nature genetics

Copy number variation at the *GL7* locus contributes to grain size diversity in rice

Yuexing Wang[1–3,5], Guosheng Xiong[2–5], Jiang Hu[1,5], Liang Jiang[2,3], Hong Yu[2,3], Jie Xu[1], Yunxia Fang[1], Longjun Zeng[4], Erbo Xu[1], Jing Xu[1], Weijun Ye[1], Xiangbing Meng[2,3], Ruifang Liu[1,4], Hongqi Chen[1], Yanhui Jing[2,3], Yonghong Wang[2,3], Xudong Zhu[1], Jiayang Li[2,3] & Qian Qian[1,4]

© 2015 Nature America, Inc. All rights reserved.

Copy number variants (CNVs) are associated with changes in gene expression levels and contribute to various adaptive traits[1,2]. Here we show that a CNV at the *Grain Length on Chromosome 7* (*GL7*) locus contributes to grain size diversity in rice (*Oryza sativa* L.). *GL7* encodes a protein homologous to *Arabidopsis thaliana* LONGIFOLIA proteins, which regulate longitudinal cell elongation. Tandem duplication of a 17.1-kb segment at the *GL7* locus leads to upregulation of *GL7* and downregulation of its nearby negative regulator, resulting in an increase in grain length and improvement of grain appearance quality. Sequence analysis indicates that allelic variants of *GL7* and its negative regulator are associated with grain size diversity and that the CNV at the *GL7* locus was selected for and used in breeding. Our work suggests that pyramiding beneficial alleles of *GL7* and other yield- and quality-related genes may improve the breeding of elite rice varieties.

Naturally occurring CNVs in individuals have the potential to drive phenotypic diversity in populations[3]. CNVs in plants tend to occur within the loci of stress-related genes[4,5]. Genome-wide studies have suggested that CNVs in rice are relatively recent events that arose within breeding populations[6,7]. However, whether and how CNVs are associated with specific traits still remains to be elucidated. Grain appearance is a major quality parameter that is determined by grain size and endosperm translucency[8]. The cloning and characterization of quantitative trait loci (QTLs), including *GS3*, *GW2*, *GW5* (*qSW5*), *GS5*, *GW8*, *GL3.1* (*qGL3*), *TGW6* and *Chalk5* (refs. 9–17), has enriched knowledge of how grain yield and quality are regulated and enables breeders to develop high-yield and good-quality varieties[18].

The landrace Ping13 (P13) is an *indica* variety with superior grain length and appearance quality (**Supplementary Fig. 1a**). Using an F_2 population derived from a cross between P13 and Nipponbare (NPB; a *japonica* cultivar), we detected five QTLs associated with grain length and/or width in P13, including a major QTL affecting grain length on chromosome 7, *GL7* (**Supplementary Fig. 1b**). To clone *GL7*, we developed a near-isogenic line for *GL7* (NIL-*GL7*), containing an 82.9-kb chromosome segment from P13, through repeated backcrossing with NPB. The longer and narrower spikelet hull of NIL-*GL7* plants in comparison to NPB plants was the result of an increase in cell length and a decrease in cell width for epidermal cells of the outer and inner glumes (**Fig. 1a–c** and **Supplementary Fig. 2**). Comparison of grain size and weight among NPB, NPB × NIL-*GL7* (F_1) and NIL-*GL7* populations showed that *GL7* is a semidominant locus (**Fig. 1d–h** and **Supplementary Fig. 3a**). We found no significant difference in grain yield per plant, head-milled rice rate, amylose content, gel consistency or protein content between the NPB and NIL-*GL7* populations (**Supplementary Fig. 3b–f**). The chalky grain rate and chalkiness degree were significantly decreased in grain from NIL-*GL7* plants (**Fig. 1i,j**), and the starch granules in the NPB-*GL7* endosperm were larger and more tightly packed than those in the NPB endosperm (**Supplementary Fig. 3g**). These results suggest that *GL7* regulates grain appearance quality mainly through affecting the grain length to width ratio and the formation of starch granules in endosperm.

We cloned *GL7* through a map-based cloning approach. Briefly, we mapped *GL7* to an interval of 20.4 kb between the CAPS1 and 210Q markers, similar to the intervals described for the previously reported *qSS7* and *GS7* loci[19,20]; the newly identified locus contained two candidate genes, *Os07g0603300* (*LOC_ Os07g41200*) and *Os07g0603400* (*LOC_ Os07g41210*) (**Supplementary Fig. 4** and **Supplementary Table 1**). *Os07g0603400* encodes a protein of uncharacterized function, and *Os07g0603300* encodes a TON1 RECRUIT MOTIF (TRM)-containing protein that has 20–22% amino acid sequence identity with the LONGIFOLIA1 and LONGIFOLIA2 proteins from *Arabidopsis* (**Supplementary Fig. 5**), which regulate longitudinal cell elongation in leaves and siliques[21]. LONGIFOLIA2 (also known as TRM1) is able to bind microtubules and may be involved in directing cell expansion by recruiting TON1 to cortical microtubule arrays[22].

[1]State Key Laboratory of Rice Biology, China National Rice Research Institute, Hangzhou, China. [2]State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. [3]National Center for Plant Gene Research (Beijing), Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. [4]Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China. [5]These authors contributed equally to this work. Correspondence should be addressed to Q.Q. (qianqian188@hotmail.com), J.L. (jyli@genetics.ac.cn) or X.Z. (ricezxd@126.com).

Received 18 April; accepted 1 June; published online 6 July 2015; doi:10.1038/ng.3346

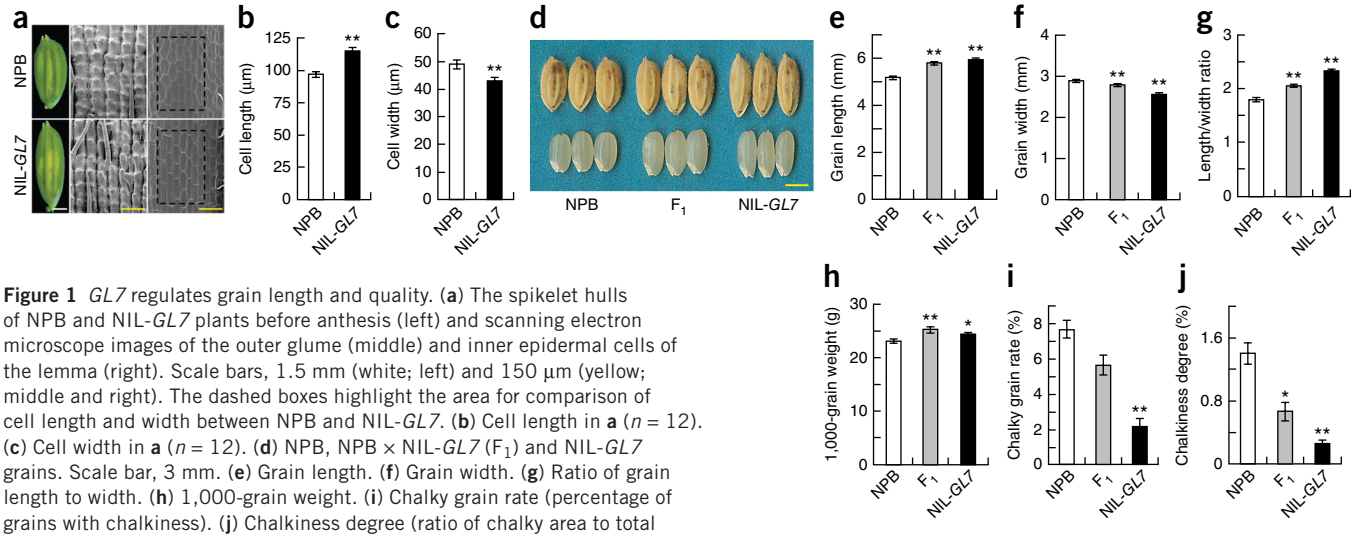

Figure 1 *GL7* regulates grain length and quality. (**a**) The spikelet hulls of NPB and NIL-*GL7* plants before anthesis (left) and scanning electron microscope images of the outer glume (middle) and inner epidermal cells of the lemma (right). Scale bars, 1.5 mm (white; left) and 150 μm (yellow; middle and right). The dashed boxes highlight the area for comparison of cell length and width between NPB and NIL-*GL7*. (**b**) Cell length in **a** ($n = 12$). (**c**) Cell width in **a** ($n = 12$). (**d**) NPB, NPB × NIL-*GL7* (F_1) and NIL-*GL7* grains. Scale bar, 3 mm. (**e**) Grain length. (**f**) Grain width. (**g**) Ratio of grain length to width. (**h**) 1,000-grain weight. (**i**) Chalky grain rate (percentage of grains with chalkiness). (**j**) Chalkiness degree (ratio of chalky area to total area of grain). All phenotypic data in **e**–**j** were measured from plants grown with 16.7 × 23 cm spacing in paddies under normal cultivation conditions. $n = 30$ in **e**–**h**, and $n = 3$ in **i** and **j**. All data represent means ± s.e.m. $^*P < 0.05$, $^{**}P < 0.01$, Student's *t* test.

We found two full-length cDNAs transcribed from *Os07g0603300* in P13 using a RACE assay, which we named *GL7-S1* and *GL7-S2* (**Supplementary Fig. 6a**,**b**). *GL7-S1* and *GL7-S2* encode exactly the same polypeptide and exhibited similar transcript levels (**Supplementary Figs. 5** and **6c**). To further understand the genomic structure underlying these two transcripts, we sequenced the DNA interval between the CAPS1 and 210Q markers and found a 17.1-kb tandem duplication containing *GL7* in NIL-*GL7* (**Fig. 2a** and **Supplementary Fig. 6d**–**f**), which explains how a 1,471-bp fragment could be amplified from P13 with primers NGSP11F and 210QCF but

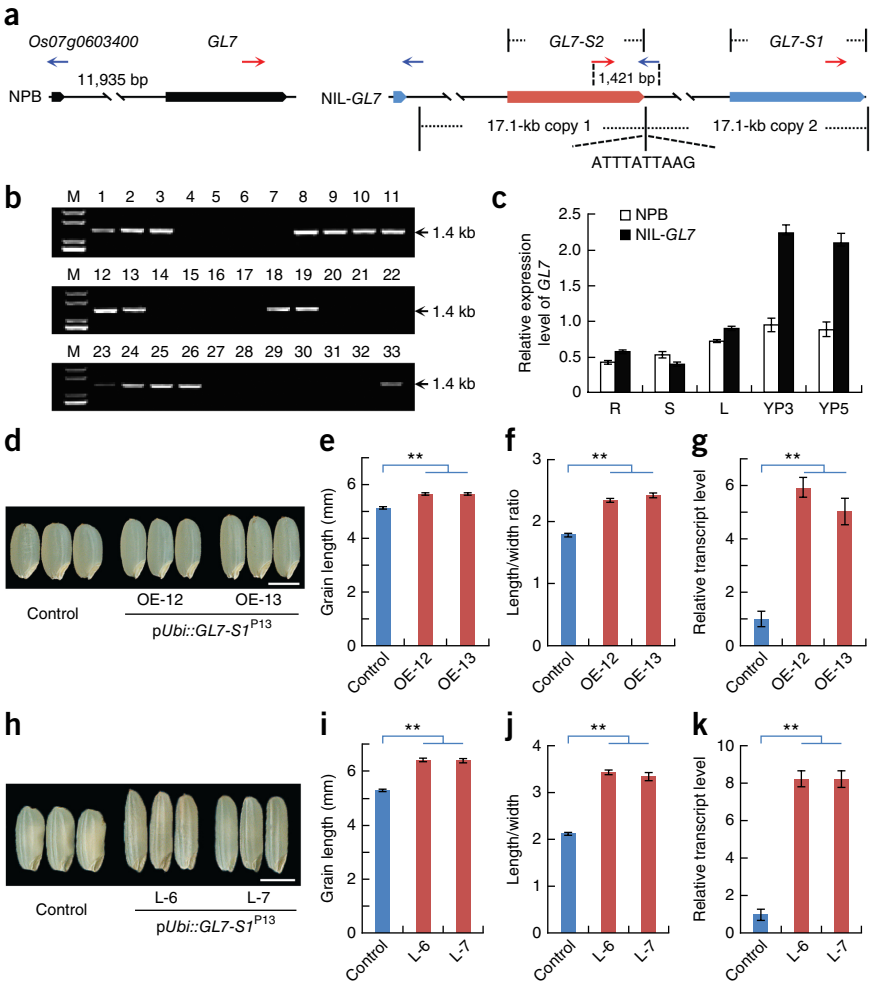

Figure 2 Elevated expression of *GL7* increases grain length. (**a**) Structures of the *GL7* loci in NPB and NIL-*GL7*. A 17.1-kb tandem duplication with a 10-bp insertion between the duplicates was found in NIL-*GL7*. The red arrows represent primer NGSP11F, and the blue arrows represent primer 201QCF. (**b**) PCR amplification with the 210QCF and NGSP11F primer pair using DNA samples from the 33 varieties listed in **Supplementary Table 3**. M, size marker. (**c**) *GL7* expression levels in organs from NPB and NIL-*GL7* plants. R, root; S, stem; L, leaf; YP3, young panicle at 3 cm in length; YP5, young panicle at 5 cm in length; $n = 3$. (**d**) Brown grains from two *GL7-S1*^{P13}–overexpressing lines and the control line, constituting NPB transformed with empty vector. Scale bar, 3 mm. (**e**) Grain length of the transgenic and control lines ($n = 30$). (**f**) Grain length to width ratio in the transgenic and control lines ($n = 30$). (**g**) Relative expression levels of *GL7* in young panicles of the transgenic and control lines were detected by qPCR, with data normalized to *Ubi* levels ($n = 3$). (**h**) Brown grains from two *GL7-S1*^{P13}–overexpressing lines and the control line, constituting ZF802 transformed with empty vector. Scale bar, 3 mm. (**i**) Grain length in the transgenic and control lines ($n = 30$). (**j**) Grain length to width ratio in the transgenic and control lines ($n = 30$). (**k**) Relative expression levels of *GL7* in young panicles of the transgenic and control lines were detected by qPCR, with data normalized to *Ubi* levels ($n = 3$). All data in **c**, **e**–**g** and **i**–**k** are presented as means ± s.e.m. $^{**}P < 0.01$, Student's *t* test.

© 2015 Nature America, Inc. All rights reserved.

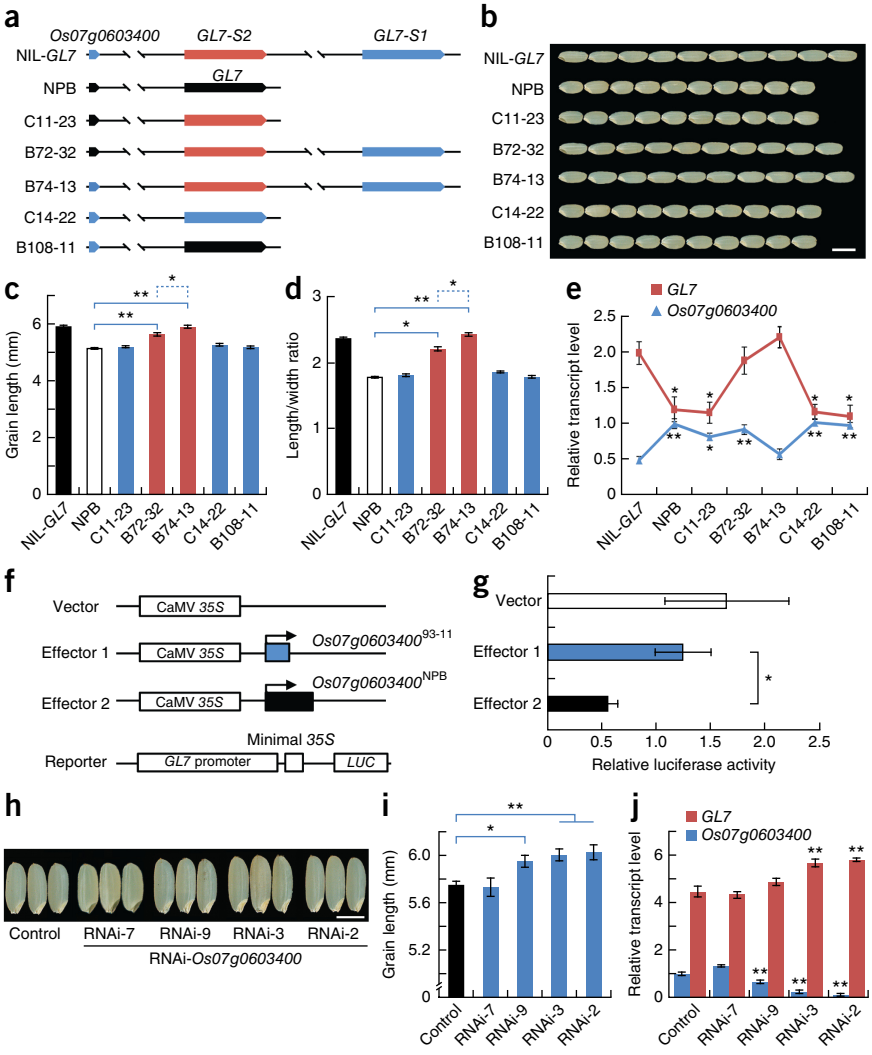

Figure 3 Correlation between genomic structure variations and grain length diversity. (**a**) Genomic structures at the *GL7* locus in NIL-*GL7*, NPB and BC_6F_3 recombinant lines. (**b**) Brown grains from NIL-*GL7*, NPB and recombinant lines, showing grain differences. Scale bar, 5 mm. (**c**) Grain length ($n = 30$). (**d**) Grain length to width ratio ($n = 30$). (**e**) Relative expression levels of *GL7* and *Os07g0603400* in 3-cm young panicles of the recombinant lines, NPB and NIL-*GL7* were detected by qPCR, with data normalized to *Ubi* levels ($n = 3$). (**f**) Plasmid constructs for the transient expression assay. CaMV, cauliflower mosaic virus. (**g**) Relative firefly luciferase activities in rice protoplasts, with data normalized to activity for cotransformed constitutively expressed *Renilla* luciferase. Data are shown as means ± s.e.m. ($n = 5$ technical replicates). (**h**) Brown grains from control plants and ones transgenic for RNA interference (RNAi) targeting *Os07g0603400* on the NIL-*GL7* background. Scale bar, 3 mm. (**i**) Grain length ($n = 30$). (**j**) Relative expression levels of *GL7* and *Os07g0603400* in 3-cm young panicles of the transgenic and control lines were detected by qPCR. The expression levels of *GL7* and *Os07g0603400* were normalized to *Ubi* levels ($n = 3$). All data in **c**–**e**, **i** and **j** are presented as means ± s.e.m. $^*P < 0.05$, $^{**}P < 0.01$, Student's *t* test.

not from NPB (**Fig. 2a** and **Supplementary Table 2**). These results suggest that the CNV at the *GL7* locus in P13 is very likely responsible for the long-grain phenotype. This hypothesis was further confirmed through analyzing a total of 96 varieties, of which all 17 varieties containing the CNV at the *GL7* locus had long grains (**Fig. 2b** and **Supplementary Table 3**). The CNV at the *GL7* locus explained the unexpected recombinant events observed in the mapping populations for *qSS7* and *GS7* (refs. 19,20).

To understand the function of *GL7*, we analyzed its expression pattern and found that NIL-*GL7* had a higher level of *GL7* expression in young panicles than NPB (**Fig. 2c**), which was similar to the level found in a previous study of *qSS7* (ref. 19). We transformed NPB with either *Ubiquitin* promoter–driven *GL7-S1* (p*Ubi*::*GL7-S1*^{P13}) or *GL7-S2* (p*Ubi*::*GL7-S2*^{P13}), which resulted in ~10% longer grain length in transgenic plants (**Fig. 2d**–**g** and **Supplementary Fig. 7**). Furthermore, we found that the transformation of p*Ubi*::*GL7-S1*^{P13} into an *indica* variety, Zhefu802 (ZF802), resulted in a ~21% increase in grain length (**Fig. 2h**–**k**). The degree of the increase in grain length was well correlated with the expression level of *GL7* in transgenic plants (**Supplementary Fig. 8**), and the increase in grain length could be explained by an increase in the length of epidermal cells in glumes (**Supplementary Fig. 9a**,**b**). Moreover, overexpression of *GL7* promoted the formation of densely and regularly packed starch granules (**Supplementary Fig. 9c**,**d**). It is possible that the elevated expression of *GL7* might affect the arrangement of microtubules, thereby leading to extension of the grain shape and alteration of starch granule stacking. We also found cell length increased in the flag leaves but not the main stems of the *GL7* overexpression lines, thereby explaining why the lengths of flag leaves were increased but those of the main stems were not (**Supplementary Figs. 10** and **11**).

Detailed analysis of recombinant lines showed that lines B72, B74 and C20 with the *GL7* duplication had long grains (**Supplementary Fig. 12**), suggesting that elevated expression of *GL7* through increased copy number is sufficient to increase grain length. Further characterization of the homozygous progeny of recombinant lines (**Fig. 3a**–**e** and **Supplementary Fig. 13**) indicated that grain length diversity in these lines is more likely determined by the CNV than the SNPs in the *GL7* locus differing between NPB and NIL-*GL7*.

To our surprise, we found that the expression level of *GL7* was negatively correlated with the *Os07g0603400* expression level in the recombinant lines (**Fig. 3e** and **Supplementary Fig. 13**), which is consistent with the finding that NIL-*GL7* had a lower *Os07g0603400* expression level in young panicles than NPB (**Supplementary Fig. 14**). Comparison of the *Os07g0603400* DNA sequence between short-grain varieties (such as NPB) and long-grain varieties (such as 93-11) identified a null mutation that results in a truncated protein in long-grain varieties (**Supplementary Fig. 15**). Overexpression of NPB *Os07g0603400* indeed repressed expression from a luciferase reporter gene under the control of the *GL7* promoter (p*GL7*::*LUC*) in rice protoplasts, but overexpression of 93-11 *Os07g0603400* encoding a truncated protein showed no effect (**Fig. 3f**,**g**). These results suggest that *Os07g0603400* may function as a negative regulator of *GL7*. This hypothesis was further confirmed by knocking down *Os07g0603400* expression in NIL-*GL7*, which increased the expression level of *GL7*

© 2015 Nature America, Inc. All rights reserved.

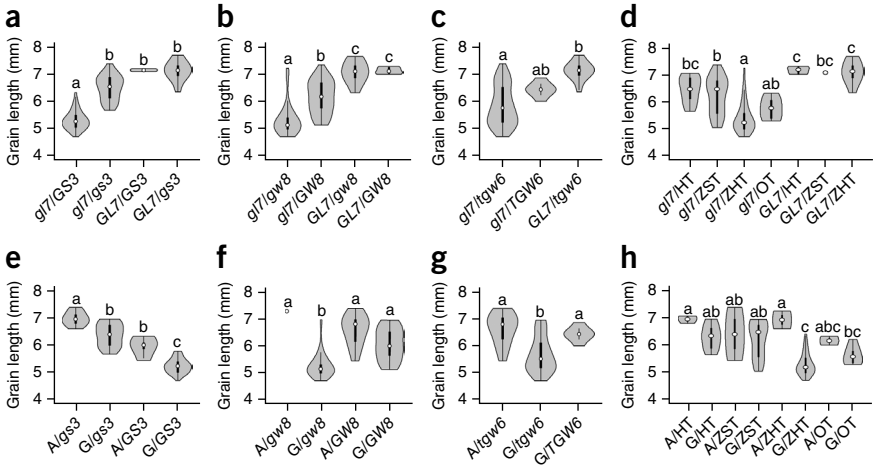

Figure 4 Relationship between *GL7* or *Os07g0603400* and other grain size–related loci. A box plot and a kernel density plot were generated as violin plots for different groups. (**a**–**d**) Relationship between *GL7* and other grain size–related genes, including *GS3* (**a**), *GW8* (**b**), *TGW6* (**c**) and *GS5* (**d**). HT, ZST, ZHT or OT indicates the different haplotype of *GS5*. (**e**–**h**) Relationship between *Os07g0603400* and other grain size–related genes, including *GS3* (**e**), *GW8* (**f**), *TGW6* (**g**) and *GS5* (**h**). A and G indicate the haplotypes of *Os07g0603400* in 93-11 and NPB, respectively. Different letters above columns indicate statistically significant differences between groups (Tukey's honestly significant difference (HSD) test, $P < 0.05$). Landraces, genotypes and phenotypes are listed in **Supplementary Table 3**.

and led to an increase in grain length and improvement of grain appearance quality (**Fig. 3h–j** and **Supplementary Fig. 16**). However, when *Os07g0603400* was knocked down in NPB or overexpressed in NIL-*GL7*, transgenic lines showed no significant changes in cell or grain length, probably owing to little effect on the expression of *GL7* (**Supplementary Figs. 17** and **18**).

To dissect the genetic interaction between *GL7* and other genes controlling grain length, we analyzed the coding sequences of *GL7* and *Os07g0603400* and the haplotypes of *GS3*, *GW8*, *GS5*, *TGW6*, *GW2* and *GL3.1* (*qGL3*) in 96 accessions (**Supplementary Table 4**). Phylogenetic analysis indicated that these loci might cooperatively regulate grain length in rice, as accessions with the same haplotype for a given gene exhibited grain length diversity (**Supplementary Fig. 19**). To identify the major contributors and dissect their interactions, we analyzed the association between grain length and allelic variants of grain length–related genes. Our results suggest that the beneficial alleles of *GL7*, *GS3*, *GW8*, *TGW6* and *GS5* contribute to grain length in these accessions and that variations at *GS3* and *GL7* are major factors affecting grain length diversity (**Fig. 4a–d**, **Supplementary Fig. 19** and **Supplementary Table 3**). Furthermore, we found that the beneficial allele of *GL7* had an epistatic effect on grain length (**Fig. 4a–d**) and that the effect of *Os07g0603400* was independent of that for *GS3* (**Fig. 4e–h** and **Supplementary Fig. 20**). On the basis of understanding how different alleles of the grain length–related genes interact, it is possible to breed varieties with optimal grain shape by pyramiding different combinations of beneficial alleles of these genes using marker-assisted selection.

Among the 96 accessions studied, the allelic variant of *Os07g0603400* affecting grain length could be found even in *Oryza rufipogon* accessions, indicating that natural variation in *Os07g0603400* could be an ancient event (**Supplementary Table 3**). Moreover, we found that the US long-grain accessions CPSLO17 and Cypress contained a *GL7* locus that was identical to the one in P13, suggesting that P13, CPSLO17 and Cypress may be derived from a common ancestor (**Supplementary Fig. 21** and **Supplementary Table 4**). Interestingly, we found that a short-grain accession, Kanto146 (KT146), derived from CPSLO17 (ref. 23; **Supplementary Fig. 22a**), had a cDNA sequence for *GL7* identical to *GL7-S1* in

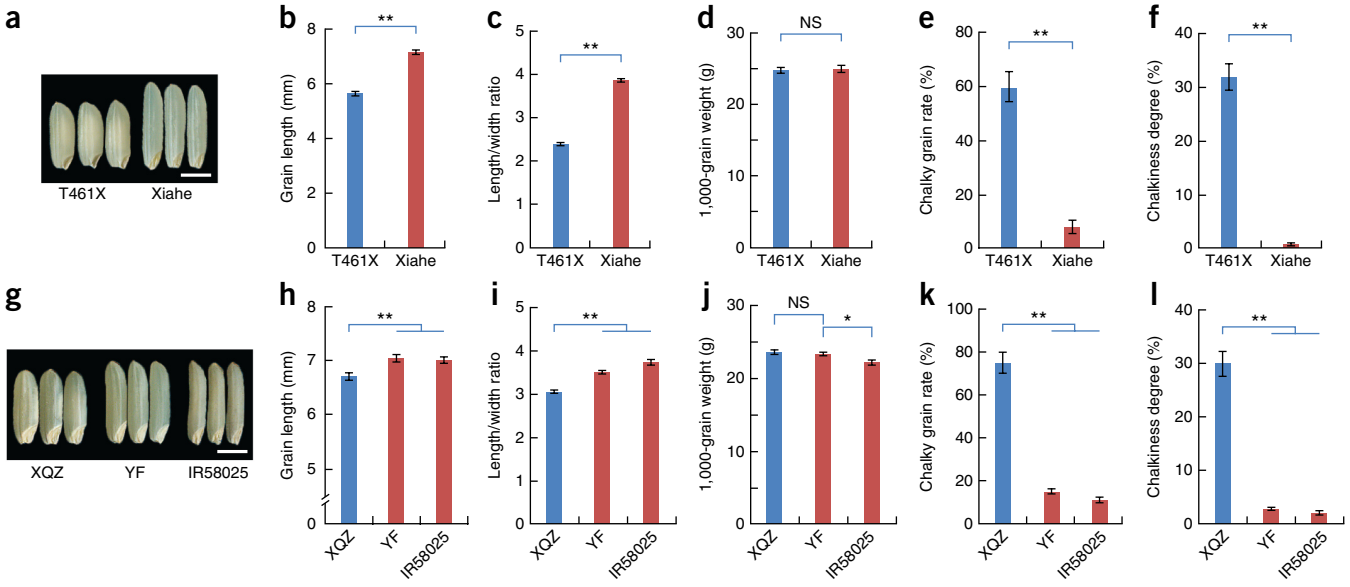

Figure 5 Application of *GL7* in breeding. (**a**) Brown grains from the T461X and Xiahe lines. Scale bar, 3 mm. (**b**) Grain length. (**c**) Ratio of grain length to width. (**d**) 1,000-grain weight. (**e**) Chalky grain rate. (**f**) Chalkiness degree. (**g**) Brown grains from the XQZ, YF and IR58025 lines. Scale bar, 3 mm. (**h**) Grain length. (**i**) Ratio of grain length to width. (**j**) 1,000-grain weight. (**k**) Chalky grain rate. (**l**) Chalkiness degree. Data for **b**–**f** and **h**–**l** are shown as means ± s.e.m. $n = 30$ in **b**–**d** and **h**–**j**, and $n = 3$ in **e**, **f**, **k** and **l**. NS, not significant, $^{*}P < 0.05$, $^{**}P < 0.01$, Student's *t* test.

© 2015 Nature America, Inc. All rights reserved.

P13 but had lost the 17.1-kb duplication (**Supplementary Fig. 21**). This genomic structure was similar to that found in the recombinant line C14-22 (**Fig. 3a–e** and **Supplementary Fig. 21**), indicating that KT146 might have lost one copy of the 17.1-kb duplication from CPSLO17 during the breeding process. Given that Americans and Japanese have differing preferences for grain length[9], the CNV at the *GL7* locus that confers different grain lengths in CPSLO17 and KT146 might have been selected by breeders in the United States and Japan, respectively, providing an example of how naturally occurring CNVs controlling important agronomic traits could be selected and employed by breeders.

We crossed a relatively high-yield restorer line, T461X, with P13 and found that introduction of *GL7* alone could significantly improve grain quality without compromise of grain yield (**Fig. 5a–f** and **Supplementary Fig. 22b**). Moreover, Yuefeng (YF), an *indica* sterile line widely used in southeast China because of its particularly good appearance quality[24], carried the beneficial allele of *GL7* deriving from IR58025, whereas Xieqingzao (XQZ), a formerly widely used high-yield sterile line, has relatively poor grain quality (**Supplementary Fig. 22c** and **Supplementary Table 3**). Combination of the beneficial *GL7* and *gs3* alleles with the allele for an unidentified grain weight–related gene derived from XQZ significantly improved grain appearance quality and grain weight for YF plants (**Fig. 5g–l**), suggesting that the beneficial allele of *GL7* had been applied in breeding and that pyramiding this beneficial allele with those for other yield- and quality-related genes could contribute to the breeding of elite rice varieties.

Our study demonstrates that the CNV occurring at the *GL7* locus affects the expression of two linked genes and contributes to grain size diversity. We also found that this locus has been selected and used in rice breeding. Further exploration of the association between CNVs and complex agronomic traits and dissection of the underlying molecular mechanisms would shed new light on the domestication and improvement of crops and provide more resources and approaches for breeding elite varieties.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. The sequence of the *GL7* coding region has been deposited in the GenBank database under accession KP899557.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank the National Medium Rice Genebank at the China National Rice Research Institute for providing rice accessions. This work was supported by grants from the Ministry of Science and Technology of China (2013CBA01400) and the National Natural Science Foundation of China (31221004, 31201004 and 31201194).

AUTHOR CONTRIBUTIONS

Yuexing Wang, G.X. and J.H. designed research, performed experiments, analyzed data and wrote the manuscript. L.J., Jie Xu, Y.F., L.Z., E.X., Jing Xu, W.Y., X.M., H.C. and Y.J. performed the experiments. H.Y., R.L. and Yonghong Wang designed research and analyzed data. J.L., Q.Q. and X.Z. conceived the project and designed research. J.L. and Q.Q. supervised the project, analyzed data and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Stranger, B.E. *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848–853 (2007).
2. Tang, Y.C. & Amon, A. Gene copy-number alterations: a cost-benefit analysis. *Cell* **152**, 394–405 (2013).
3. Weischenfeldt, J., Symmons, O., Spitz, F. & Korbel, J.O. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat. Rev. Genet.* **14**, 125–138 (2013).
4. McHale, L.K. *et al.* Structural variants in the soybean genome localize to clusters of biotic stress-response genes. *Plant Physiol.* **159**, 1295–1308 (2012).
5. Cook, D.E. *et al.* Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean. *Science* **338**, 1206–1209 (2012).
6. Yu, P. *et al.* Detection of copy number variations in rice using array-based comparative genomic hybridization. *BMC Genomics* **12**, 372 (2011).
7. Yu, P. *et al.* Genome-wide copy number variations in *Oryza sativa* L. *BMC Genomics* **14**, 649 (2013).
8. Fitzgerald, M.A., McCouch, S.R. & Hall, R.D. Not just a grain of rice: the quest for quality. *Trends Plant Sci.* **14**, 133–139 (2009).
9. 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).
10. Zhang, X. *et al.* Rare allele of *OsPPKL1* associated with grain length causes extra-large grain and a significant yield increase in rice. *Proc. Natl. Acad. Sci. USA* **109**, 21534–21539 (2012).
11. 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).
12. Shomura, A. *et al.* Deletion in a gene associated with grain size increased yields during rice domestication. *Nat. Genet.* **40**, 1023–1028 (2008).
13. Li, Y. *et al.* Natural variation in *GS5* plays an important role in regulating grain size and yield in rice. *Nat. Genet.* **43**, 1266–1269 (2011).
14. Ishimaru, K. *et al.* Loss of function of the IAA-glucose hydrolase gene *TGW6* enhances rice grain weight and increases yield. *Nat. Genet.* **45**, 707–711 (2013).
15. Qi, P. *et al.* The novel quantitative trait locus *GL3.1* controls rice grain size and yield by regulating Cyclin-T1;3. *Cell Res.* **22**, 1666–1680 (2012).
16. Wang, S. *et al.* Control of grain size, shape and quality by *OsSPL16* in rice. *Nat. Genet.* **44**, 950–954 (2012).
17. Li, Y. *et al. Chalk5* encodes a vacuolar H^+-translocating pyrophosphatase influencing grain chalkiness in rice. *Nat. Genet.* **46**, 398–404 (2014).
18. Zuo, J. & Li, J. Molecular genetic dissection of quantitative trait loci regulating rice grain size. *Annu. Rev. Genet.* **48**, 99–118 (2014).
19. Qiu, X., Gong, R., Tan, Y. & Yu, S. Mapping and characterization of the major quantitative trait locus *qSS7* associated with increased length and decreased width of rice seeds. *Theor. Appl. Genet.* **125**, 1717–1726 (2012).
20. Shao, G. *et al.* Allelic variation for a candidate gene for *GS7*, responsible for grain shape in rice. *Theor. Appl. Genet.* **125**, 1303–1312 (2012).
21. Lee, Y.K. *et al. LONGIFOLIA1* and *LONGIFOLIA2*, two homologous genes, regulate longitudinal cell elongation in *Arabidopsis*. *Development* **133**, 4305–4314 (2006).
22. Drevensek, S. *et al.* The *Arabidopsis* TRM1-TON1 interaction reveals a recruitment network common to plant cortical microtubule arrays and eukaryotic centrosomes. *Plant Cell* **24**, 178–191 (2012).
23. Imbe, T. *et al.* Development of a multipurpose high yielding rice variety "*Takanari*". *Bull. Natl. Inst. Crop Sci.* **5**, 35–51 (2004).
24. Cheng, J., Li, C., Wu, Y., Zheng, H. & Liang, S. Features and properties of *indica* rice sterile line Yuefeng A with high quality and its utilization. *Guangdong Agr. Sci.* **4**, 5–6 (2002).

© 2015 Nature America, Inc. All rights reserved.

ONLINE METHODS

Plant materials. Rice plants were cultivated at the Experimental Stations of the China National Rice Research Institute in Hangzhou, China, during the natural growing season and under field conditions with an interplant spacing of 16.7 × 23.3 cm for transplanting. An F_2 population was derived from the cross between a superior long-grain landrace cultivar from Sichuan province, P13, and a traditional *japonica* cultivar, NPB. We measured the grain length and width of 165 individuals in the F_2 population, and 5 major loci were identified with 115 simple sequences repeat (SSR) markers. NIL-*GL7* was developed from the BC_6F_2 generation by repetitive backcrossing to NPB with marker-assisted selection.

Trait measurement. Fully filled grains were used to measure 1,000-grain weight, and 30 randomly chosen brown grains were measured by electronic digital-display Vernier caliper to obtain grain length, width and thickness. Grain yield per plant was measured for 30 plants from 3 completely randomized blocks. Chalky grain rate, chalkiness degree, head-milled rice rate, amylose content, gel consistency and protein content were measured by the Rice Product Quality Supervision and Inspection Center, Ministry of Agriculture, China, according to the standard of NY/T 593-2002.

Histological analysis. Young spikelet hulls were fixed in FAA fixation solution (5% formaldehyde, 5% glacial acetic acid and 70% ethanol) for more than 24 h. Samples were dehydrated via a graded ethanol series (70%, 80%, 90% and 100%) over ~1.5 h, treated with dimethylbenzene and embedded in paraffin. Samples were then sectioned by a rotary microtome (Thermo Scientific, HM340E) and stained with sarranine and toluidine blue. Segments (2–3 mm in length) of the fully mature region of leaves were collected and fixed in FAA overnight, washed twice with PBS (pH 7.4) and dehydrated via a graded ethanol series. Segments were transferred to a solution of 50% ethanol and 50% clearing solution (20% lactic acid, 20% trichloroacetaldehyde monohydrate, 20% phenol, 20% dibutyl phthalate, 10% xylene and 10% benzyl benzoate), mixed for ~3 h and placed in clearing solution for more than 3 h. For scanning electron microscopy, samples were fixed in 2.5% glutaraldehyde solution (pH 7.4) for at least 24 h and then processed according to the manual supplied with the device (Hitachi, S-3000N).

Mapping and cloning of *GL7*. Primary QTL analysis was carried out using 112 BC_3F_2 plants and 9 molecular markers (**Supplementary Table 1**). To fine map the *GL7* locus, recombinants were screened for the RM21930 and RM234 markers using 2,236 plants from the BC_4F_2 population. Seven new markers were developed (**Supplementary Table 1**), and the *GL7* locus was finally mapped to the interval between CAPS1 and 210Q using an additional 20,160 BC_6F_2 individuals. The genomic DNA fragments corresponding to the candidate gene in P13 and NPB were sequenced and analyzed.

Conventional molecular biology methods. Total RNA was extracted from various plant tissues from P13, NPB, NIL-*GL7* and the transgenic lines using the AxyPrep Multisource Total RNA Miniprep kit (Axygen). The full-length *GL7-S1* (3,117 bp) and *GL7-S2* (3,285 bp) transcripts were amplified from the first-strand cDNA derived from P13 and sequenced. Quantitative RT-PCR was carried out using SYBR Green QPCR mix (Bio-Rad), and *Ubi* was used as a control. For the RACE assay, the full-length transcripts of *GL7* were amplified by nested PCR with the SMARTer RACE cDNA Amplification kit (Clontech). High-purity template genomic DNA from P13 and NPB was extracted with the improved QIAamp DNA Mini kit (Qiagen). Southern bolt analysis was performed with DIG High Prime DNA Labeling and Detection Starter Kit II (Roche) according to the user's instructions. To construct p*Ubi*::*GL7-S1*^{P13} and p*Ubi*::*GL7-S2*^{P13}, the corresponding coding regions of *GL7* were amplified from P13 and cloned into the binary pTCK303 vector to generate the overexpression constructs[25]. Full-length cDNA for the *Os07g0603400* gene was amplified and cloned into the binary pTCK303 vector to generate the overexpression vector p*Ubi*::*Os07g0603400*^{P13}. A 289-bp fragment of *Os07g0603400* cDNA was used to construct the RNAi-*Os07g0603400* transgenic vector. All primers used are listed in **Supplementary Table 2**.

Rice transformation. *Agrobacterium tumefaciens*–mediated rice transformation was carried out with the transformation vectors as described previously[26], and a total of 5, 9 and 13 independent transgenic lines of p*Ubi*::*GL7-S1*^{P13}, p*Ubi*::*GL7-S2*^{P13} and RNAi-*Os07g0603400*, respectively, in the NPB background and 10 independent lines of p*Ubi*::*GL7-S1*^{P13} in the *indica* accession ZF802 were generated. In addition, a total of 12 and 10 independent transgenic lines of p*Ubi*::*Os07g0603400*^{P13} and RNAi-*Os07g0603400* were generated, respectively, on the NIL-*GL7* background.

Transient expression assays in rice protoplasts. The *GL7* promoter was amplified and inserted upstream of the CaMV minimal *35S* promoter (–90 to +1, with respect to the transcription start site) to drive expression of luciferase. The full-length coding sequence of *Os07g0603400* from NPB or 93-11 was cloned into the pUC-SPYCE vector[27]. Luciferase activity was measured using the Dual-Luciferase reporter kit (Promega) on a TD-20/20 Luminometer (Turner Designs).

25. Wang, Z. *et al.* A practical vector for efficient knockdown of gene expression in rice (*Oryza sativa* L.). *Plant Mol. Biol. Rep.* **22**, 409–417 (2004).
26. Hiei, Y., Ohta, S., Komari, T. & Kumashiro, T. Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J.* **6**, 271–282 (1994).
27. Walter, M. *et al.* Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. *Plant J.* **40**, 428–438 (2004).

© 2015 Nature America, Inc. All rights reserved.