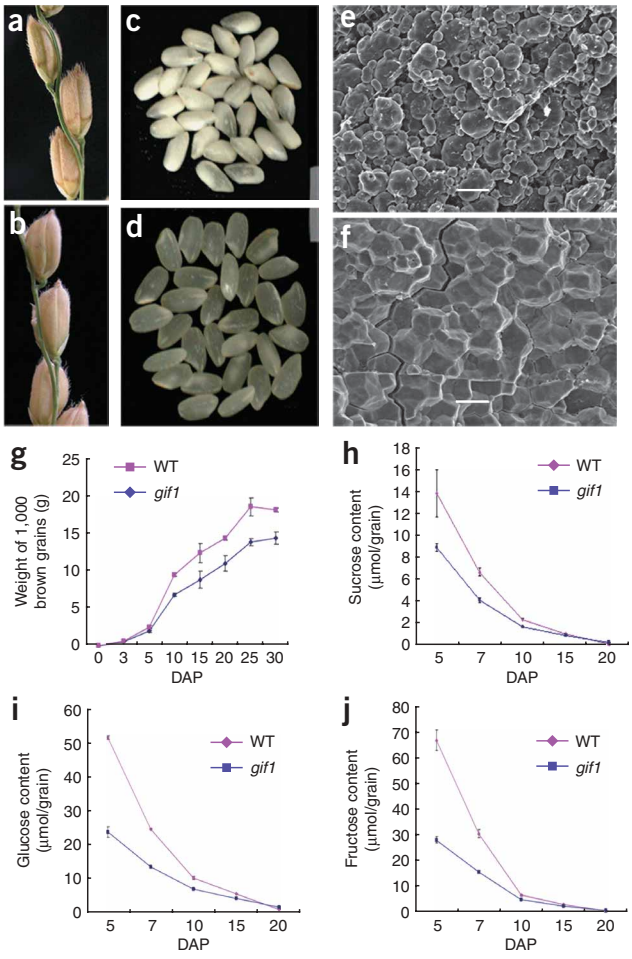


Figure 1 Grain-filling and sugar content of *gif1* mutant and wild-type rice. (**a**,**b**) Grains of *gif1* (**a**) and wild-type (**b**) rice at 25 DAP (see **Supplementary Fig. 1** for seed development). (**c**,**d**) White grains of *gif1* (**c**) and wild-type (**d**) rice. (**e**,**f**) Scanning electron microscope analysis of *gif1* (**e**) and wild-type (**f**) grains. Starch granules developed abnormally and were packed loosely in *gif1* grains. Magnification, ×1,500. Scale bars represent 10 μm. (**g**) Grain-filling process (weight in grams of 1,000 brown grains) of *gif1* and wild-type (WT) rice. (**h**–**j**) Sucrose, glucose and fructose contents of *gif1* and wild-type grains. Data in **g**–**j** are shown as means ± s.e.m.

grains compared to wild-type grains at the early filling stage (**Fig. 1h**–**j**). We then generated transgenic rice plants (GIF1-OE) that ectopically express *GIF1* using the cauliflower mosaic virus *35S* promoter. The insoluble invertase activity was greatly increased in the GIF1-OE root, which were badly filled and severely shrunken (**Supplementary Figs. 6** and **7** online). Furthermore, a fusion protein of GIF1 and green fluorescent protein (GFP) localized to the cell wall of the transgenic root (**Fig. 2**), indicating that GIF1 is indeed a cell-wall invertase.

To investigate the site of GIF1 action, we examined its expression using a pGIF1-β-glucuronidase (GUS) reporter transgene (*pGIF1-GUS*). GUS activity was mainly detected in growing roots, the node and the rapidly elongating zone of the internode, similar to its transcript accumulation (**Fig. 2a**,**b** and **Supplementary Fig. 8** online). Particularly during early grain-filling (**Fig. 2c**,**d**), strong GUS activity was observed in the ovular vascular trace end of the grain at 3 and 5 DAP. At 10 DAP, strong GUS activity was constrained in the ovular vascular and lateral stylar vascular traces (**Fig. 2e**,**g**,**h**). At 20 DAP, GUS activity was observed in the ovular vascular trace but not in the stylar vasculature (**Fig. 2f**). Consistently, RT-PCR analysis showed that the *GIF1* transcript accumulated in filling grains at 0–15 DAP and decreased at 20 DAP (**Fig. 2i**). This *GIF1* expression pattern is consistent with the *gif1* phenotypes that show defects in grain-filling mostly at 3–15 DAP (**Fig. 1g**). It is well known that assimilated carbon partitioning and sucrose metabolism are important to grain-filling in crops[17,18]. Our results suggest that sucrose is unloaded by GIF1 in the ovular and stylar vascular tissues for starch synthesis in the endosperm during grain-filling.

The *gif1* mutant grains accumulated lower levels of glucose and fructose, as well as sucrose, than did the wild-type grains (**Fig. 1h**–**j**), suggesting that sugar metabolism, including sucrose synthesis and unloading, is homeostatically regulated. We used microarray analysis to examine the regulation of genes involved in sugar metabolism (see **Supplementary Methods** online). A total of 44 genes related to starch synthesis and carbohydrate metabolism were significantly up- or downregulated in *gif1* grains compared to wild-type grains at 7 DAP (**Supplementary Tables 2** and **3** online). We propose that sucrose partitioning is altered by the *gif1* mutation, leading to lower sucrose levels in the *gif1* grain, as similarly reported for carrot[19].

Cultivated rice has undergone intensive selection from its progenitor, *Oryza rufipogon*. It has been proposed that many of the agronomically important traits were subjected to selection during rice

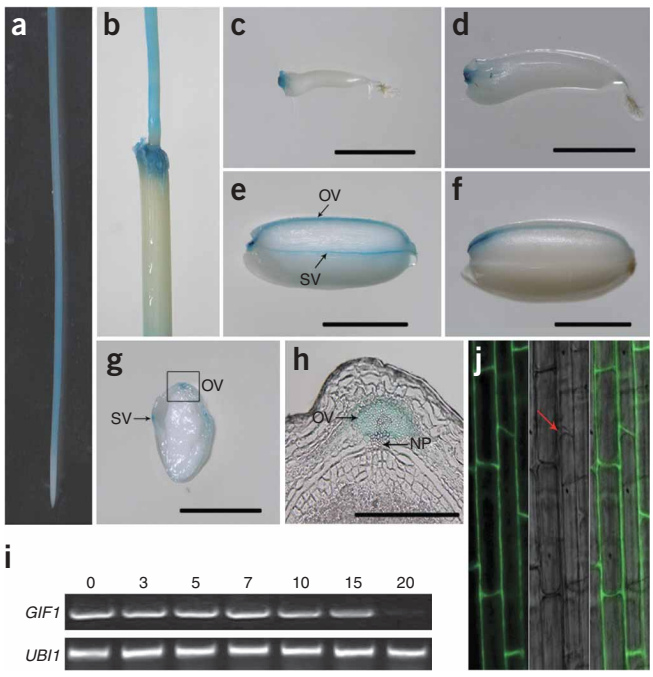
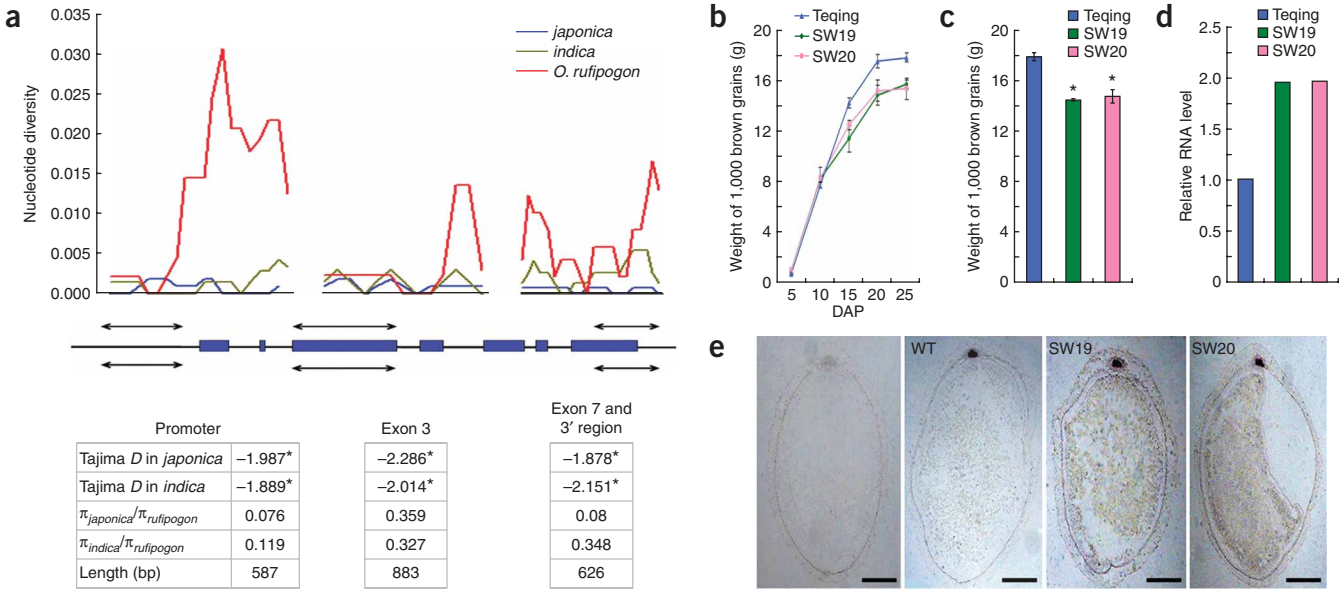


Figure 2 Expression pattern and localization of GIF1. (**a**,**b**) GUS activity in the growing root (**a**) and the node and elongating zone of the internode (**b**). (**c**,**d**) GUS activity in developing grains was restricted to the ovular vascular trace (OV) end at 3 DAP (**c**) and 5 DAP (**d**). (**e**) GUS activity in the OV and lateral stylar vascular trace (SV) of grains at 10 DAP. (**f**) GUS activity in the OV at 20 DAP. (**g**) Cross-section of the grain at 10 DAP. (**h**) Boxed area in **g** observed under a microscope, showing GUS activity in the OV. NP, nucellar projection. (**i**) *GIF1* transcript levels detected by RT-PCR. *UBI1* was used as a loading control. The analysis was repeated twice with similar results. (**j**) GIF-GFP fusion protein localized to the cell wall in a transgenic rice root tip. Arrow indicates plasmolysis. Scale bars represent 3 mm (**c**–**g**) or 100 μm (**h**).



	Promoter	Exon 3	Exon 7 and 3′ region
Tajima *D* in *japonica*	–1.987*	–2.286*	–1.878*
Tajima *D* in *indica*	–1.889*	–2.014*	–2.151*
$\pi_{japonica}/\pi_{rufipogon}$	0.076	0.359	0.08
$\pi_{indica}/\pi_{rufipogon}$	0.119	0.327	0.348
Length (bp)	587	883	626

Figure 3 Molecular domestication and grain-filling of introgression lines. (**a**) Nucleotide diversity (π) was measured in the *GIF1* promoter, exon 3, exon 7 and the 3′ UTR regions of 9 wild rice (*O. rufipogon*), 22 *japonica* and 14 *indica* varieties. Tajima *D* statistic and ratios of $\pi_{japonica}$ to $\pi_{rufipogon}$ and π_{indica} to $\pi_{rufipogon}$ were determined. *$P < 0.05$. (**b**) Grain-filling process (weight in grams of 1,000 brown grains) of wild-type Teqing and introgression lines SW19 and SW20 containing the *GIF1* locus from *O. rufipogon* (Hainan 1), which were developed from a backcrossing breeding program with Teqing as the recurrent parent. (**c**) Grain weight of wild-type Teqing and SW19 and SW20 introgression lines. *$P < 0.05$. (**d**) Real-time PCR detection of *GIF1* transcript levels in the recurrent parent Teqing and the introgression lines SW19 and SW20. Data in **b**–**d** are shown as means ± s.e.m. (**e**) *GIF1* transcripts detected by *in situ* hybridization in Teqing and the introgression lines SW19 and SW20. The sense probe was used as the control (left). For **d** and **e**, three independent experiments were repeated with similar results. Scale bars represent 100 μm.

domestication[2,4]. Among the domestication traits, the genes controlling seed shattering[20,21] and color[22] have been isolated. Recently, the *Waxy* gene controlling glutinous grain was also shown to have been artificially selected[23]. Because grain-filling directly contributes to grain weight, the *GIF1* gene might also have been a target for artificial selection. To test this theory, we analyzed the signature of past selection in *GIF1* sequences from a panel of *O. sativa indica* and *japonica* rice varieties and wild rice (*O. rufipogon*) with different origins (**Supplementary Table 4** online). The *GIF1* promoter region showed strong evidence of a past selective sweep, with a significant Tajima *D* statistic in both *indica* and *japonica* rice (**Fig. 3a**) compared to the Tajima *D* in 111 random gene fragments from *indica* (–0.3382) and *japonica* (–0.2642)[24]. The silent-site diversity (π) at the *GIF1* promoter regions in *japonica* and *indica* was only 8.1% and 12.7%, respectively, of the diversity in *O. rufipogon*. These levels were far below the 42% (*japonica*) and 48% (*indica*) observed for random gene fragments across the *O. sativa* genome[24]. Less evidence for a selective sweep was observed in exon 3, exon 7 and the 3′ region compared to the random gene fragments. Similar domestication signatures have been observed for other domestication genes, including rice *Waxy* and maize *tb* and *tga*[23,25,26].

Because wild rice flowers asynchronously and sets only a few small seeds that are easily shattered, it is difficult to study its grain-filling process. To further evaluate the contribution of *GIF1* to rice domestication, we generated introgression lines SW19 and SW20 carrying the *GIF1* locus of *O. rufipogon* (Hainan 1) with the Teqing variety in the backcross progeny (BC3F8)[27]. The grain-filling and weight of the two introgression lines were significantly lower (by 13.4–16.0% and 17.6–19.1%, respectively) than those of the parent Teqing (**Fig. 3b**,**c**). Furthermore, the wild rice *GIF1* allele was semidominant over the cultivated allele (**Supplementary Fig. 9** online). To further investigate the difference between the wild rice gene and its cultivated counterpart, we compared the expression pattern of *GIF1* in the Teqing and introgression lines by *in situ* hybridization. The wild rice *GIF1* showed a broader expression pattern in filling grains, being detected not only in ovular vascular trace but also in the pericarp and endosperm tissues, whereas the domesticated *GIF1* was mainly confined to the ovular trace during the filling stage (**Fig. 3**). Real-time PCR quantitatively confirmed expression levels of the wild and cultivated alleles (**Fig. 3d**). Similarly elevated expression of wild rice *GIF1* was found in other tissues of the introgression lines (data not shown). These results strongly suggest that *GIF1* is a domestication gene and that the restricted expression pattern of the cultivated *GIF1* gene was caused by nucleotide changes in its promoter during rice domestication.

We next conducted fine mapping of the locus responsible for reduced grain weight in the introgression lines, using a large population of the introgression line backcross (BC4F2/F3). On the basis of phenotyping (grain weight) and genotyping, we narrowed the locus decreasing grain weight down to an ~86-kb region that is flanked by the markers *SSLP1* and *CAPS8* and cosegregates with *CAPS1*, where *GIF1* is centered (**Supplementary Fig. 10** and **Supplementary Table 5** online). The mapping experiment revealed that the wild rice *GIF1* is most likely to be the gene that decreases grain weight in the introgression lines. We further analyzed the effect of cultivated alleles on grain-filling using three introgression lines in a recurrent Huajingxian 74 (*indica*) background with two *japonica* alleles and one *indica* allele with distinct origins (in total, two *japonica* alleles and two *indica* alleles). We found that these introgression lines showed no difference in grain weight, suggesting that the cultivated rice alleles of *indica* and *japonica* have the same effect on grain-filling (**Supplementary Fig. 11** online). Taken together, these data strongly suggest that *GIF1* is a target of domestication selection. Similarly, a tomato invertase gene

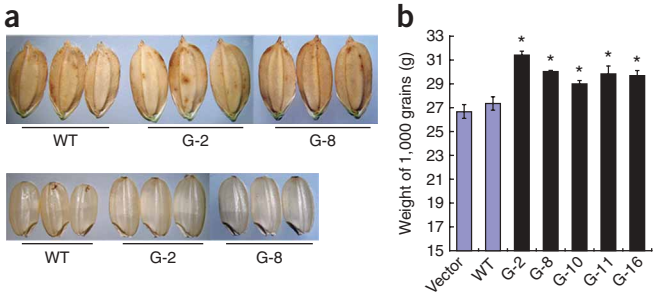


Figure 4 Increased grain size and weight in transgenic rice overexpressing *GIF1*. (**a**) Grains of two representative transgenic lines and wild-type (WT) rice. Transgenic plants had larger grains (see **Supplementary Fig. 12** for gene expression and grain sizes). (**b**) Grain weight (weight in grams of 1,000 grains) of five representative transgenic lines, wild-type rice and the transgenic control carrying the empty vector. Data are shown as means ± s.e.m. $^*P < 0.05$.

regulating sugar content and yield[2,14] and a maize starch synthesis gene[28] are also proposed to have been subjected to domestication.

Because *GIF1* contributes to grain weight, we further investigated whether it could improve grain yield. We generated transgenic rice lines that overexpress *GIF1* from its native promoter. The transgenic lines had larger and heavier grains compared to the wild-type rice (**Fig. 4** and **Supplementary Fig. 12** online), in sharp contrast to the transgenic plants that ectopically expressed *GIF1* from the *35S* promoter (**Supplementary Fig. 7**). We propose that the restricted expression pattern of the *GIF1* gene in the ovular vascular trace is the key to increased grain weight. Consistent with our interpretation, transgenic rice plants that ectopically expressed the cultivated *GIF1* gene from the rice *Waxy* promoter also had smaller grains (data not shown). Thus, our study suggests that crop yield can be enhanced by manipulating a sucrose-partitioning gene.

Grain-filling is a key determinant of rice production. In this study, we showed that the rice grain-filling gene, *GIF1*, encodes a cell-wall invertase that regulates sugar levels in specific tissues, most evidently in the ovular vascular and lateral stylar vascular traces of the developing grain and in the rapidly elongating internodes and roots, where a large amount of sugars is required to support cell division and growth. Assimilated carbon, mainly in the form of sucrose, is transported from the leaf (source) to the vascular trace of seed (sink)[17,18], where sucrose is hydrolyzed in the extracellular space into monosaccharides, which are then transported into the endosperm for starch synthesis. Our data reveal that GIF1 is a key regulator of this process and has a role in sucrose unloading, which is important in grain development. The rice genome contains eight predicted *CIN* genes[16]. Although other *CIN* genes contribute to unloading sucrose or maintaining sucrose homeostasis, this mechanism seems to be insufficient to fully circumvent the effect of the *gif1* mutation, resulting in the major phenotype of the *gif1* mutant, similar to the *flo4* mutant[29]. This suggests that the *OsCIN* genes are functionally differentiated, probably because of differing expression patterns[16].

Our study indicates that *GIF1* was most likely subjected to selection for better grain-filling to achieve a good harvest in cultivated rice. The domestication process probably selected for the accumulation of mutations in the *GIF1* regulatory region. Changes in expression patterns have been frequently observed in domestication genes[2,30]. An extensive microarray survey of gene expression indicates that many genes encoding proteins involved in sugar and energy metabolism have higher expression levels in *O. rufipogon* than in cultivated rice (X. Deng, Yale University and National Institute of Biological Sciences, China, personal communication). The restricted expression pattern of *GIF1* in the vascular bundle should facilitate sucrose unloading favoring grain-filling, whereas the wild rice *GIF1* allele might promote energy metabolism. The *GIF1* gene could increase yield potential through improved grain-filling in the transgenic lines (**Fig. 4a**,**b**), providing experimental evidence that a domestication-like agronomic trait gene can still be altered by molecular breeding to improve yield in a modern variety. *GIF1* might be particularly useful for breeding high-yield hybrid rice.

METHODS

Plant materials and sugar and starch assays. The *gif1* mutant was obtained from the *japonica* Zhonghua 11 gamma radiation–induced mutant population[12]. A mapping population was generated from the cross between *gif1* and Zhenshan97 (*indica*). All of the plants were grown in the paddy field to ensure the grain-filling and weight phenotypes. Developing grains were harvested, immediately frozen in liquid nitrogen and stored at –80 °C until use. Sugar content in grains without hulls and starch levels of mature grains were measured.

Mapping and cloning of *GIF1*. *GIF1* was primarily mapped with simple sequence repeat markers using 300 F2 *gif1* individuals and was further placed in a 32-kb region between the markers *CAPS4* and *CAPS8*. *GIF1* was also shown to cosegregate with *CAPS7* using 900 F2 mutant plants. The candidate gene was amplified and sequenced from both the *gif1* and wild-type genomic DNA using the sequencing primers (**Supplementary Table 6** online).

Complementation test and transgenic expression. The rice Nipponbare (*japonica*) BAC al662945 bearing *GIF1* was digested to isolate a 9-kb genomic DNA fragment containing the entire *GIF1* coding region, the 3-kb promoter region and the 800-bp 3′ region. The fragment was inserted into the binary vector pCAMBIA1301 to generate the plasmid *pGIF-GIF1*. The plasmid and the empty vector were introduced into the *gif1* mutant by *Agrobacterium tumefaciens*–mediated transformation. More than 20 independent lines were obtained for *pGIF1-GIF1* and five for pCAMBIA1301. The plasmid *pGIF1-GIF1* was also transformed into the wild-type cultivar TP309 (*japonica*) to generate more than 15 independent overexpression lines. The plasmid *p35S-GIF1*, containing the 2-kb *GIF1* full-length cDNA inserted into the overexpression vector 35S-C1301 (ref. 12), was transformed into TP309 to generate 20 independent lines ectopically expressing *GIF1*. The *GIF1* coding region and 3′ region were fused to the rice *Waxy* promoter to generate the chimeric fusion *pWx-GIF1* in pCAMBIA1301, which was then transformed into TP309 to generate 15 independent ectopic expression lines. All transgenic materials were assayed in the second (T1) or third (T2) generations with 24 sibling plants.

Subcellular localization of GIF1. The GIF1-GFP fusion was made by in-frame fusion of the 1.8-kb full-length *GIF1* cDNA with GFP (GenBank accession no. U87973). The fusion gene was inserted into the vector 35S-C1301. The construct was introduced into TP309 to generate 12 independent transgenic lines. The root tips of the transgenic plants were incubated in 25% sucrose for cell plasmolysis and then observed under a confocal laser microscope (LSM510, Zeiss).

Promoter activity. A 2.4-kb *GIF1* promoter region was fused to the *GUS* reporter gene with the nopaline synthase terminator and cloned into pCAMBIA1300 to generate the plasmid *pGIF1-GUS*, which was introduced into TP309 to generate ten independent transgenic lines. GUS activity in transgenic plants was detected by histochemical assay.

Invertase activity assay. The caryopses were ground in extraction buffer, and the extract was centrifuged at 12,000*g* for 10 min. The pellet was washed twice and then resuspended in extraction buffer. Insoluble invertase activity was assayed.

RNA analysis and *in situ* hybridization. Total RNA was prepared from rice tissues using TRIzol reagent (GIBCO BRL) according to the manufacturer's protocol. For RT-PCR, 1–5 μg of total RNA was used for cDNA synthesis with the SuperScript III System (Invitrogen). Real-time PCR was carried out with

primers and SYBR Premix Ex Taq system (Takara; **Supplementary Table 6**). For *in situ* hybridization, the 3′ end of the *GIF1* cDNA was subcloned into pGEM-Teasy (Sigma) and used as the template to generate sense and antisense RNA probes. RNA *in situ* hybridization was carried out using digoxigenin-labeled sense and antisense probes on 8-mm sections of Teqing and introgression line grains (7 DAP). The slides were observed and photographed under a bright-field microscope.

Evaluation of domestication. The promoter, exon 3, exon 7 and 3′ regions were amplified and sequenced using the PCR primers (**Supplementary Table 6**) in a set of 14 *indica*, 22 *japonica* and 9 wild rice germplasms (**Supplementary Table 4**). Haplotype diversity was calculated for *O. rufipogon*. Nucleotide diversity (π) and Tajima *D* statistic were calculated using DnaSP software version 4.0.

Introgression line screening and fine mapping. The introgression lines SW19 and SW20 containing the wild rice *GIF1* allele were screened from the introgression line population using simple sequence repeat markers[27]. The introgression lines progenies (BC3F8) were compared with the recurrent parent Teqing for grain-filling. Plants homozygous for the cultivated *GIF1* allele, homozygous for the wild rice allele, and heterozygous for both alleles were selected by PCR from the backcross population (BC4F2). Grain weights of these plants were measured to determine the effects of the alleles. A total of 5,384 progenies of BC4F2/F3 heterozygous plants were used for genotype and phenotype analysis to fine-map the QTL responsible for decreased grain weight. Additional introgression lines in the recurrent Huajingxian 74 (*indica*) background with two *japonica* alleles and two *indica* alleles were also selected, and grain weights were measured to further evaluate the effect of the cultivated *GIF1* alleles on grain-filling.

Additional analysis. Microarray analysis to determine gene expression changes in the mutant plants during grain-filling is described in **Supplementary Methods**.

Accession codes. GenBank: sequences have been deposited under accession codes EU095553, EU095554, EU095555, EU095556, EU095557, EU095558, EU095559, EU095559, EU095560, EU95561, EU095562, EU095563, EU095564, EU095565, EU095566, EU095567, EU095568, EU095569, EU095570, EU095571, EU095572, EU095573, EU095574, EU095575, EU095576, EU095577, EU095578, EU095579, EU095580, EU095581, EU095582, EU095583, EU095584, EU095585, EU095586, EU095587, EU095588, EU095589, EU095590, EU095591, EU095592, EU095593, EU095594, EU095595, EU095596. NCBI GEO: the complete set of microarray data has been deposited in a MIAME-compliant format under accession codes GSE9498, GSM240994, GSM240995, GSM240996, GSM240997, GSM240998, GSM240999.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank J.Y. Li, T. Sang and J.M. Li for critical reading of the manuscript and helpful suggestions; B. Han for the rice BAC clone; Z.Y. Wang for the *Waxy* promoter; D. Luo for help with *in situ* hybridization; X.Y. Gao and X.S. Gao for assistance with scanning electron microscopy and confocal laser microscopy; X.M. Zhang, L.J. Zeng and S.H. Ye for rice growth; and H.Q. Zheng for assistance with sugar measurement. This work was supported by grants from the Ministry of Science and Technology of China (2007AA02Z162, 2006AA10A102 and 2007AA10Z187), grants from the National Natural Science Foundation of China (30721061 and 30623006) and the Reproductive Development Project of the Shanghai Institutes for Biological Sciences.

AUTHOR CONTRIBUTIONS

E.W. and Z.H. conceived the research project, designed experiments and analyzed the data. E.W. carried out field phenotyping, genetics, gene cloning and functional and molecular evolution experiments. X.Z. screened the mutant. J.W. and L.W. conducted the genetic and field phenotype analyses. W. Hao, H.L. and G.Z. developed the introgression lines. Q.L. helped with the microarray assay. L.Z. helped with field testing. W. He helped with *in situ* hybridization. H.M. contributed to the funding and discussed the experiments. B.L. helped with wild rice analysis. Z.H. oversaw the entire study.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Takai, T., Fukuta, Y., Shiraiwa, T. & Horie, T. Time-related mapping of quantitative trait loci controlling grain-filling in rice (*Oryza sativa* L.). *J. Exp. Bot.* **56**, 2107–2118 (2005).
2. Doebley, J.F., Gaut, B.S. & Smith, B.D. The molecular genetics of crop domestication. *Cell* **127**, 1309–1321 (2006).
3. Yano, M. Genetic and molecular dissection of naturally occurring variation. *Curr. Opin. Plant Biol.* **4**, 130–135 (2001).
4. Sweeney, M. & McCouch, S. The complex history of the domestication of rice. *Ann. Bot. (Lond.)* **100**, 951–957 (2007).
5. International Rice Genome Sequencing Project. The map based sequence of the rice genome. *Nature* **436**, 793–800 (2005).
6. Xue, W. *et al.* Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. *Nat. Genet.* **40**, 761–767 (2008).
7. Fan, C. *et al. GS3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor. Appl. Genet.* **112**, 1164–1171 (2006).
8. Li, J., Thomson, M. & McCouch, S.R. Fine mapping of a grain-weight quantitative trait locus in the pericentromeric region of rice chromosome 3. *Genetics* **168**, 2187–2195 (2004).
9. Song, X.J., Huang, W., Shi, M., Zhu, M.Z. & Lin, H.X.A. QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* **39**, 623–630 (2007).
10. Ashikari, M. *et al.* Cytokinin oxidase regulates rice grain production. *Science* **309**, 741–745 (2005).
11. Nagata, K., Yoshinaga, S., Takanashi, J. & Terao, T. Effects of dry matter production, translocation of nonstructural carbohydrates and nitrogen application on grain-filling in rice cultivar Takanari, a cultivar bearing a large number of spikelets. *Plant Prod. Sci.* **4**, 173–183 (2001).
12. Zhu, Y. *et al. ELONGATED UPPERMOST INTERNODE* encodes a cytochrome P450 monooxygenase that epoxidizes gibberellins in a novel deactivation reaction in rice. *Plant Cell* **18**, 442–456 (2006).
13. Cheng, W.-H., Taliercio, E.W. & Chourey, P.S. The *miniature1* seed locus of maize encodes a cell wall invertase required for normal development of endosperm and maternal cells in the pedicel. *Plant Cell* **8**, 971–983 (1996).
14. Fridman, E., Carrari, F., Liu, Y.S., Fernie, A.R. & Zamir, D. Zooming in on a quantitative trait for tomato yield using interspecific introgressions. *Science* **305**, 1786–1789 (2004).
15. Roitsch, T. & Gonzalez, M.C. Function and regulation of plant invertases: sweet sensations. *Trends Plant Sci.* **9**, 606–613 (2004).
16. Cho, J.I. *et al.* Molecular cloning and expression analysis of the cell wall invertase gene family in rice (*Oryza sativa* L.). *Plant Cell Rep.* **24**, 225–236 (2005).
17. Sturm, A. & Tang, G.Q. The sucrose-cleaving enzymes of plants are crucial for development, growth and carbon partitioning. *Trends Plant Sci.* **4**, 401–407 (1999).
18. Krishnan, S. & Dayanandan, P. Structural and histochemical studies on grain-filling in the caryopsis of rice (*Oryza sativa* L.). *J. Biosci.* **28**, 455–469 (2003).
19. Tang, G.Q., Luscher, M. & Sturm, A. Antisense repression of vacuolar and cell wall invertase in transgenic carrot alters early plant development and sucrose partitioning. *Plant Cell* **11**, 177–189 (1999).
20. Konishi, S. *et al.* An SNP caused loss of seed shattering during rice domestication. *Science* **312**, 1392–1396 (2006).
21. Li, C., Zhou, A. & Sang, T. Rice domestication by reducing shattering. *Science* **311**, 1936–1939 (2006).
22. Sweeney, M.T., Thomson, M.J., Pfeil, B.E. & McCouch, S. Caught red-handed: *Rc* encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell* **18**, 283–294 (2006).
23. Olsen, K.M. *et al.* Selection under domestication: evidence for a sweep in the rice *waxy* genomic region. *Genetics* **173**, 975–983 (2006).
24. Caicedo, A.L. *et al.* Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genet.* **3**, 1745–1756 (2007).
25. Wang, R.L., Stec, A., Hey, J., Lukens, L. & Doebley, J. The limits of selection during maize domestication. *Nature* **398**, 236–239 (1999).
26. Wang, H. *et al.* The origin of the naked grains of maize. *Nature* **436**, 714–719 (2005).
27. Hao, W., Jin, J., Sun, S.Y., Zhu, M.Z. & Lin, H.X. Construction of chromosome segment substitution lines carrying overlapping chromosome segments of the whole wild rice genome and identification of quantitative trait loci for rice quality. *Zhi Wu Sheng Li Yu Fen Zi Sheng Wu Xue Xue Bao* **32**, 354–362 (2006).
28. Whitt, S.R., Wilson, L.M., Tenaillon, M.I., Gaut, B.S. & Buckler, E.S. Genetic diversity and selection in the maize starch pathway. *Proc. Natl. Acad. Sci. USA* **99**, 12959–12962 (2002).
29. Kang, H.G., Park, S., Matsuoka, M. & An, G. White-core endosperm *floury endosperm-4* in rice is generated by knockout mutations in the C-type pyruvate orthophosphate dikinase gene (OsPPDKB). *Plant J.* **42**, 901–911 (2005).
30. Clark, R.M., Wagler, T.N., Quijada, P. & Doebley, J. A distant upstream enhancer at the maize domestication gene, *tb1*, has pleiotropic effects on plant and inflorescent architecture. *Nat. Genet.* **38**, 594–597 (2006).